Using insects to detect, monitor and predict the distribution of Xylella fastidiosa: a case study in Corsica

Astrid Cruaud, Anne Alicia Gonzalez, Martin Godefroid, Sabine Nidelet,
Jean-Claude Streito, Jean-Marc Thuillier, Jean-Pierre Rossi, Sylvain Santoni,
Jean Yves Rasplus

SCIENTIFIC REP**O**RTS

Using insects to detect, monitor and predict the distribution of *Xylella fastidiosa*: a case study in Corsica

Astrid Cruaud[1], Anne-Alicia Gonzalez[1,2], Martin Godefroid[1], Sabine Nidelet[1], Jean-Claude Streito[1], Jean-Marc Thuillier[1], Jean-Pierre Rossi[1], Sylvain Santoni[2] & Jean-Yves Rasplus[1]

We sampled *ca* 2500 specimens of *Philaenus spumarius* (Hemiptera: Aphrophoridae) throughout Corsica without *a priori* knowledge on the presence of symptoms on plants. We screened 448 specimens for the presence of *Xylella fastidiosa (Xf)* using qPCR and a custom nested PCR. qPCR appeared versatile and under-estimated the prevalence of *Xf*. Nested PCR showed that *Xf* was present in all populations. Molecular results were validated by prediction on the distribution of *Xf* made from tests conducted on plants, which shows the pertinence of using vectors in risk assessment studies. *Xf* was detected in tenerals and adults. Thus, *P. spumarius* could acquire *Xf* from its host plant, mostly *Cistus monspeliensis* in Corsica, which may act as reservoir for the next season. This contrasts with other observations and suggests that management strategies may have to be adapted on a case-by-case basis. At least two genetic entities and several variants of *Xf* not yet identified on plants were present in the insects, which suggests ancient introductions of *Xf* and a probable underestimation of the current diversity of the strains present in Corsica. Interestingly 6% of the specimens carried two subspecies of *Xf*. Studies are required to better characterize the strains present in Corsica and to determine how the disease was introduced, spread and why no sign of a potential epidemic was detected earlier. This study shows that, when sensitive enough methods are implemented, spittlebugs (and more specifically *P. spumarius* for which species distribution modelling shows it could be a good sentinel for Europe) can be used to predict and better assess the exact distribution of *Xf*. Furthermore, *Xf* multiply only in their foregut and does not become circulative, which facilitates its detection.

Xylella fastidiosa (*Xf*) (Xanthomonadaceae, Gammaproteobacteria) is a xylem-limited Gram-negative bacterium that causes disease in important crops and ornamental plants, such as Pierce's disease of grapevine, *Citrus* variegated chlorosis disease, phony peach disease, plum leaf scald as well as leaf scorch on almond or elm and *Quercus*[1,2]. *Xf* infects a large number of plants (more than 300 species from more than 60 plant families)[3]. However, the different genetic lineages exhibit narrower host-plant ranges[4]. The disease is endemic and widespread on the American continent and its biology, ecology and epidemiology have been extensively studied in the last forty years (reviewed in[1,2,5–8]).

In the last few years several subspecies of *Xf* have been detected in Europe (https://gd.eppo.int/taxon/XYLEFA/distribution, https://ec.europa.eu/food/plant/plant_health_biosecurity/legislation/emergency_measures/xylella-fastidiosa/latest-developments_en). An outbreak of *Xf pauca* was first identified in Apulia (southeastern Italy) in olive groves[9]. *Xf multiplex* was then detected on *Polygala myrtifolia* in Corsica[10] and subsequently in continental southern France. Large-scale studies conducted in 2015 further revealed that *Xf pauca* and *Xf fastidiosa-sandyi* (ST76) as well as possible recombinants were also present in France[11]. Recently, *Xf multiplex* was detected in eastern parts of the Iberian peninsula (region of Alicante) and *Xf multiplex*, *Xf pauca* and *Xf fastidiosa* were detected in the Balearic Islands[12]. *Xf fastidiosa* was also detected in Germany on a potted *Nerium*

[1]CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, University Montpellier, Montpellier, France. [2]INRA, UMR1334 AGAP, F-34398, Montpellier, France. Correspondence and requests for materials should be addressed to A.C. (email: astrid.cruaud@inra.fr)

oleander kept in glasshouse in winter (http://pflanzengesundheit.jki.bund.de/dokumente/upload/3a817_xylella-fastidiosa_pest-report.pdf) and several interception of infected coffee plants have been reported in Europe ([13,14], https://gd.eppo.int/taxon/XYLEFA/distribution).

The bacterium is transmitted to plants by xylem-sap feeding leafhoppers (Hemiptera, Cicadomorpha) and members of several families are known to transmit the disease from plants to plants[6]. In the Americas, Cicadellidae[15,16], spittlebugs (Cercopidae, Clastopteridae, Aphrophoridae)[17–20], and cicadas (Cicadidae)[19,21] have been shown to efficiently transmit *Xf*. In Europe, little is known about the vectors that efficiently transmit the bacterium. So far, of the 119 potential vectors that feed on xylem sap[10] only *Philaenus spumarius* (Linnaeus, 1758), the meadow spittlebug, has been identified as an effective vector of *Xf* in southern Italy[22,23]. Other studies are thus clearly needed to clarify whether other insects may play an important role in the epidemiology of *Xf*.

Usually, epidemiological survey of *Xf* is conducted on symptomatic plant material. Most frequently, the presence of the bacterium is assessed using qPCR targeting a small part of the gene *rimM* as designed in Harper *et al*.[24] (see Baldi & La Porta[25] for a review of the available methods and their advantages and drawbacks). Then, if wanted, fragments of seven housekeeping genes are sequenced as defined in Yuan *et al*.[26] and sequences are compared to a reference database (http://pubmlst.org/xfastidiosa/) to assign the strain to subspecies or detect recombinants (e.g.[11,14,27–29]). Large scale and unbiased survey of the disease requires exhaustive sampling of plants (both symptomatic and asymptomatic) in multiple habitats, which is fastidious. Furthermore, the heterogeneous distribution of the bacterium in the plant[3] as well as PCR inhibitors (e.g. polyphenols[30]) may induce false-negative results. To the contrary, most insect vectors can be easily sampled through sweeping (among vectors only cicadas are relatively difficult to sample and may require acoustic tools to locate them). Insects are also known to contain PCR inhibitors ([31,32]; R. Krügner USDA USA *pers. comm.*) but colonization of insects by *Xf* occurs in a non-circulative manner with bacterial colonies located in the foregut (Purcell & Finlay[33]; observations on *Graphocephala atropunctata* (Signoret, 1854)). Thus, it is either possible to dissect the foregut of the insects or extract DNA from an entire specimen to make sure having access to the bacterium. Therefore, as suggested by the "spy insect" approach set up in buffer zones and symptom-less areas closed to contaminated olive groves in Italy[34–36], implementing massive surveys to test whether or not insect populations in different ecosystems do carry *Xf* would efficiently complement studies on plants and improve the early detection and monitoring of the disease.

Here we propose to go one step further on this idea and provide a first assessment of the use of insects to detect, monitor or predict the distribution of *Xf* in Europe, using Corsica as a case study. In a first step, we propose to test the feasibility of a large screening of insect populations for the presence of *Xf* but also for the possible characterisation of the carried strains via PCR amplification and sequencing of the loci included in the MLST of *Xf*. We sampled 62 populations of *Philaenus spumarius* throughout Corsica (Fig. 1, Table S1) from early June (when there was still a mix of larvae and adults, Fig. S1) to late October (before the adults are presumably killed by winter). We then tested for the presence of *Xf* in a subset of 11 populations (448 specimens, Fig. 1, Table S1) using a qPCR approach and a nested PCR protocol designed for the purpose of the study. Indeed, targeting *Xf* using qPCR appeared inconclusive and did not allow assessing the genetic identity of the strains. In a second step, we compared the results of our molecular tests to the potential range of *Xf* as estimated using species distribution modelling based on presence/absence tests conducted on plants. In a third step, we collected occurrence data of *P. spumarius* throughout Europe and estimated its geographical range using species distribution modelling to discuss the interest of using this insect species as a sentinel for the early detection and monitoring of *Xf* in Europe.

Materials and Methods

Sampling.   We sampled adults of *Philaenus spumarius* in 62 localities in Corsica from early June to late October (Fig. 1, Table S1) by passing a sweep net through the vegetation using alternate backhand and forehand strokes. Specimens were collected in the net with an aspirator, killed on site with Ethyl Acetate and stored in 8 mL vials containing 70% EtOH. Vials were stored in a freezer (−20 °C) until DNA extraction. We mostly sampled in natural environments. The distance between sampling areas and the closest infested area (as identified by the national survey on plant material) ranged from ca. 20 m to 30000 m (Fig. 1). To optimize our sampling and test the feasibility of a large survey, we spent no more than 30 min sweeping in each locality. Specimens and plants on which they were successfully collected were identified to species. Molecular tests were conducted on a subset of 11 populations of *P. spumarius* (32 specimens per population, Fig. 1, Table S1). Three of these populations were sampled both in early June and late October to test for a possible seasonal variation of the prevalence of *Xf*. Other populations were sampled either in June (early/late) or October. A total of 448 specimens were screened for *Xf*.

DNA extraction.   It is infeasible to proceed to the dissection of the foregut of each specimen for a mass-survey. Moreover, dissection of insects can generate cross-contamination. Thus, we tried to improve each steps of the DNA extraction protocols classically used[37,38] to reduce the impact of PCR inhibitors that may be contained in the insects[31,32] and increase yield in bacterial DNA. The complete protocol is available in Supplementary data (Appendix 1). Briefly, insects were placed in lysis buffer that contained PVP (Polyvinylpyrrolidone, to absorb polyphenols and polyamins thus preventing them from interacting with DNA which could inhibit PCR) and Sodium Bisulfite (to prevent oxydation of polyphenols, that, when oxidized covalently bind to DNA making it useless for further application). Insects were crushed using garnet crystals and ceramic beads coated with zirconium. Then, lysozyme was added to facilitate lysis of the bacteria. After 30 min incubation, Proteinase K and extraction buffer that contained guanidium chlorure (to denature proteins and increase lysis of bacterial cells) and sodium bisulfite (antioxydant) were added to the mix. After one-hour incubation, deproteneisation using potassium acetate was performed. Finally, DNA extracts were purified using a KingFisher robot and Chemagic beads.
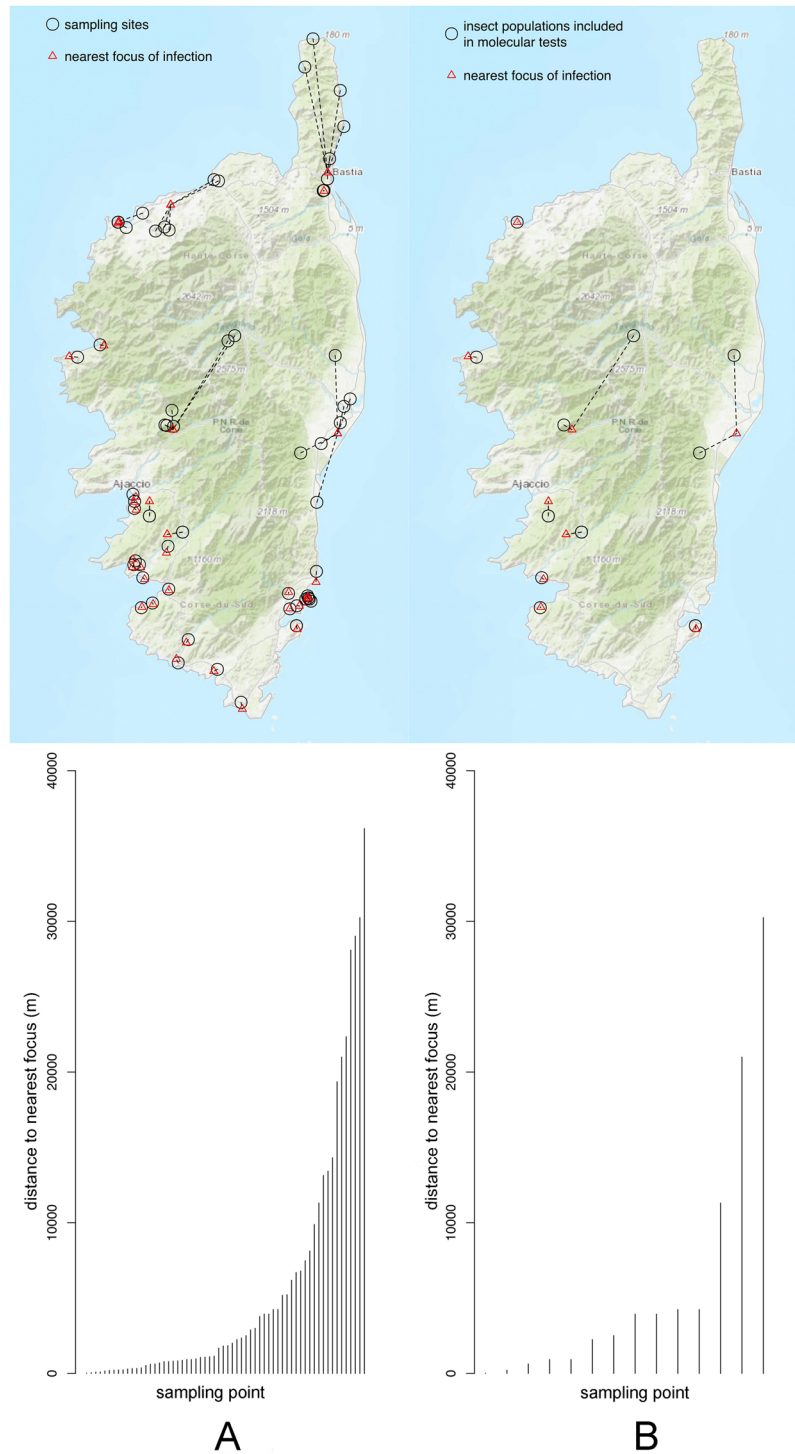
Figure 1. Sampling sites and distance to the nearest focus of infection determined by molecular tests on plant material (national survey). (**A**) = All sampling sites, (**B**) = sampling sites where insect populations were tested for the presence of *Xf*. See Table S1 for more information on sampling sites. The maps show the sampling sites and the foci of infection. Graphs are plots of distance between each sampling site and the nearest focus of infection. Maps were created using the R package OpenStreetMap[75] with data copyright to OpenStreetMap contributors.

Quantitative PCR (qPCR). We used the method proposed by Harper *et al.*[24], which is listed as one of the official detection methods of *Xf* in plant material[27] and recognized as the most sensitive method available to date for the detection of *Xf* in plants[24,25]. We followed recommendations by the ANSES[39] and the EPPO[27] but re-evaluated a cycle threshold using negative controls (ultrapure water and 2 μg of phage lambda purified DNA)

to better fit with our experimental conditions. To estimate the sensitivity of the qPCR approach, we used incremental dilution of an inactivated bacterial suspension of known concentration provided by B. Legendre (LSV, ANSES, Angers, France). Two replicates of qPCR were performed on each insect specimen.

Nested PCR and Sequencing. Our attempt to amplify the seven loci included in the MLST of *Xf* using conventional PCR and primers/conditions described in the original protocol[26], https://pubmlst.org/xfastidiosa/) were unsuccessful, probably because of the low amount of bacteria. We thus switched to a nested PCR approach. Sequences of the different alleles of each locus were downloaded from, https://pubmlst.org/xfastidiosa/ (last access October 19th 2017) and aligned. Internal primers for each locus were designed from these alignments (Table S2) and primers were M13 tailed to simplify the sequencing reaction. We ensured that the nested PCR approach did not preclude discrimination among genetic entities by comparing maximum likelihood phylogenetic trees obtained from a concatenation of the 7 loci originally included in the MLST of *Xf* and their reduced sequences as included in the nested PCR scheme (Fig. S2). Loci extracted from all genomes available on Genbank (last access October 19th 2017) were used as input. To test for the presence of *Xf* in the insects, we first targeted *holC*. When the amplification of *holC* was successful, a nested PCR to amplify the six other loci was attempted. *HolC* was first amplified using the primers listed in Yuan *et al.*[26] and the mastermix and PCR conditions described in Tables S3 and S5. Five microliters of PCR product were then used to perform a nested PCR with the mastermix and PCR conditions described in Tables S3 and S5. For the six other loci, we first performed a triplex PCR (gltT/leuA/petC and cysG/malF/nuoL) using the primers listed in Yuan *et al.*[26] and the mastermix and PCR conditions described in Tables S4 and S5. Five microliters of the PCR product were then used to perform a simplex nested PCR with the mastermix and PCR conditions described in Tables S4 and S5. The strict procedure implemented to avoid carry-over contamination is detailed in the Appendix 2 of the supplementary data file. To estimate the sensitivity of the nested PCR approach, we used incremental dilution of the same inactivated bacterial suspension as for qPCR. Sequencing of the PCR products was performed at AGAP on an Applied Biosystems 3500 Genetic Analyser. Allele assignation was performed using, http://pubmlst.org/xfastidiosa/. Phylogenetic inferences were performed using raxmlHPC-PTHREADS-AVX[40]. Given that α and the proportion of invariable sites cannot be optimized independently from each other[41] and following Stamatakis' personal recommendations (RAxML manual), a GTR $+ \Gamma$ model was applied to each gene region. We used a discrete gamma approximation[16] with four categories. GTRCAT approximation of models was used for ML boostrapping[42] (1000 replicates). Resulting trees were visualised using Figtree[43]. SplitsTree v.4.14.4[44] was used to build NeighborNet phylogenetic networks.

Species distribution modelling of *Xf* at the Corsican scale. The potential distribution of *Xf* subsp. *multiplex* (ST6 & ST7) in Corsica was modelled using BIOCLIM[45,46] and DOMAIN[47]. Methodology followed Godefroid *et al.*[48]. Data collected in France from 2015 to 2017 by the national survey on plant material were used as input. Three different climatic datasets were used to fit the models. The results consisted in a map depicting the percentage of models predicting the presence of *Xf* subsp. *multiplex* (ST6 & ST7).

Species distribution modelling of *Philaenus spumarius* at the European scale. *Occurrence dataset.* A total of 1323 occurrences were used to model the distribution of *P. spumarius* in Europe. Off these, 471 originated from the GBIF database (GBIF.org (2017), *GBIF Home Page*. Available from: http://gbif.org [1rd November 2017]). The remaining 852 occurrences corresponded to our own observation records or were taken from the literature (List of references in Supplementary material, Appendix 3).

Modelling framework. We modelled the distribution of *P. spumarius* by means of Maxent, the most widely used species distribution model[49,50]. Maxent models the potential species distribution based on the principle of maximum entropy. It relates species occurrence records and background data with environmental descriptors to get insight into the environmental conditions that best reflect ecological requirements of the species. Species responses to environmental constraints are often complex, which implies using nonlinear functions[49]. For that reason the parametrization of Maxent involves choosing among several transformations (referred to as feature classes or FCs) of original environmental descriptors i.e. linear, quadratic, product, hinge and threshold[51]. The parametrization of Maxent also involves a regularization multiplier (RM) introduced to reduce overfitting[52]. Various authors have highlighted that the default settings of Maxent are not optimal in all situations and might lead to poorly performing models in certain cases[52–54] and it is sensible to search for the best parameters given the dataset at hand[53].

We built models with RM values ranging from 0.5 to 4 with increments of 0.5 and 6 FC combinations (L, LQ, H, LQH, LQHP, LQHPT with L = linear, Q = quadratic, H = hinge, P = product and T = threshold) using the R package ENMeval[55]. This led to 48 different models. The models were fitted using all the available occurrence points and 10,000 randomly positioned background points. The optimal settings corresponded to the models giving the minimum AICc values (see[55] for details). The resulting optimal FC and RM values were used to fit the final Maxent model based on a training dataset constituted by a random subset of 80% of the occurrences and 10,000 randomly positioned background points. The resulting Maxent model was then used to create a map of suitability scores (i.e. habitat suitability) across Western Europe (logistic output of Maxent[51]). Suitability values were transformed into presence/absence by applying two thresholds a) the value at which the sum of the sensitivity (true positive rate) and specificity (true negative rate) is the highest and b) the lowest presence threshold (LPT)[56]. These analyses were performed using the R package dismo[57].

We evaluated the performance of the model using the AUC metric[58] and the Boyce index[59]. The Boyce index is a reliable presence-only evaluation measure that varies between −1 and +1. Positive values indicate a model which predictions are consistent with the distribution of the presences in the evaluation dataset. The Boyce index was calculated using the R package ecospat[60].
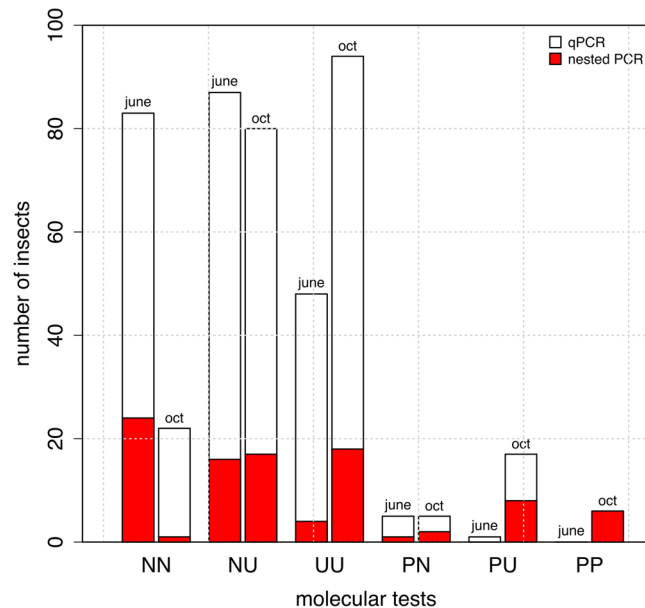
Figure 2. Results of the molecular tests performed on the insects (qPCR and nested PCR). The white histogram shows the distribution of the results of the qPCR replicates for the 448 insects. N = negative, P = positive, U = undetermined (see text for the Ct). Thus, NN indicates that the two replicates were negative. The red histogram indicates that positive nested PCR on *holC* were obtained. The figure was created with R[76].

Environmental descriptors (bioclimatic variables). Our modelling strategy relies on a set of bioclimatic descriptors hosted in the Worldclim database[61]. Each variable is available in the form of a raster map and represent the average climate conditions for the period 1970–2000. We used raster layers of 2.5-minute spatial resolution, which corresponds to about 4.5 km at the equator. The choice of environmental descriptors to be involved is crucial and it is widely acknowledged that using reduced number of variables improves transferability and avoids model overparametrization[62,63]. We used the variables referred to as bio5, bio7 and bio19 corresponding to the maximum temperature of the warmest month, the temperature annual range and the precipitation of the coldest quarter see details in[64]. Both temperature and precipitation were considered as proxy for the main environmental features constraining the insect distribution.

Results

Detection of *Xf* in insect vectors using qPCR.
Based on the results obtained with the negative controls, we fixed the cycle threshold (Ct) to 32.5. Thus, results were considered positive when Ct < 32.5 and an exponential amplification curve was observed, results were considered negative when Ct > 36 and results were considered undetermined when 32.5 < Ct < 36. Sensitivity tests on the inactivated bacterial suspension indicated that the signal was lost when i) less than 100 bacteria were present in the reaction mix or ii) less than 250 bacteria mixed with 2 µg of insect DNA were present in the reaction mix. The results of the two qPCR replicates were different for 43.8% of the insects (Fig. 2). A single positive qPCR was obtained for 2.7% of the insects collected in June. Four of the seven populations (sites A, B, D, J, Fig. 1) contained positive insects (from one to two positive insects per population). A single positive qPCR was obtained for 12.5% of the insects collected in October. Six of the seven populations (sites A, D, C, F, G, I) contained positive insects (from 3 to 7 positive insects per population). The two qPCR replicates were positive for 1.3% of the insects all collected in October from two populations (sites I: 2 positive insects; and G: 8 positive insects).

Detection of *Xf* in insect vectors using nested PCR and sequencing.
Sensitivity tests on the inactivated bacterial suspension indicated that the signal was lost when i) less than 5 bacteria were present in the reaction mix or ii) less than 50 bacteria mixed with 2 µg of insect DNA were present in the reaction mix. As compared with the qPCR approach, the nested PCR approach on *holC* revealed that *Xf* was present in all populations both in June and October with higher prevalence rates (Figs 2 and 3). Positive nested PCR were always obtained when the results of the two qPCR replicates were positive, i.e. presumably from the insects with the highest bacterial load (collected in late October). However, positive nested PCR were also obtained when the results of the two qPCR replicates were negative (5.6% of the samples). The rate of false negative as compared to the nested PCR approach was 8.7% for the first replicate of qPCR and 10.5% for the second replicate. Notably, 11.2% (resp. 7.1%) of the qPCR that gave an undetermined result led to a positive nested PCR for the first replicate of qPCR (resp. the second replicate of qPCR). With the nested PCR approach, an average of 20.1% (23.2%) of the specimens were found positive to *Xf* in June (October). The prevalence of *Xf* in the different populations varied from 0.0% to 43.7% in June and 12.5–34.4% in October. No significant seasonal variation of the presence of *Xf* was observed. Analysis of the sequences obtained for *holC* revealed that 56.7% of the insects tested positive for *Xf* carried allele *holC_3*, and
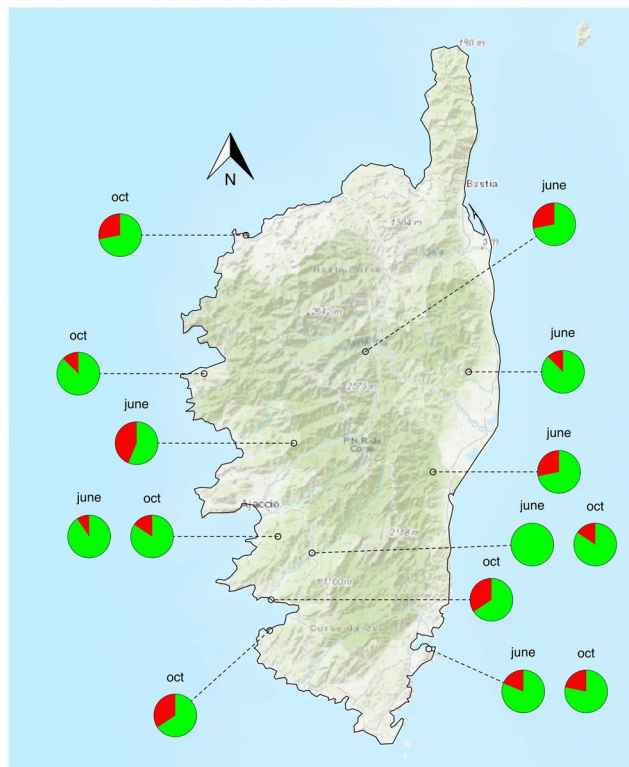
Figure 3. Prevalence of *Xf* in the different populations as revealed by the nested PCR approach targeting *holC*. Red (green) color indicates the percentage of insects tested positive (negative) for the presence of *Xf*. Tests were conducted on 32 specimens in each sample site. The map was created using the R package OpenStreetMap[75] with data copyright to OpenStreetMap contributors.

	holC_3	holC_1
Site A (June)	1	1
Site A (October)	2	2
Site C (June)	5 + 1 variant	0
Site C (October)	5	1
Site D (June)	0	0
Site D (October)	3 + 1 variant	0
Site B	3	0
Site E	1	7
Site J	4	3
Site K	6 + 1 variant	2
Site F	10	1
Site G	9	1
Site H	0	1 + 2 variants
Site I	6	2

Table 1. *HolC* alleles present in each population of *P. spumarius*.

21.6% carried allele *holC_1* (Table 1, Fig. 4a). It is noteworthy that two yet undescribed variants of *holC_3* as well as two yet undescribed variants of *holC_1* were found in the screened populations. This result is interesting *per se* and indicates that the probability that our results are due to carry-over contamination is reduced. Interestingly for a few specimens (6% of the positive samples), double peaks were observed on the diagnostic sites for allele *holC_3* versus allele *holC_1*, which suggest that they may carry two subspecies of *Xf*.

Sequences for the seven loci of the MLST have been obtained for the specimens with the highest bacterial load (2 specimens collected in site G in October). The complete typing indicates that the carried strain was *Xf multiplex* ST7. Only partial typing (at most 2 loci) could be obtained for other specimens. Thus, we could not conclude without doubt on the identity of the strains they carried. For two specimens that carried *holC_1*, one collected in late June in site K and one collected in October in site E, sequencing of *gltT* indicates that the carried allele was
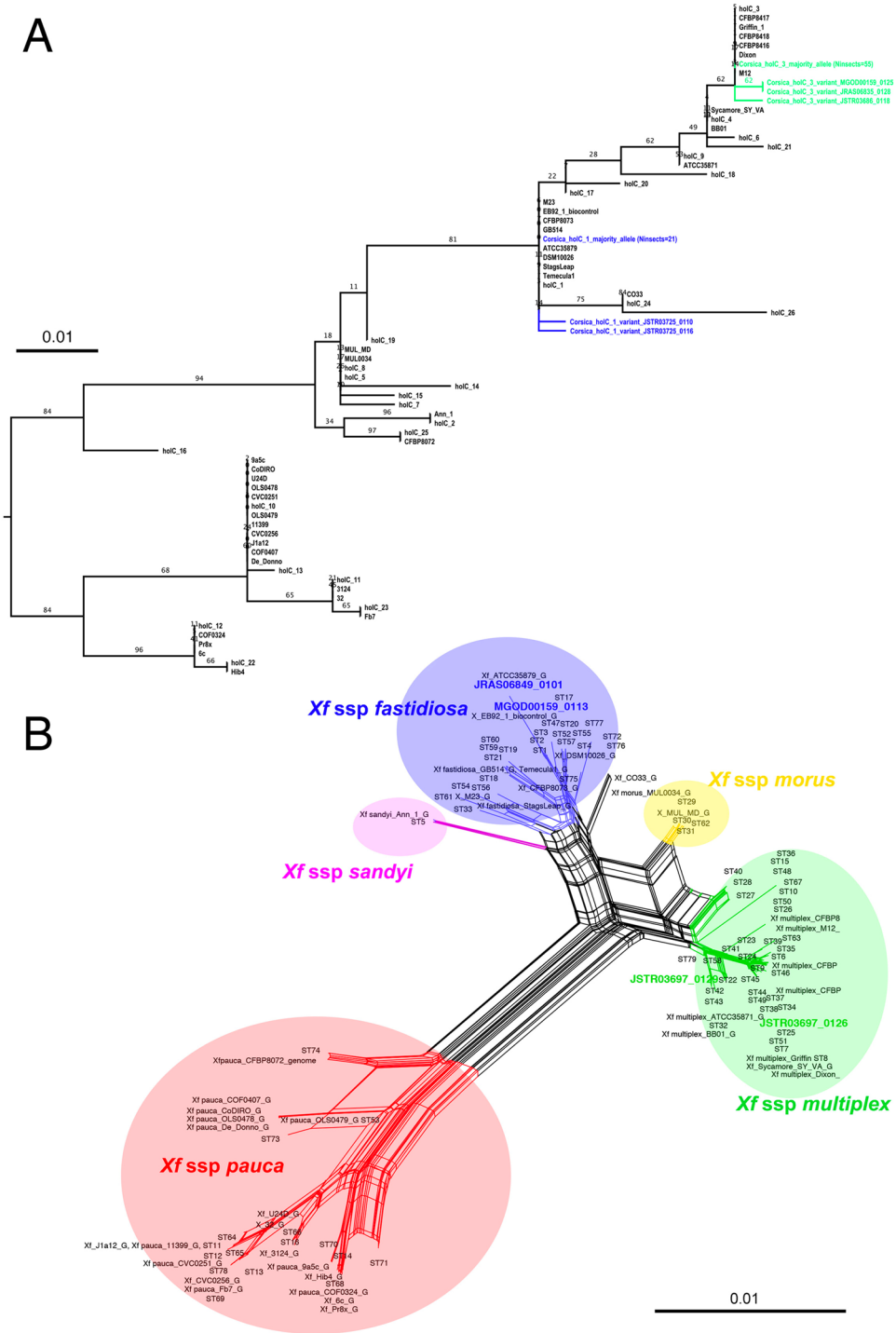
Figure 4. Position of the strains characterized in populations of insects in Corsica. (**A**) RAxML tree inferred from the analysis of the reduced *holC* sequences targeted with the nested PCR approach (BP at nodes, 1000 replicates). Alleles present in, https://pubmlst.org/xfastidiosa/ (last access October 19th 2017), genomes available on Genbank (last access October 19th 2017) and alleles obtained from insect samples are included in the analysis. The tree was exported from Figtree[43] and modified in Adobe Photoshop CS4. (**B**) NeighborNet phylogenetic network. Inferences are based on a concatenation of the reduced *leuA*, *petC*, *malF*, *cysG*, *holC*, *nuoL*, *gltT* targeted with the nested PCR approach. STs present in, https://pubmlst.org/xfastidiosa/ as well as genomes available on Genbank are included in the analysis (last access October 19th 2017). Insect samples for which at least two loci could be sequenced (JSTR03697_0126 & JSTR03697_0129: 7 loci sequenced and MGOD00159_0113 & JRAS06849_0101: *holC* and *gltT* sequenced) are included in the network. Note: CO33/ST72[13,77]; ST76[13]; CFBP8073/ST75[14] 14 as well as ST79[11] are genetically related to isolates belonging to different subspecies of *Xf*. The network was built with SplitsTree v.4.14.4[44] and modified with Adobe Photoshop CS4.
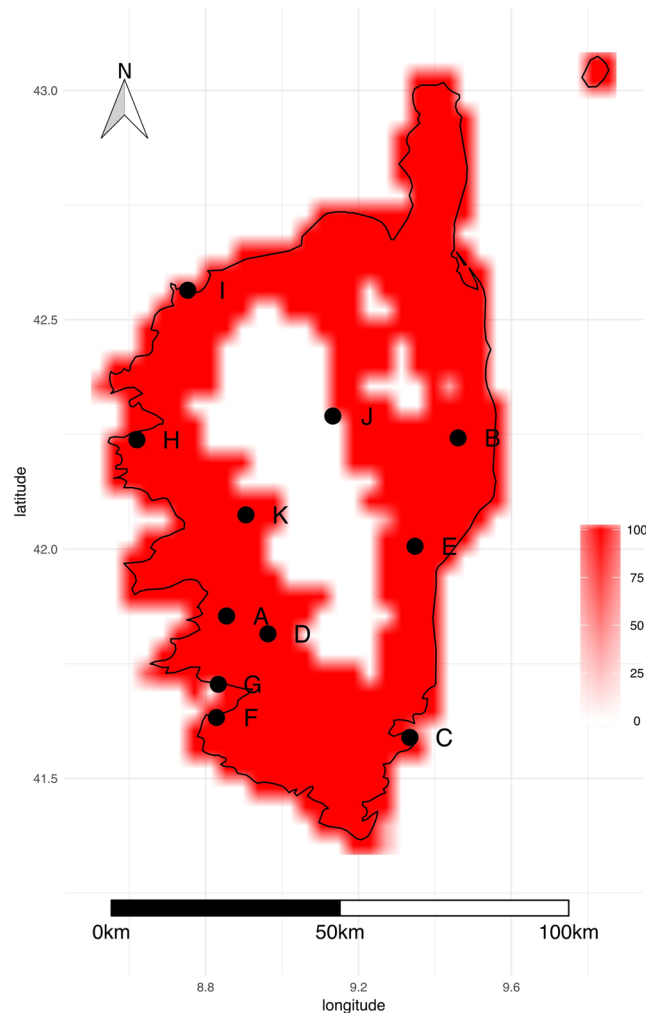
Figure 5. Potential geographical distribution of *Xf multiplex* (ST6/ST7) in Corsica. DOMAIN and Bioclim models were fitted with different climate datasets and used to predict the distribution of the bacterium. The map depicts the percentage of predictions indicating the presence of the bacterium. Black dots indicate insect populations tested for the presence of *Xf*. The figure was created with R[76].

gltT_1 and a variant of *gltT_1* respectively, which suggests that the subspecies *Xf fastidiosa* may be also present in Corsica, though this result needs to be confirmed by a complete typing. The NeighborNet network inferred from the concatenation of the reduced sequences of *leuA*, *petC*, *malF*, *cysG*, *holC*, *nuoL*, *gltT* targeted by the nested PCR approach and including reported STs, available genomes as well as Corsican strains characterized on more than one locus is presented in Fig. 4b. All sequences have been deposited on Genbank (accession numbers: MH628341-MH628361).

Match between the molecular results and the predicted distribution of *Xf multiplex* ST6 & ST7.
All sampling sites fall within the predicted distribution of the bacterium, which encompasses the entire island apart from mountainous regions in the centre (Fig. 5). Two sampling sites (stations J & K) fall near the edge of the predicted potential distribution area of *Xf*.

Distribution of Philaenus spumarius *at the European scale.*
Occurrence data used in the study are presented in Fig. 6a. The regularization multiplier of the Maxent model giving the minimum AICc values was 1.5 and it was associated to feature class combining linear, quadratic, hinge, product and threshold features (LQHPT). The value of the AUC of the Maxent model fitted using the latter optimal settings was 0.89 and the Boyce index was 0.986. Both metrics indicated that the Maxent model performed satisfactorily. Figure 6b displays the prediction of the Maxent model (logistic output) i.e. the habitat suitability for *P. spumarius* across the study area. Most of the Western Europe appeared to be associated to high habitat suitability values. The presence/absence maps derived from the conversion of habitat suitability using the threshold value maximizing the sensitivity and the specificity or the lowest presence threshold[56] are given in Fig. S3. In both cases nearly all the Western Europe appeared to host *P. spumarius*.
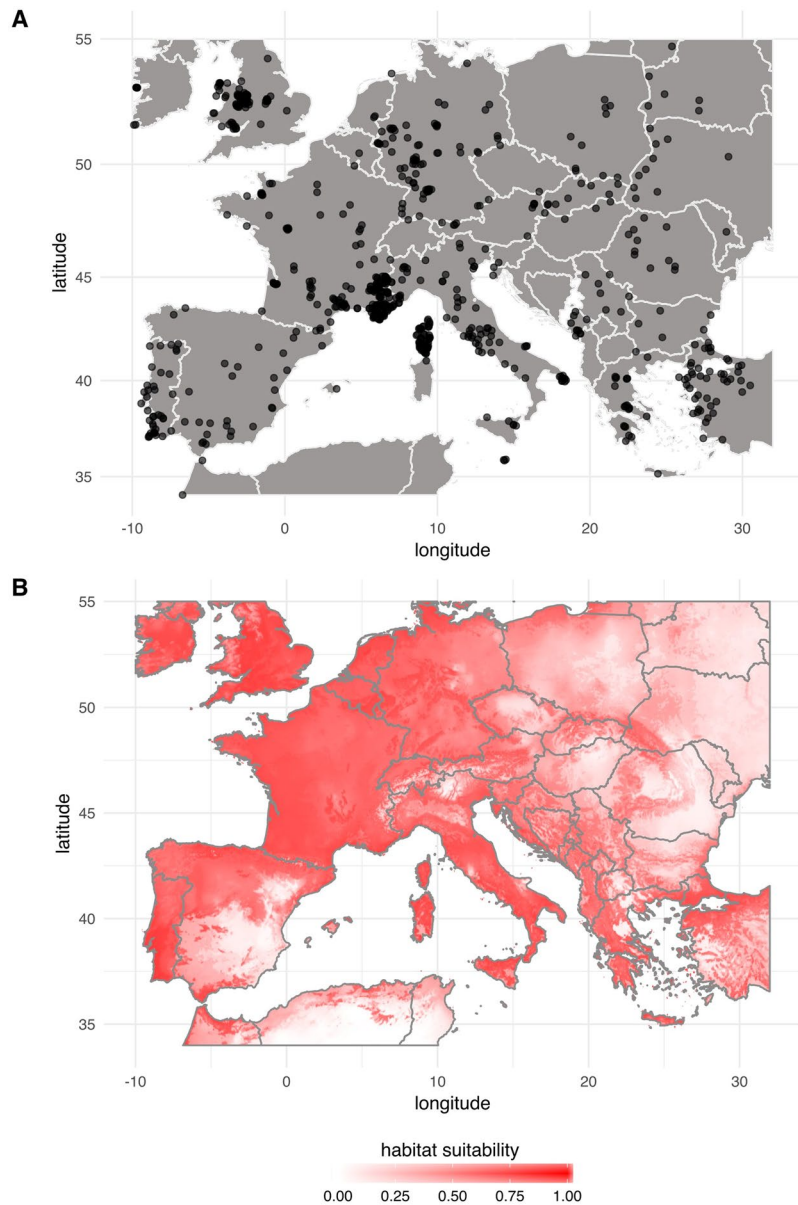
Figure 6. Occurrences and predicted distribution of *P. spumarius* in Europe. (**A**) Plot of the occurrences recorded from GBIF, the literature and our own observations (**B**). Habitat suitability corresponding to the logistic output of the Maxent model (from white/low to red/high habitat suitability). The figure was created with R[76].

Discussion

The idea of using insect as "spies" for the early detection of *Xf* in buffer zones and symptom-less areas is not new[34-36]. Here, we tried to go one step further than what is done on olive groves in Southern Italy and propose to test whether insects could be used to detect, monitor or predict the distribution of *Xf*. We also propose to make a large-scale preliminary identification of the subspecies/strains of *Xf* that may be present in the ecosystem. If needed, whole genome sequencing could refine the identification.

Importantly, our study reveals some limitation of the qPCR approach *sensu* Harper *et al.*[24] to detect *Xf* in the insect vectors. The approach appears not sensitive enough to detect low bacterial load, which questions its use for the early detection of the bacterium in insects. Indeed, if we consider a result as positive only when two replicates of qPCR are positive, *Xf* will be considered as present in only two of the eleven populations and only in October. When a more sensitive approach was used, no significant differences in the proportion of insects carrying *Xf* in June and in October was observed. Furthermore, the rate of undetermined results obtained with the qPCR approach is high (73.0% of the replicates with at least one undetermined result), which is unsatisfactory when it comes to the detection of plant pathogens. Although we have not formally tested a loop-mediated isothermal amplification (LAMP) approach for the detection of *Xf* in the insects (an approach increasingly used on the

field), our study indicates that the results obtained with this approach should be interpreted with caution. Indeed, LAMP has been shown to be less sensitive than qPCR[24,25].

This result suggests that the lower prevalence of *Xf* in *P. spumarius* observed in early season (winter-spring, 12.6%)[34] as compared to late season (October-December, 40%)[65] in an olive grove in Italy may be artefactual. This difference may be due to the relatively poor ability of PCR and LAMP to detect *Xf* in insects with the lowest *Xf* load (more frequent in June). Consequently, we strongly advocate the use of highly sensitive methods to monitor *Xf* within insects, especially in *Xf*-free areas to avoid false negative results.

The nested PCR approach targeting *holC* optimized for the purpose of this study appeared much more sensitive than the qPCR approach and allows a first assessment of the diversity of the strains present in the environment. With this approach, all insect populations appeared to carry *Xf*, which shows that the bacterium is widely distributed in Corsica, not only in the area where *Xf* is supposed to have been recently introduced. The sampling sites of the 11 populations of *P. spumarius* tested positive for the presence of *Xf* all fall within the predicted potential distribution of the bacterium, which validates the plausibility of our nested PCR results and shows that molecular tests on insects could be used for risk assessment. It is noteworthy that while they were not visible when we performed sample collection in 2016, leaf scorch symptoms could be clearly observed in all localities tested for the presence of *Xf* when we went back to the field in October 2017. However, weather data indicated that the summer of 2017 has been the driest in 15 years and it is acknowledged that symptoms due to *Xf* are not easy to differentiate from drought symptoms[3]. Thus, observed symptoms may be due to summer drought itself. However, the possibility that at least part of these symptoms are due to *Xf* cannot be ruled out as i) all populations of *P. spumarius* were tested positive for the presence of the bacterium, ii) plants were found positive to *Xf* close to certain prospected sites (Fig. 1), iii) all sites were predicted as favourable for the bacterium (Fig. 5). The mechanisms underlying the interaction of water stress and infection by/sensibility to *Xf* and a possible causal relationship between these two parameters is a constant area of research[66–68]. It is difficult to assess whether or not the severe drought may have favoured the spread of *Xf* or revealed its presence. Regarding the vectors, studies conducted in the US on *Homalodisca vitripennis* Germar, 1821 have shown that the insects will take longer meal and feed more frequently on fully irrigated plants, both events that favour the acquisition and transmission of *Xf*[69]. This led to the conclusion that even low levels of water stress may reduce the spread of *Xf* by *H. vitripennis*. However, nothing is known about the feeding behaviour of *P. spumarius* or other European insect vectors under severe drought conditions. One can hypothesize that probing behaviour may vary with more switch from one plant to another as xylem fluid tension is reduced in all plant species. However, our sampling campaigns show that *P. spumarius* may rarely switch to woody plants. Furthermore, the spittlebug is subject to aestivation. Consequently, the role of other potential vectors in the spread of the disease should be investigated in the future (especially cicadas).

The wide distribution of two subspecies of *Xf* in Corsica highlighted by our molecular tests suggests that the introductions of the bacterium to Corsica may be ancient and multiple. Indeed, it appears unlikely that the bacterium spread into insect populations all over Corsica in such a short time lapse since the first detection (less than 2 years). Another argument in favour of an ancient/multiple introduction of *Xf* to Corsica is the presence of several STs, including variants either highlighted on plants[11] and/or on insects (this study) and co-occurrence of strains/subspecies in the same matrix (plant/insect). Thus, *Polygala myrtifolia*, on which *Xf multiplex* was detected during the summer 2015, might not have been a key actor in the spread of *Xf*. This detection might just have served as a trigger for large-scale surveillance and studies that now reveal a much more complex situation than expected. Notably, the co-occurrence of subspecies/strains in the same host plant raises doubt about which entity produced symptoms and therefore on subspecies/strain occurrences used for risk assessment. Furthermore, as co-occurrence of subspecies/strains in the same host insect or plant may favour recombination, and, as a consequence enlarge host range, disease management may be further complicated[70–72].

Our results indicate that the number of bacterial cells in the cibarium of *P. spumarius* may be low, even in the late season, which complicates molecular detection. Thus, our results may still underestimate the prevalence of *Xf* in insect populations. This low number of bacteria makes possible the PCR amplification and sequencing of all the loci included in the MLST of *Xf* only on the insects in which the bacterial load is the highest. Progress should be made to circumvent this issue and hybridization-based capture of genomic regions or other approaches that are more sensitive than PCR and nested PCR may be soon implemented. *Xf multiplex* was recovered both in plants and in insects. However, we highlighted rare or widespread subspecies/variants not yet detected in plants. To the contrary, we did not detect *Xf pauca* ST53 in the targeted insect populations. These results may be explained by the followings: i) *Xf pauca* or our rare variants may be restricted to some areas where insect/plant prospection was not yet conducted, ii) some strains are harder to detect in plants (competition between PCR primers or differences in development). With survey effort, results on plants and insects should become consistent.

Obviously, we did not aim to study the entire community of vectors of *Xf* occurring in Corsica. Our goal was to gather huge populations of the same species of vector in each locality to perform molecular tests. Doing this, *P. spumarius*, which is, by the way, the only efficient vector known in Europe so far[22,23], was identified as the perfect candidate. While we did not aim at studying the exact phenology and host preferences of *P. spumarius*, we made a couple of biological observations that will need to be formally tested but seem nevertheless relevant to propose a few immediate solutions to decrease its potential impact on cultivated plants.

Interestingly, specimens of *P. spumarius* were easily and almost exclusively collected from *Cistus monspeliensis* with only a few specimens collected in grasses and clover. In 2016, adults started to emerge in early June, were impossible to collect in summer as mentioned by Chauvel et al.[10] and huge adult populations reappeared in early October with the first rains. No survey was conducted in winter. We are thus unable to state whether or not adults of *P. spumarius* may survive winter in some areas. The mix of larvae and tenerals observed in many localities in early June suggest they may not survive winter but this needs to be confirmed especially in a context of global warming. Our observations contrast with observations made in Apulian olive groves (Southern Italy) by Ben Moussa et al.[73] where *P. spumarius* begins to moult in April-May, is abundant in summer and seems more

polyphagous as it moves from herbaceous plants to olive trees. It is noteworthy that *P. spumarius* is also polyphagous in Southern France as larvae were found on *ca* 120 different species of plants in the area of Montpellier (JCS, *pers. obs.*). Therefore, the situation in Corsica appears different from what has been observed elsewhere in Europe. While host preferences, exact phenology and tolerance to temperature variation of the Corsican populations of *P. spumarius* require to be precisely assessed in the future, performing a genetic analysis to evaluate their status appears also relevant. On a more general note, the contrasted observations on *P. spumarius* suggest that strategies that would need to be set up to monitor the spread of the disease would differ from areas to areas. Our field samplings and observations of huge numbers of spitlike foam on *Cistus monspeliensis* in the spring suggest that this widespread plant may play a critical role in the spread of the bacterium in Corsica. *Xf* was detected in young and older adults of *P. spumarius* which suggests that the insects could acquire the bacterium from *C. monspeliensis*, which may act as reservoir for the next season. Thus, *C. monspeliensis*, which, from our observation, seems mostly asymptomatic to *Xf* (but was already tested positive to *Xf* on plants[11]), could have favoured an initially invisible spread of the disease throughout Corsica. *Cistus monspeliensis* may thus have played the role of the hidden compartment suggested by one of the model selected by Soubeyrand *et al*.[74] to best explain current observations on plants. If a key role of *P. spumarius* or *C. monspeliensis* is confirmed by further studies, disease management in Corsica may be trickier in natural ecosystems than in agro-ecosystems. Indeed, *Cistus* spp. are largely distributed in anthropized and open semi-natural habitats. They are major components of the spontaneous natural reforesting that generally follows abandonment of agriculture and grazing and are among the first colonizers after a fire. While it seems feasible to remove *Cistus* spp. from the vicinity of crops, other strategies should be implemented to control the spread of *Xf* in natural environments.

Interestingly, the species distribution model of *P. spumarius* at the European scale indicates that it may be the perfect sentinel to detect the presence of *Xf* and make a preliminary assessment of the subspecies/strains present in the environment. As a conclusion, we suggest that a study of this type in Europe will provide a better picture of the distribution of the bacterium and set up a global strategy to control it. There is an urgent need to take stock of the current situation with a large-scale, blind survey and using effective/sensitive enough molecular tools. This may allow finding out why the current epidemic appears so recent, understand what could be the triggers, better design management strategies, and avoid the unnecessary economic pressure on certain geographical areas and agricultural sectors. It is even more urgent as global warming may favour the (re)-emergence of *Xf* and predictions at the European scale suggest that *Xf* may be more widespread that what is currently thought[48].

References

1. Almeida, R. P. & Nunney, L. How do plant diseases caused by *Xylella fastidiosa* emerge? *Plant Dis.* **99**, 1457–1467 (2015).
2. Retchless, A. C. *et al*. In *Genomics of Plant-Associated Bacteria (*ed. Springer Berlin Heidelberg) 177–202 (2014).
3. EFSA. Panel on Plant Health (PLH), European Food Safety Authority (EFSA), Scientific Opinion on the risk to plant health posed by *Xylella fastidiosa* in the EU territory, with the identification and evaluation of risk reduction, EFSA Journal 2015;13(1):3989 (2015).
4. Nunney, L. *et al*. Recent evolutionary radiation and host plant specialization in the *Xylella fastidiosa* subspecies native to the United States. *Appl. Environ. Microbiol.* **79**, 2189–2200 (2013).
5. Chatterjee, S., Almeida, R. P. P. & Lindow, S. Living in two worlds: the plant and insect lifestyles of *Xylella Fastidiosa*. *Annu. Rev. Phytopathol.* **46**, 243–271 (2008).
6. Redak, R. A. *et al*. The biology of Xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology. *Annu. Rev. Entomol.* **49**, 243–270 (2004).
7. Purcell, A. H. Paradigms: examples from the bacterium *Xylella fastidiosa*. *Annu. Rev. Phytopathol.* **51**, 339–356 (2013).
8. Janse, J. & Obradovic, A. *Xylella fastidiosa*: its biology, diagnosis, control and risks. *J. Plant Pathol.* S35–S48 (2010).
9. Saponari, M., Boscia, D., Nigro, F. & Martelli, G. P. Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (Southern Italy). *J. Plant Pathol.* **95**, 668 (2013).
10. Chauvel, G., Cruaud, A., Legendre, B., Germain, J.-F. & Rasplus, J.-Y. Rapport de mission d'expertise sur *Xylella fastidiosa* en Corse. Available at, http://agriculture.gouv.fr/sites/minagri/files/20150908_rapport_mission_corse_xylella_31082015b.pdf (2015).
11. Denancé, N. *et al*. Several subspecies and sequence types are associated with the emergence of *Xylella fastidiosa* in natural settings in France. *Plant Pathology* **66**, 1054–1064 (2017).
12. Olmo, D. *et al*. First detection of *Xylella fastidiosa* infecting cherry (*Prunus avium*) and *Polygala myrtifolia* plants, in Mallorca Island, Spain. *Plant Dis.* **101**, 1820–1820 (2017).
13. Loconsole, G. *et al*. Intercepted isolates of *Xylella fastidiosa* in Europe reveal novel genetic diversity. *Eur. J. Plant Pathol.* **146**, 85 (2016).
14. Jacques, M.-A. *et al*. New coffee-infecting *Xylella fastidiosa* variants derived via homologous recombination. *Appl. Environ. Microbiol.* **82**, 1556–1568 (2016).
15. Dellapé, G., Paradell, S., Semorile, L. & Delfederico, L. Potential vectors of *Xylella fastidiosa*: a study of leafhoppers and treehoppers in citrus agroecosystems affected by Citrus Variegated Chlorosis. *Entomologia Experimentalis et Applicata* **161**, 92–103 (2016).
16. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
17. Severin, H. H. P. Transmission of the virus of Pierce's disease by leafhoppers. *Hilgardia* **19**, 190–202 (1949).
18. Severin, H. H. P. Spittle-insect vectors of Pierce's disease virus. II. *Life history and virus transmission. Hilgardia* **19**, 357–382 (1950).
19. Krell, R. K., Boyd, E. A., Nay, J. E., Park, Y. L. & Perring, T. M. Mechanical and insect transmission of *Xylella fastidiosa* to *Vitis vinifera*. *Am. J. Enol. Vitic.* **58**, 211–216 (2007).
20. Almeida, R. P., Blua, M. J., Lopes, J. R. S. & Purcell, A. H. Vector transmission of *Xylella fastidiosa*: Applying fundamental knowledge to generate disease management strategies. *Ann. Entomol. Soc. Am.* **98**, 775–786 (2005).
21. Paião, F., Meneguim, A., Casagrande, E. & Leite, R. Envolvimento de cigarras (Homoptera, Cicadidae) na transmissão de *Xylella fastidiosa* em cafeeiro. *Fitopatol. Bras.* **27**, 67 (2002).
22. Cornara, D. *et al*. Transmission of *Xylella fastidiosa* by naturally infected *Philaenus spumarius* (Hemiptera, Aphrophoridae) to different host plants. *J. Appl. Entomol.* **141**, 80–87 (2016).
23. Saponari, M. *et al*. Infectivity and transmission of *Xylella fastidiosa* by *Philaenus spumarius* (Hemiptera: Aphrophoridae) in Apulia, Italy. *J. Econ. Entomol.* **107**, 1316–1319 (2014).
24. Harper, S. J., Ward, L. I. & Clover, G. R. G. Development of LAMP and real-time PCR methods for the rapid detection of *Xylella fastidiosa* for quarantine and field applications. *Phytopathology* **12**, 1282–1288 (2010 erratum 2013).

25. Baldi, P. & La Porta, N. *Xylella fastidiosa*: host range and advance in molecular identification techniques. *Frontiers in Plant Science* **8**, 944, https://doi.org/10.3389/fpls.2017.00944 (2017).
26. Yuan, X. *et al.* Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's disease and oleander leaf scorch in the United States. *Phytopathology* **100**, 601–611 (2010).
27. European Plant Protection Organization. EPPO Standards PM 7 – Diagnostics PM 7/24 (2) *Xylella fastidiosa*. *Bull. OEPP/EPPO Bull.* **46**, 463–500 (2016).
28. Nunney, L., Elfekih, S. & Stouthamer, R. The importance of multilocus sequence typing: Cautionary tales from the bacterium *Xylella fastidiosa*. *Phytopathology* **102**, 456–460 (2012).
29. Nunney, L., Yuan, X., Bromley, R. E. & Stouthamer, R. Detecting genetic introgression: high levels of intersubspecific recombination found in *Xylella fastidiosa* in Brazil. *Appl. Environ. Microbiol.* **78**, 4702–4714 (2012).
30. Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. PCR inhibitors - occurrence, properties and removal. *J. Appl. Microbiol.* **113**, 1014–1026 (2012).
31. Boncristiani, H., Li, J., Evans, J., Pettis, J. & Chen, Y. Scientific note on PCR inhibitors in the compound eyes of honey bees. *Apis mellifera*. *Apidologie* **42**, 457–460 (2011).
32. Shamim, G., Ranjan, S. K., Pandey, D. M. & Ramani, R. Biochemistry and biosynthesis of insect pigments. *Eur. J. Entomol.* **111**, 149–164 (2014).
33. Purcell, A. H. & Finlay, A. Evidence for non-circulative transmission of Pierce's disease bacterium by sharpshooter leafhoppers. *Phytopathology* **69**, 393–395 (1979).
34. Yaseen, T. *et al.* On site detection of *Xylella fastidiosa* in host plants and in "spy insects" using the real-time loop-mediated isothermal amplification method. *Phytopathol. Mediterr.* **54**, 488–496 (2015).
35. Yaseen, T., Valentini, F., Santoro, F., Lorusso, D. & D'Onghia, A. M. The "Spy Insect" approach for monitoring Xylella fastidiosa in absence of symptomatic plants In: *D'Onghia, A. M., Brunel, S. & Valentini, F., eds. Xylella fastidiosa & the Olive Quick Decline Syndrome (OQDS) A serious worldwide challenge for the safeguard of olive trees - IAM Bari: CIHEAM (Centre International de Hautes Etudes Agronomiques Méditerranéennes)*, 2017–172 p (Série A Mediterranean Seminars, N° 121, Options Méditerranéennes) (2017).
36. D'onghia, A. M. CIHEAM/IAMB innovative tools for early surveillance and detection of Xylella fastidiosa. In: *D'Onghia, A. M., Brunel, S. & Valentini, F., eds. Xylella fastidiosa & the Olive Quick Decline Syndrome (OQDS) A serious worldwide challenge for the safeguard of olive trees - IAM Bari: CIHEAM (Centre International de Hautes Etudes Agronomiques Méditerranéennes)*, 2017–172 p (Série A Mediterranean Seminars, N° 121, Options Méditerranéennes) (2017).
37. Bextine, B., Tuan, S. J., Shaikh, H., Blua, M. & Miller, T. A. Evaluation of methods for extracting *Xylella fastidiosa* DNA from the glassy-winged sharpshooter. *J. Econ. Entomol.* **97**, 757–763 (2004).
38. Brady, J. A., Faske, J. B., Castañeda-Gill, J. M., King, J. L. & Mitchell, F. L. High-throughput DNA isolation method for detection of *Xylella fastidiosa* in plant and insect samples. *J. Microbiol. Methods* **86**, 310–312 (2011).
39. ANSES. Détection de *Xylella fastidiosa* par PCR en temps réel sur plantes hôtes. ANSES/ LSV/MA 039 version 1. Available from, https://www.anses.fr/fr/system/files/ANSES_MA039_Xylellafastidiosa_final.pdf (2015).
40. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, https://doi.org/10.1093/bioinformatics/btu033 (2014).
41. Gu, X. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* **12**, 546–557 (1995).
42. Stamatakis, A. Phylogenetic models of rate heterogeneity: A High Performance Computing Perspective. *International Parallel and Distributed Processing Symposium (IPDPS 2006)*, *Rhodes Island*, *Greece*, 8 pp (2006).
43. FigTree. Available online at, http://tree.bio.ed.ac.uk/software/figtree/ (2006).
44. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
45. Busby, J. R. BIOCLIM: a bioclimate analysis and prediction system. *Plant Protection Quarterly* **6**, 8–9 (1991).
46. Booth, T. H., Nix, H. A., Busby, J. R. & Hutchinson, M. F. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions* **20**, 1–9 (2014).
47. Carpenter, G., Gillison, A. N. & Winter, J. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and conservation* **2**, 667–680 (1993).
48. Godefroid, M., Cruaud, A., Streito, J.-C., Rasplus, J.-Y. & Rossi, J.-P. Climate change and the potential distribution of *Xylella fastidiosa* inEurope. *bioRxi*v, https://doi.org/10.1101/289876 (2018).
49. Elith, J. *et al.* Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151 (2006).
50. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 (2006).
51. Phillips, S. J. & Dudík, M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175 (2008).
52. Merow, C., Smith, M. J. & Silander, J. A. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* **36**, 1058–1069, https://doi.org/10.1111/j.1600-0587.2013.07872.x (2013).
53. Radosavljevic, A. & Anderson, R. P. Making better Maxent models of species distributions: complexity, overfitting and evaluation. *J. Biogeogr.* **41**, 629–643, https://doi.org/10.1111/jbi.12227 (2014).
54. Shcheglovitova, M. & Anderson, R. P. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. *Ecol. Model.* **269**, 9–17 (2013).
55. Muscarella, R. *et al.* ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity forMaxent ecological niche models. *Methods in Ecology and Evolution* **5**, 1198–1205, https://doi.org/10.1111/2041-210x.12261 (2014).
56. Liu, C., Berry, P. M., Dawson, T. P. & Pearson, R. G. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* **28**, 385–393 (2005).
57. Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. dismo: Species Distribution Modeling. R package version 1.1–4, https://cran.r/project.org/package=dismo (2016).
58. Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**, 38–49 (1997).
59. Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* **199**, 142–152 (2006).
60. ecospat: Spatial Ecology Miscellaneous Methods. version 2.1.1. https://cran.r-project.org/package=ecospat. (2016).
61. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302–4315, https://doi.org/10.1002/joc.5086 (2017).
62. Jiménez-Valverde, A. *et al.* Use of niche models in invasive species risk assessments. *Biol. Invasions* **13**, 2785–2797, https://doi.org/10.1007/s10530-011-9963-4 (2011).
63. Peterson, A. T. & Nakazawa, Y. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Glob. Ecol. Biogeogr.* **17**, 135–144, https://doi.org/10.1111/j.1466-8238.2007.00347.x (2008).
64. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–1978 (2005).

65. Elbeaino, T. *et al*. Identification of three potential insect vectors of *Xylella fastidiosa* in southern Italy. *Phytopathol. Mediterr.* **53**, 328–332 (2014).
66. Thorne, E. T., Stevenson, J. F., Rost, T. L., Labavitch, J. M. & Matthews, M. A. Pierce's disease symptoms: comparison with symptoms of water deficit and the impact of water deficits. *Am. J. Enol. Vitic.* **57**, 1–11 (2006).
67. Choi, H.-K., Iandolino, A., Silva, F. G. & Cook, D. R. Water deficit modulates the response of *Vitis vinifera* to the Pierce's disease pathogen *Xylella fastidiosa. Mol. Plant-Microbe Interact.* **26**, 643–657 (2013).
68. Daugherty, M. P., Lopes, J. R. S. & Almeida, R. P. P. Strain-specific alfalfa water stress induced by *Xylella fastidiosa. Eur. J. Plant Pathol.* **127**, 333–340 (2010).
69. Krugner, R. & Backus, E. A. Plant water stress effects on stylet probing behaviors of *Homalodisca vitripennis* (Hemiptera: Cicadellidae) associated with acquisition and inoculation of the bacterium *Xylella fastidiosa. J. Econ. Entomol.* **107**, 66–74, https://doi.org/10.1603/ec13219 (2014).
70. Kandel, P. P., Almeida, R. P., Cobine, P. A. & De La Fuente, L. Natural competence rates are variable among *Xylella fastidiosa* strains and homologous recombination occurs *in vitro* between subspecies *fastidiosa* and *multiplex. Mol. Plant-Microbe Interact.* **30**, 589–600 (2017).
71. Nunney, L., Schuenzel, E. L., Scally, M., Bromley, R. E. & Stouthamer, R. Large-scale intersubspecific recombination in the plant-pathogenic bacterium *Xylella fastidiosa* is associated with the host shift to mulberry. *Appl. Environ. Microbiol.* **80**, 3025–3033 (2014).
72. Nunney, L., Hopkins, D. L., Morano, L. D., Russell, S. E. & Stouthamer, R. Intersubspecific recombination in *Xylella fastidiosa* strains native to the United States: Infection of novel hosts associated with an unsuccessful invasion. *Appl. Environ. Microbiol.* **80**, 1159–1169 (2014).
73. Ben Moussa, I. E. *et al*. Seasonal fluctuations of sap-feeding insect species infected by *Xylella fastidiosa* in Puglia olive groves of Southern Italy. *J. Econ. Entomol.* **109**, 1512–1518 (2016).
74. Soubeyrand, S. *et al*. Inferring pathogen dynamics from temporal count data: the emergence of *Xylella fastidiosa* in France is probably not recent. *New Phytol.* **219**, 824–836, https://doi.org/10.1111/nph.15177 (2018).
75. OpenStreetMap: Access to Open Street Map Raster Images. R package version 0.3.3, https://cran.r-project.org/package=OpenStreetMap (2016).
76. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.r-project.org/ (2018 version 3.5.1 (Feather Spray)).
77. Giampetruzzi, A. *et al*. Draft genome sequence of CO33, a coffee-infecting isolate of *Xylella fastidiosa. Genome Announc* **3**, e01472–01415 (2015).

Acknowledgements

Author Contributions

Designed the study: A.C., J.Y.R. collected samples: A.C., J.Y.R., M.G., J.C.S., J.M.T.; developed/wrote protocols and conducted laboratory work: A.A.G., S.S. performed species distribution modelling: M.G., J.P.R.; discuss the results: A.C., A.A.G., S.N., J.P.R., S.S., J.Y.R. wrote the manuscript: A.C., J.Y.R. All authors commented on the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-33957-z.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.