Pre-visual symptoms of *Xylella fastidiosa* infection revealed in spectral plant-trait alterations

Zarco-Tejada[1*], P.J., Camino[2], C., Beck[1], P.S.A., Calderon[2], R., Hornero[2], A., Hernández-Clemente, R[3]., Kattenborn[4], T., Montes-Borrego[2], M., Susca[5], L., Morelli[6], M., Gonzalez-Dugo[2], V., North[3], P.R.J., Landa[2], B.B., Boscia[6], D., Saponari[6], M., Navas-Cortes[2], J.A.

[1] European Commission (EC), Joint Research Centre (JRC), Directorate D-Sustainable Resources, Via E. Fermi 2749 – TP 261, 26a/043, I-21027 Ispra (VA), Italy
[2] Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), Alameda del Obispo s/n, 14004 Córdoba, Spain
[3] Department of Geography, Swansea University, SA2 8PP Swansea, United Kingdom
[4] Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Kaiserstraße 12, 76131 Karlsruhe, Germany
[5] Dip. di Scienze del Suolo, della Pianta e degli Alimenti dell'Università di Bari, Bari, Italy
[6] CNR, Istituto per la Protezione Sostenibile delle Piante, Bari, Italy
[*] *Corresponding author*

ABSTRACT

Plant pathogens cause significant losses to agricultural yields, and increasingly threaten food security[1], ecosystem integrity, and societies in general[2–5]. *Xylella fastidiosa* (*Xf*) is one of the most dangerous plant bacteria worldwide, causing several diseases with profound impacts on agriculture and the environment[6]. Primarily occurring in the Americas, its recent discovery in Asia and Europe demonstrates that *Xf*'s geographic range has broadened considerably, positioning *Xf* as a re-emerging global threat that has caused socio-economic and cultural damage[7,8]. *Xf* can infect over 350 plant species worldwide[9], and early detection is critical for its eradication[8]. Here, we show that changes in plant functional traits retrieved from airborne imaging spectroscopy and thermography can reveal *Xf* infection in olive trees before symptoms are visible. We obtained accuracies of disease detection, confirmed by qPCR, exceeding 80% when high-resolution fluorescence quantified by 3D simulations and thermal stress indicators were coupled with photosynthetic traits sensitive to rapid pigment dynamics and degradation. Moreover, we found that the visually asymptomatic trees originally scored as affected via spectral plant trait alterations developed *Xf* symptoms at almost double the rate of the asymptomatic trees classified as not affected by remote sensing. We demonstrate that spectral plant trait alterations caused by *Xf* infection are detectable pre-visually at the landscape scale, a critical requirement to help eradicate some of the most devastating plant diseases worldwide.

Acronyms - **A_{nth}**: Anthocyanins; **C_{a+b}**: Chlorophyll *a+b*; **C_{x+c}**: Carotenoids; **CWSI**: Crop Water Stress Index; **E**: irradiance; **Fi**: Fluorescence efficiency; **Ft**: Leaf-level steady-state fluorescence; **FP**: False Positives; **FW1**: Intensive field work 1; **FW2**: Intensive field work 2; **κ**: Kappa coefficient; **L**: radiance; **LDA**: Linear Discriminant Analysis; **NBHI**: Narrow Band Hyperspectral Indicators; **NDVI**: Normalized Difference Vegetation Index; **NIR**: Near-infrared; **NNE**: Neural Network; **NPQI**: Chlorophyll Degradation Phaeophytinization-based Spectral Trait; **OA**: Overall Accuracy; **PS**: Pigment- and Structure-based Functional Traits; **PSFT**: Pigment-Structural-Fluorescence-Temperature Plant Functional Traits; **qPCR**: Quantitative Polymerase Chain Reaction assay; **RGB**: Red-Green-Blue; **ROC**: Receiver Operating Characteristic analysis; **RT**: Radiative Transfer; **SIF**: Solar-induced Fluorescence; **SVI**: Spectral Vegetation Indices; **SVM**: Support Vector Machine; **TN**: True Negatives; **TR**: Training dataset; **TS**: Testing dataset; **V+A+Z**: Violaxanthin (**V**), Antheraxanthin (**A**), Zeaxanthin (**Z**) pool; **VHR**: Very-High-Resolution; ***Xf***: *Xylella fastidiosa.*

Xylella fastidiosa (*Xf*) is considered one of the most dangerous plant pathogens worldwide[6]. It can infect over 350 plant species[9], causing diseases in several crops and large economic losses[8]. In America, this xylem-limited plant pathogenic bacterium is associated with detrimental diseases in high-value crops, such as Pierce's disease in grapevines and variegated chlorosis in citrus[10]. Its spread has recently gained a global dimension[11]: already widely distributed in the Americas and detected in Iran and Taiwan, *Xf* has been known to be present in Europe since 2013 after its official identification in Italy[12] causing economic and societal damage[8].

The spread of *Xf* within Europe has thus far not been contained[7]. Outbreaks detected in France and recently in Spain have raised concerns of *Xf* spreading to the world's largest olive-growing area (over 2.5 million hectares) and throughout Mediterranean agriculture[8]. The identification of all three main subspecies of *Xf* (i.e., *fastidiosa*, *multiplex*, and *pauca*) in Europe broadens the threat to several other crop plants, including almond, citrus, and grapevine, but also to ornamental trees as well as elms, oaks and sycamores. A major difficulty for *Xf* containment arises from its very wide host range, with infections that do not cause symptoms in some host–strain combinations, despite the infected hosts continuing to act as inoculum sources[9]. This threat is further exacerbated because *Xf* can be spread via xylem-sap sucking insects without any specific vector relationship[8], and due to increased global trade.

Alarms have been raised by both the international scientific community[8] and the media[7], pointing out that eradication of *Xf* will require robust monitoring and early detection of plants that show little to no signs of decline at the early stages of infection. A major limitation of standard large-scale mapping methods based on red and near-infrared (NIR) (e.g. the Normalized Difference Vegetation Index [NDVI] and its multiple variations obtained from broadband satellite sensors) is that they are useful only for detecting the advanced stages of disease damage, i.e. when canopy defoliation, leaf wilting, and chlorosis are apparent[13]. Additionally, current hyperspectral satellite sensors lack the spatial resolution to distinguish individual tree crowns. Accordingly, *Xf* eradication efforts involving its early detection necessitate high spatial resolution (i.e. sub-meter) imaging spectroscopy and thermal data to assess subtle changes in spectral features and traits, a technology that can be potentially deployed at large scales with airborne platforms[14].

We carried out intensive multi-year *in-situ* inspections of >7000 trees and airborne imaging data in 15 olive orchards, finding that physiological alterations caused by *Xf* infection at the pre-visual stage were detectable in functional plant traits assessed remotely by hyperspectral and thermal sensors. We confirmed the presence of *Xf* infection in all selected orchards by testing at least two symptomatic trees per plot by quantitative polymerase chain reaction assay[15] (qPCR). Additionally, we sampled one of the olive fields more extensively for an orchard-level validation of the remote sensing model testing, by qPCR assays, 67 out of the 157 trees spanning the full range of symptoms, i.e. from asymptomatic to severely affected. Although quantitative PCR is considered the most sensitive diagnostic approach, its accuracy under field conditions for the detection of the *Xf* in host plants is affected by the period of sampling and the uneven distribution of the bacterium in the large canopy of the olive trees (especially at the early stage of infection). Moreover, this type of laboratory assay is time consuming and costly, and requires skilled and trained personnel. For these reasons, we evaluated non-destructive remote sensing methods comprising the acquisition of spectroscopy data to build 40 cm radiance and reflectance scenes in 260 narrow spectral bands (Fig. 1a;b) and in the thermal spectral region (Fig. 1a;c). The entire flight campaigns covered three areas within the *Xf*-affected olive growth region in Southern Italy and scanned ca. 200,000 individual trees in 2016 and 2017, quantifying tree-level physiology-related narrow-band spectral traits, Solar-induced Fluorescence (SIF) and fluorescence efficiency (Fi) by Monte Carlo 3-D scene generation (Fig. 1d) that modelled the individual tree fluorescence emissions (Fig. 1e) at the tree radiance level (Fig. 1f).

We used a multi-layered functional plant trait scheme to extract the alterations caused by *Xf* from a pool of physiology-related narrow-band hyperspectral indicators (NBHI). This pool included plant physiological traits specifically related to rapid changes in photosynthetic pigments and leaf processes not simulated by any existing radiative transfer (RT) model, e.g. the de-epoxidation state of the xanthophyll-cycle pigments via the violaxanthin (V), antheraxanthin (A), and zeaxanthin (Z) pool[16], and chlorophyll degradation via phaeophytinization[17,18] (see Supplementary Table 1[13]). In addition, we assessed traits sensitive to *Xf* infection (i.e. anthocyanins and carotenoid / chlorophyll ratios) by a hybrid wavelet-inverted model inversion method (Supplementary Table 2; Supplementary Fig. 1), and quantified SIF emission and Fi by a multi-step LUT-based inversion scheme (Supplementary Table 3, Supplementary Fig. 2). The inversion of radiative transfer models enables the simultaneous and independent retrieval of multiple leaf and canopy traits linked to physiological processes in plants. Thus, compared to single-band and index-based relationships from radiance or reflectance spectra which simultaneously relate to several traits (e.g. both photosynthetic pigments and structure), the model-inverted traits space is more likely to reveal the physiological processes associated with the disease. Furthermore, the process-based retrieval of traits by physical models increases the potential transferability of findings to other data sets, diseases, and plant species. Nevertheless, specific narrow-band spectral indices that track processes currently not simulated by any radiative transfer simulations can complement model-estimated traits.

To reveal the gas exchange dynamics associated with *Xf* symptoms, we incorporated a functional trait consisting of temperature-based plant stress indicators linked to stomatal conductance and tree transpiration alterations. Linear, as well as machine- and deep-learning algorithms (linear discriminant analysis, LDA; support vector machine, SVM; neural network ensemble, NNE, see Methods in Extended Material) fed by the pool of functional plant traits via receiver operating characteristic (ROC) analysis revealed that the chlorophyll degradation phaeophytinization-based spectral trait (NPQI)[17,18] calculated in the blue region, and the thermal-based stress trait (CWSI, Crop Water Stress Index) best distinguished *Xf*-symptomatic from asymptomatic trees (Fig. 2a) in both years (Fig. 2b), followed by anthocyanins (A_{nth}), carotenoids (C_{x+c}) and solar-induced fluorescence. Notably, the importance of the functional traits varied as a function of *Xf*-symptom severity: NPQI and CWSI most reliably distinguished symptomatic from asymptomatic material (Fig. 2a, left-side bars), but were of lesser importance to discriminate between initial and advanced stages of the disease. For these symptomatic trees, solar-induced fluorescence was the most sensitive functional trait to detect the severity of *Xf* symptoms (Fig. 2a, right-side bars).

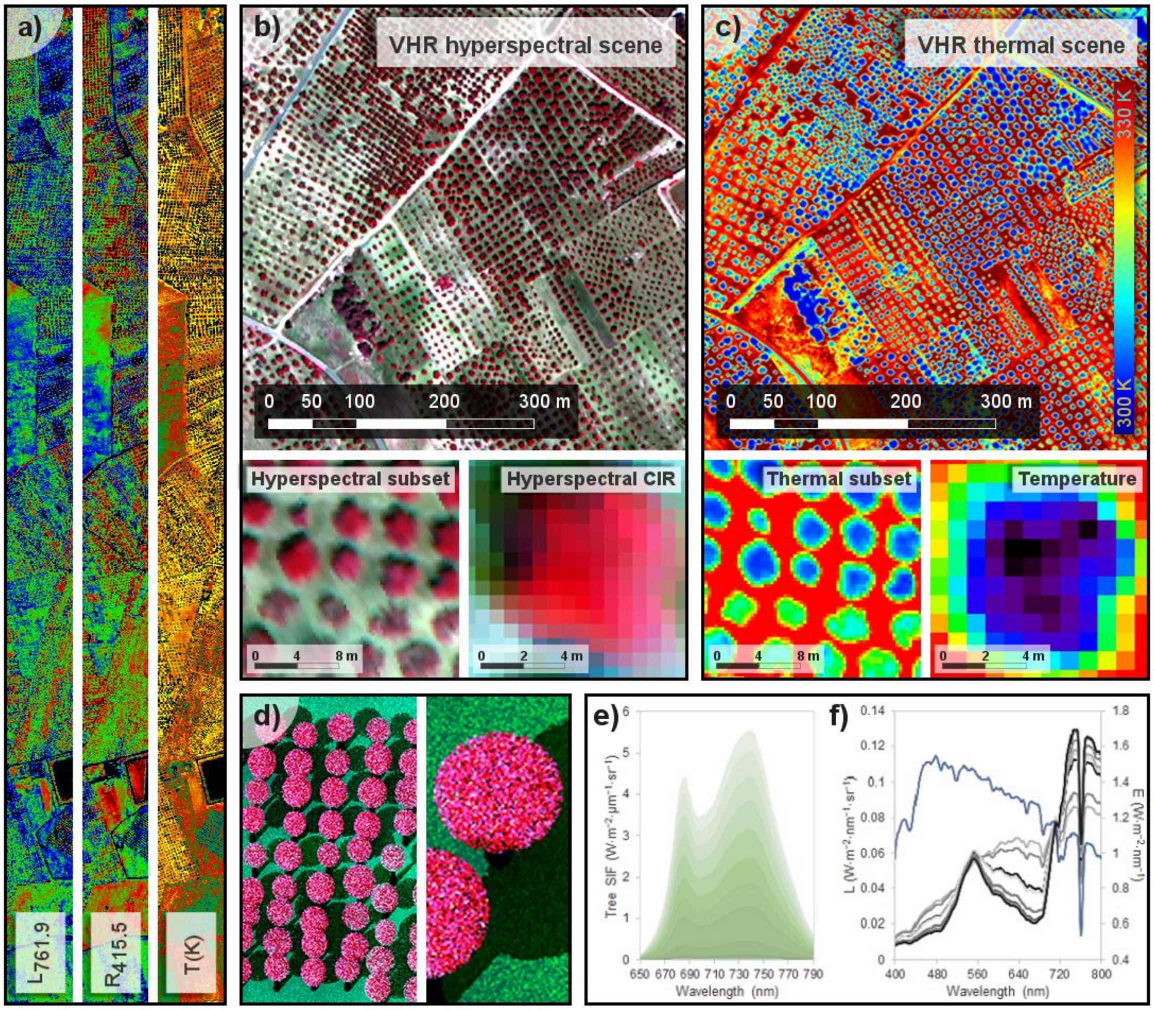

Fig. 1. Imagery acquisition and plant-trait fluorescence retrievals. a, Strips of airborne images of 40-cm hyperspectral radiance collected at the O_2-A band, reflectance at 415 nm (used to calculate NPQI), and temperature (in K). Subsets of the very-high-resolution (VHR) hyperspectral (**b**) and thermal imagery (**c**) enable the identification of single trees to extract tree-crown radiance (L), reflectance (R) and temperature. **d**, Monte Carlo simulation modelled solar-induced fluorescence (SIF) emission via 3-D scenes generated with FluorFLIGHT (**e**) from tree radiance (L) and irradiance (E) (**f**) to quantify fluorescence efficiency (Fi) by radiative transfer.

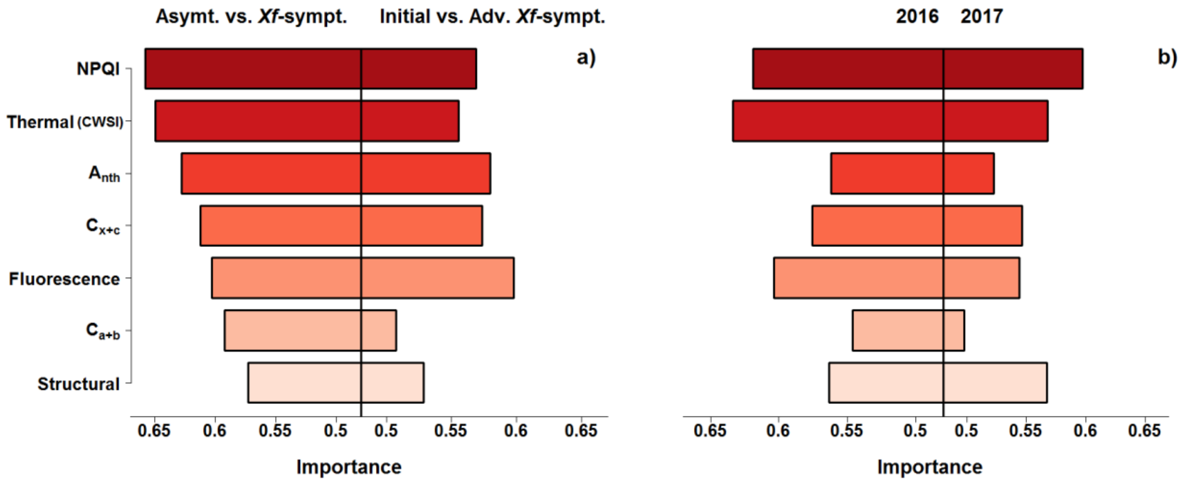

Fig. 2. Contribution of remote sensing plant traits to pre-visual *Xf* symptom detection. **a**, ROC analysis from the pool of hyperspectral and thermal plant functional traits used to detect asymptomatic vs. *Xf*-symptomatic trees (left bars) and for initial vs. advanced *Xf*-symptomatic trees (right bars). **b**, The robustness across years of the functional traits for asymptomatic vs. *Xf*-symptomatic trees. The ROC analysis was performed using the training data set (TR, n=5,852 trees).

The sensitivity of these physiology-based remote-sensed plant traits to pre-visual and early stages of the *Xf*-infection is supported in the literature by work that shows the photoprotective role of carotenoids (C_{x+c}) and the protection from damage induced by environmental stresses and plant pathogens provided by flavonoids such as anthocyanins (A_{nth})[19]. These compounds accumulate in *Xf*-infected plant material[20] and are produced by the degradation of the chlorophyll molecule into phaeophytin under stress conditions[17,18]. In addition, the alterations in stomatal regulation[21] and photosynthesis caused by plant–pathogen interactions[22] lead to decreased fluorescence[13,23] and transpiration[24], and produce phenolic plant-defense compounds[25].

The alterations of plant functional traits we detected remotely were highly consistent with *Xf*-induced leaf physiological changes measured *in-situ*. In particular, the changes we observed in the *in-situ* A_{nth}, steady-state fluorescence Ft, and temperature leaf traits (Fig. 3a; Supplementary Fig. 3) were in line with the alterations observed in the corresponding traits quantified from the imagery, such as A_{nth} (Supplementary Fig. 1), SIF and CWSI (Fig. 3d). These traits differed significantly between asymptomatic and symptomatic leaves, even when symptoms were mild (Tukey's HSD test, $P<0.05$) (Supplementary Fig. 3). Moreover, the high-resolution images revealed between- and within-tree-crown patterns of the functional traits associated with *Xf* infection (Fig. 3b;c). Although widely used in global monitoring of vegetation, NDVI did not differ significantly between asymptomatic and symptomatic trees (Fig. 3d), and was therefore unable to detect non-visual symptoms of *Xf* infection. We found that the reflectance changes in the blue region consistently tracked early and initial *Xf* symptoms, in particular the 415 and 435 nm spectral bands used to calculate the chlorophyll degradation phaeophytinization-based spectral trait NPQI[17,18], which was the NBHI indicator most sensitive to *Xf* infection. The SIF calculated from the airborne radiance imagery and CWSI calculated from the remotely sensed tree crown temperature, showed statistically-significant ($P<0.001$) and consistent trends for early *Xf* symptoms.

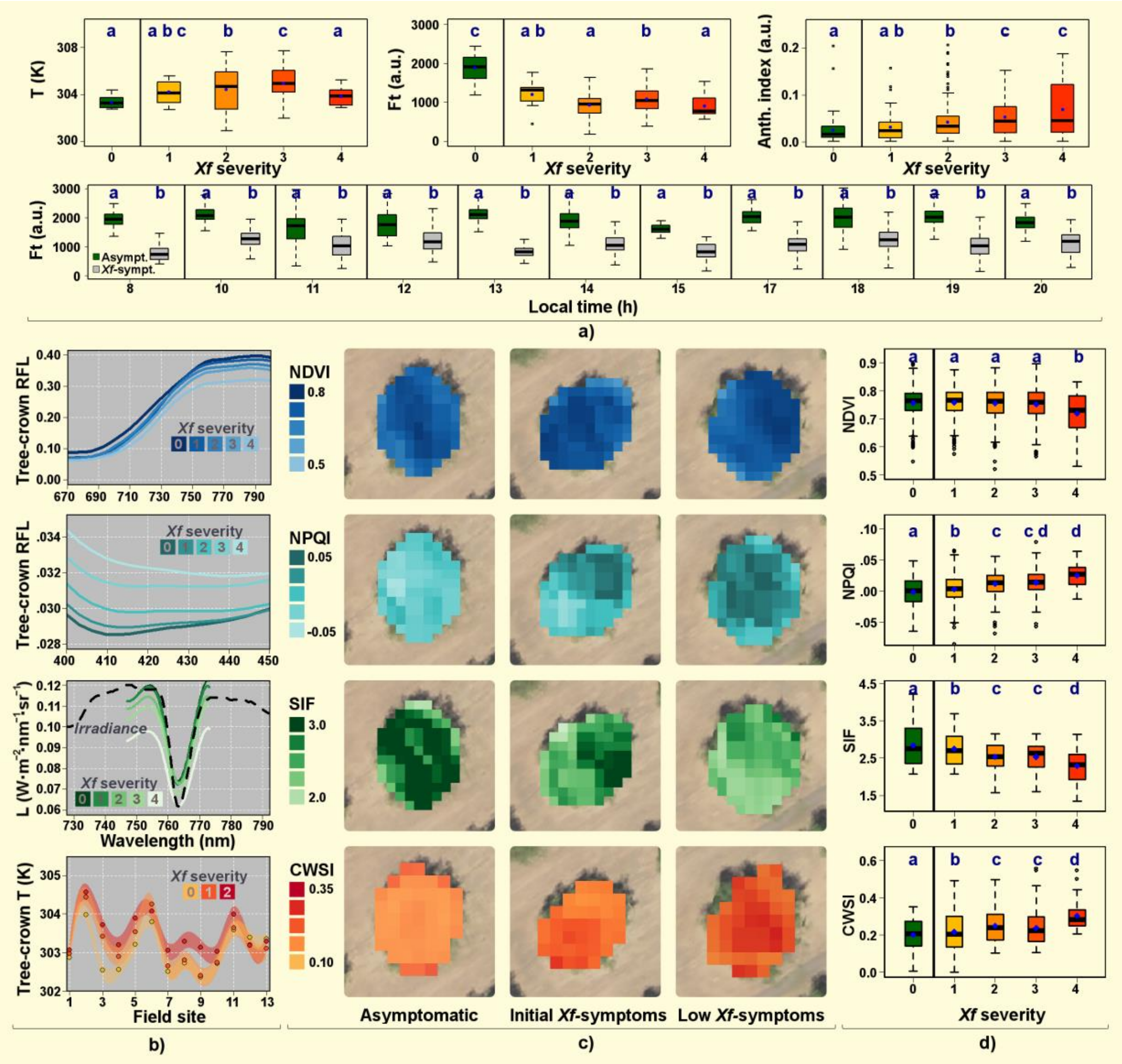

Fig. 3. Relationships between remote-sensed functional plant traits and *Xf* disease severity levels at leaf and canopy levels. a, Temperature (T, n=922 leaves), fluorescence (Ft, n=1,197 leaves) and anthocyanins (A_{nth}, n=939 leaves), as well as hourly Ft (n=2,863 leaves), measured in asymptomatic (*Xf* severity = 0) and increasingly symptomatic leaves of *Xf*-infected olive trees. **b**, Mean tree-crown reflectance for trees with increasing severity of *Xf* symptoms in the red-NIR region (n=923 trees), blue region (n=923 trees), O_2-A radiance region for SIF quantification (n=923 trees), and temperature (n=1,493 trees). The standard deviation for the tree-crown T data is represented as shaded. **c,** Respective associated maps of NDVI, NPQI, SIF and CWSI, showing the within-crown variation of traits in asymptomatic, initial, and low *Xf*-symptomatic trees. **d**, Trait values across the entire sample of trees for NDVI, NPQI and SIF (n=1,493 trees) and CWSI (n=1,446 trees). The disease severity at leaf and canopy levels was compared by one-sided Tukey's HSD test at 5%. Severity levels with same letter are not significantly different (Tukey HSD test, *p-value* <0.05). In the box plots, the black line represents the median, and the top and bottom are the 75th and 25th quartiles. The whiskers are the upper and lower limits based on the interquartile ranges (Q±1.5xIQR). Average values are shown with a blue point. The outliers (circles), are the values out of the upper and lower limits. a.u.: arbitrary units.A

pool of plant functional traits comprising pigment and structural traits, together with a flux-based fluorescence trait and temperature (PSFT) obtained the best overall accuracy (OA) and kappa coefficient (κ) for *Xf* detection through the SVM algorithm, yielding OA = 80.9% and κ = 0.61 (Fig. 4a; Supplementary Tables 4 and 5). By contrast, models built without SIF and temperature traits (i.e. the Pigment- and Structure-based Functional Traits, PS model), and particularly one limited to standard red-green-blue (RGB)-NIR spectral vegetation indices (SVI) commonly found in satellite sensors (NDVI, and blue / green / red ratios; SVI model), obtained the lowest accuracies (OA=65.4%; κ=0.29). We obtained these results through validation with visual inspection data collected by plant pathologists from 1,332 trees per year in 15 fields, generating a large dataset with statistical robustness and ample variability in disease severity levels, tree structure and age, and agronomic management of the orchards within the *Xf*-infested zone.

We evaluated the accuracies of the remote sensing-based SVM-PSFT disease detection model and the visual inspections using quantitative PCR assay data obtained in a selected olive orchard. The assessment of the orchard-level remote sensing model validated with the tree-level qPCR dataset yielded OA=94.03% and κ=0.88. The performance of the visual inspection against qPCR (OA=77.62% and κ=0.55) showed the validity of the evaluations by the plant pathologists, but reflected a lower performance than that using remote sensing methods due to the impossibility of visually detecting the asymptomatic infections that were detected by qPCR. The validation of the remote-sensing model with qPCR data enabled the generation of a spatial map of disease incidence prediction by remote sensing, revealing infected asymptomatic trees that were missed by the visual evaluations (Fig. 5a) but detected by remote sensing (Fig. 5b). Among all trees measured in this particular orchard by qPCR (n=67), those visually considered asymptomatic by plant pathologists (n=40) but proven infected via qPCR (n=11) were detected as infected by remote sensing with 91% accuracy. When the analysis was extended to eight orchards where the qPCR-sampled trees were visible in the imagery (n=100), the accuracy of the remote sensing model validated with the tree-level qPCR dataset yielded OA=96% and κ=0.92, whereas the performance of the visual inspection against qPCR remained at the same level as the orchard-level analysis (OA=77% and κ=0.54). Moreover, the remote sensing SVM-PSFT model detected 92.9% of the infected asymptomatic trees (qPCR=1; DS=0) that were missed by visual assessment (Supplementary Table 6).

These results obtained by remote sensing and validated with qPCR data suggested the existence of trees in the very early stage of the disease that were missed by the visual evaluations. To explore whether our remote sensing model fed by plant functional traits actually detected the early symptoms at a pre-visual stage, a temporal dimension was added in the analysis. Indeed, a critical finding of this study arose from further investigation of the trees seemingly wrongly considered symptomatic by remote sensing (i.e. those initially considered ‘false positives’ based on examination by plant physiologists) over the course of two years through periodic field revisits. False-positive cases may arise from: i) error and uncertainty inherent to the remote sensing model used for detecting affected trees; and ii) trees that were indeed affected by *Xf* but did not yet display the typical visible symptoms upon which plant pathologists rely. Thus, we revisited *in situ* (Fig. 4b; Supplementary Table 7) the trees identified as symptomatic by the remote-sensing plant functional trait model applied to the 2016 image data (F1) but classified at the time as asymptomatic by plant pathologists based on the absence of visible symptoms (false positives, FP; n=178 by SVM).

During these field revisits conducted four (indicated as R1), eight (R2), eleven (R3), and twelve months (R4) after the flight at the F1 date, we recorded the development of visible *Xf* symptoms on 1,700 out of the 3,328 trees initially evaluated. Four months after F1, 61% of the false positives had developed symptoms, while only 39% of the asymptomatic trees classified as unaffected by the remote sensing-driven PSFT model had (true negatives, TN, n=818, two-sided t-test: $P < 0.001$). This difference in visible symptom development was maintained throughout the one-year post-flight evaluations (R1, R2, R3 and R4), with FP trees consistently developing symptoms sooner than TN trees. These results obtained in the multitemporal revisit scheme and via qPCR confirmed that the remote sensing-driven PSFT model based on plant functional traits was able to detect *Xf* symptoms earlier than standard visual inspections by plant pathologists. The ability to detect pre-visual infections is particularly relevant given the threat of infected but asymptomatic trees contributing to the *Xf* epidemics, as plants artificially infected with *Xf* and maintained in controlled environmental conditions take 10 to 12 months to start developing visible symptoms[8,12].

Notably, our analysis was not based just on single spectral bands or indices to feed the model. Instead, we used radiative-transfer to independently quantify physiological traits linked to photosynthesis, pigment degradation, and structural changes of trees undergoing early stress caused by *Xf* infection. This methodology permits generalization and transfer to other plant species or diseases, since the retrieved traits are closely or even directly linked to the physiological changes occurring in affected vegetation. The relative importance of these traits for disease detection will differ among pathogens and host plants, depending on the physiological effects associated with the disease. Operational remote-sensing based detections of pathogen infections should thus rely on the spectral bandsets enabling the retrieval of the most sensitive plant traits linked with a particular disease. In our case, aircraft payloads imaging <10 narrow bands (e.g. 10 nm or less) in the visible-near infrared region in tandem with a broad-band thermal sensor would reach overall accuracies exceeding 70%. As global trade increasingly exposes natural and agricultural systems to exotic pathogens, such advanced large-scale physiology-focused remote sensing methods relying on plant functional traits could prove critical to prevent and manage plant disease epidemics worldwide.

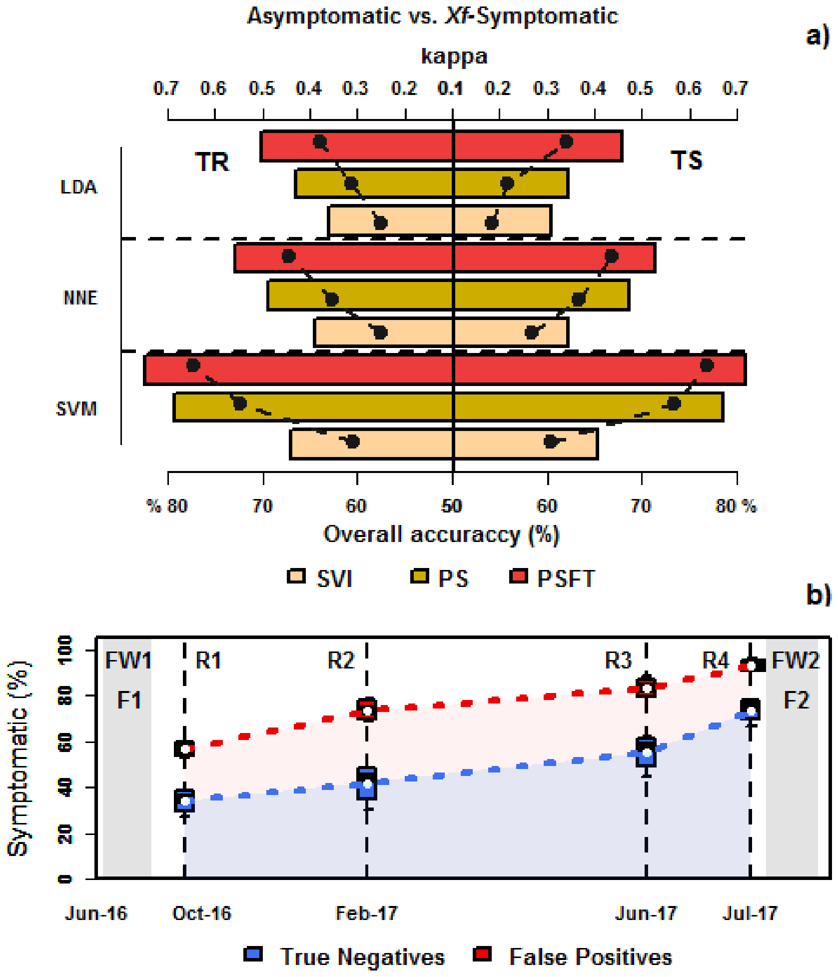

Fig. 4. Remote sensing model performance and re-visit analysis results. a, Overall accuracy (bars) and kappa coefficient (κ, bullets) of linear discriminant analysis (LDA), neural network (NNE) and support vector machine (SVM) algorithms distinguishing asymptomatic from *Xf*-symptomatic trees using as inputs standard vegetation indices calculated from RGB-NIR bands (SVI), Pigment- and Structure-based Functional Traits (PS), and Pigment-, Structure-, Fluorescence and Temperature-based Functional Traits (PSFT). Statistics are shown separately for the data (n=7,315 trees) used in training (TR, n=5,852 trees) and testing (TS, n=1,463 trees) for each of the three algorithms. **b**, Fraction of trees that were asymptomatic in June 2016 but showed visible symptoms during later revisits, for trees classified as non-symptomatic (n=818 trees for SVM, n=588 trees for NN and n=534 trees for LDA) and symptomatic (n=178 trees for SVM, n=408 trees for NN and n=462 trees for LDA) by remote sensing (true negatives, TN, and false positives, FP, respectively). F1 and F2 indicate the dates of the airborne imaging campaigns, which corresponded with intensive field work (FW1 and FW2). The field revisits conducted are indicated as R1, R2, R3 and R4. The dotted blue and red lines represent the cumulative sum of the fraction of trees that were identified as TN and FP by the three algorithms. In the box plots, the black line within the box represents the median of the predictions of the three algorithms, and the top and bottom of the box are the 75th and 25th quartiles, respectively. The whiskers represent the upper and lower limits based on the difference with the interquartile ranges (Q±1.5xIQR). The average percentage predicted by the three algorithms is shown with a white point within the boxplot.

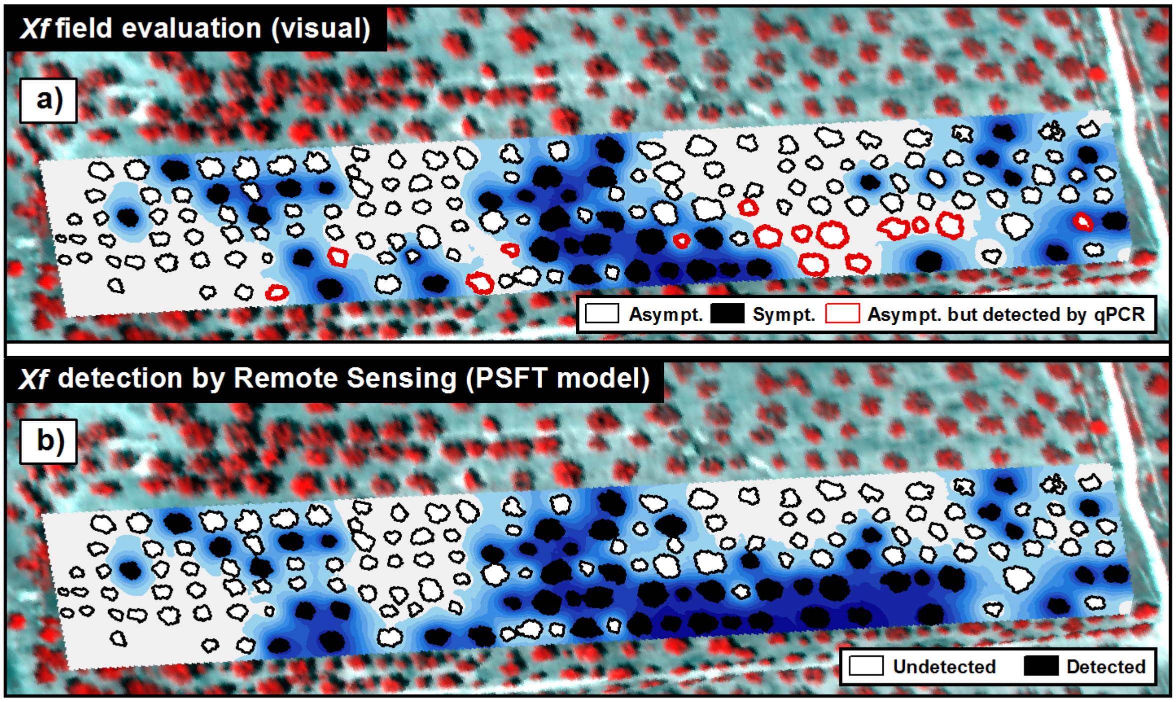

Fig. 5. Field evaluation, qPCR tests and remote sensing spatial predictions. a, Map of an olive orchard imaged by thermal and hyperspectral remote sensing showing the visual evaluation by plant pathologists in the field. **b**, Remote sensing PSFT model used to detect *Xf*-affected trees. The visually asymptomatic trees assessed as affected by qPCR (shown with red border) in (a) and therefore missed in the field evaluations by plant pathologists were detected by remote sensing using functional traits (b) with 91% accuracy. Background in stronger blue tones shows the areas more affected by *Xf*.

Methods

Field data collection. We assessed incidence and disease severity (DS) of *Xf*-induced symptoms in the field in June 2016 and July 2017 in 15 orchards in the *Xf*-infected area of Puglia, Southern Italy. Planting density and overall orchard management were highly variable within the selected area. We evaluated DS by visually inspecting every tree for symptoms of canopy desiccation and assessing it on a 0–4 rating scale according to the percentage of canopy affected by the disease symptoms; 0 indicated the absence of visually detectable symptoms (asymptomatic) and 4 referred to trees showing canopies with a prevalence of dead branches. In total, we evaluated 3,328 trees in 2016 [1,442 (DS = 0), 762 (DS = 1), 802 (DS = 2), 250 (DS = 3), and 72 (DS = 4)] and 3,987 trees in 2017 [2,607 (DS = 0), 687 (DS = 1), 555 (DS = 2), 122 (DS = 3), and 15 (DS = 4)]. Most of the olive orchards sampled had old trees (>50 years old) of cultivars Ogliarola Salentina and Cellina di Nardò, the native and widespread cultivars in the area. These cultivars have been shown to be highly susceptible to the CoDiRO strain associated with the Italian *Xf* epidemic. *Xf*-infected trees of both cultivars typically show severe desiccation that rapidly encompasses the entire canopy (within 2–3 years), and causes complete canopy die-back. Only one olive orchard consisted of trees of the Leccino cultivar (ca. 35 years old), which has genetic traits of resistance to *Xf*, as demonstrated by the lower bacterial concentrations in trees of this cultivar and the milder symptoms in infected trees[26].

During the field campaigns, we conducted different physiological measurements on leaves (Fig. 3a; Supplementary Fig. 3). Flavonoid (FLAV) concentration, chlorophyll content, anthocyanin content index, nitrogen balance index (NBI), and leaf temperature were measured on 15/25 asymptomatic/symptomatic leaves per tree using a leaf clip Dualex 4 (Force-A, Orsay, France). On the same leaves, the steady-state leaf fluorescence yield (Ft) and the leaf reflectance within the visible and near-infrared regions were measured with a FluorPen FP100 and PolyPen RP400, respectively (Photon Systems Instruments, Brno, Czech Republic), calculating leaf NPQI. We conducted a revisit assessment of disease severity in October 2016 and February, June, and July of 2017, re-evaluating 1,700 of the 3,328 trees originally evaluated in June 2016. In the 15 olive orchards selected for symptom scoring, we confirmed the presence of *Xf* infections by sampling and testing at least two symptomatic trees per plot. Diagnostic tests were performed using a quantitative PCR (qPCR) assays[15] in all orchards under study. In addition, one of the orchards was selected for a more extensive testing by qPCR assay, using 67 out of the 157 trees of this orchard. This qPCR dataset was used to validate the remote sensing and the visual evaluation methods. Based on the qPCR assays, the trees were categorized as positive (presence of infection) or negative (no bacterial infection detected) based on the resultant quantification cycle (Cq) values. Clear-cut values were consistently obtained for the trees, both symptomatic and asymptomatic, categorized as qPCR-positive (i.e. Cq ranging from 23 to 28; a positive result is considered if $Cq < 35$ and a clear exponential fluorescence curve is observed). Conversely, no fluorescence ($Cq=0$) was detected in the trees categorized as qPCR-negative. We used the data from eight orchards where the qPCR-sampled trees were visible in the imagery ($n=100$) for further statistical analysis. In particular, we evaluated the detection by the SVM-PSFT remote sensing model of the *Xf*-infected trees ($n=58$), splitting them into infected symptomatic ($qPCR=1$; $DS\geq1$; $n=44$) and infected asymptomatic trees ($qPCR=1$; $DS=0$; $n=14$) as assessed by qPCR in the laboratory.

Hyperspectral and thermal image data collection and processing. We acquired imagery on 28 June 2016 and 5 July 2017 over 1,200 ha within the *Xf*-infected area using a hyperspectral sensor and a thermal camera on board a manned aircraft. Both cameras were flown 500 m above ground level (AGL) at midday, acquiring hyperspectral and thermal imagery at 40 cm and 60 cm pixel resolution, respectively. We covered the visible and near-infrared regions with a micro-hyperspectral imager (VNIR model, Headwall Photonics, Fitchburg, MA, USA) operating in the spectral mode of 260 bands acquired at 1.85 nm/pixel and 12-bit radiometric resolution, yielding 6.4 nm full-width at half-maximum (FWHM) with a 25-micron slit in the 400–885 nm region. We set the frame storage rate on board the aircraft to 50 frames per second with 18 ms integration time. The 8-mm focal length lens yielded an instantaneous field of view (IFOV) of 0.93 mrad and an angular field of view (FOV) of 49.82°. We calibrated the hyperspectral sensor radiometrically in the laboratory with an integrating sphere (CSTM-USS-2000C Uniform Source System, LabSphere, North Sutton, NH, USA) using coefficients derived from a calibrated uniform light source at four illumination and six integration times. Atmospheric correction enabled the conversion of radiance values to reflectance using total incoming irradiance simulated with the SMARTS model[27,28]. In addition, we measured aerosol optical depth in the field at 550 nm with a Micro-Tops II Sunphotometer model 540 (Solar LIGHT Co., Philadelphia, PA, USA) during the flight. We ortho-rectified the hyperspectral imagery with PARGE (ReSe Applications Schläpfer, Wil, Switzerland), using inputs from an inertial measuring unit (IMU) (IG500 model, SBG Systems, France) installed onboard and synchronized with the micro-hyperspectral imager. Due to the high spatial resolution collected (40 cm) and the large size of most of the trees studied (>5 m) spatial binning was applied to increase the signal-to-noise ratio (SNR) of the instrument. In addition, we applied spectral binning due to the large number of spectral bands collected with oversampling (260 bands @ 1.85 nm sampling interval). After performing both spatial and spectral binning, SNR increased to values >300:1, showing radiance spectra with absence of noise (Fig. 1f) and in the reflectance spectra (Fig. 3b). The thermal camera (FLIR SC655, FLIR Systems, USA) had a resolution of 640×480 pixels and was equipped with a 24.6 mm f/1.0 lens connected to a computer via the GigaE protocol. This camera has a spectral response in the range of 7.5-14 μm and operates with a thermoelectric cooling stabilization, yielding high sensitivity below 50 mK. We calibrated the camera in the laboratory using a blackbody (model P80P, Land Instruments, Dronfield, UK) at varying target and ambient temperatures, and in the field through vicarious calibrations using surface temperature measurements obtained following Calderon *et al.*[13].

The high-resolution hyperspectral and thermal imagery acquired over the orchard allowed single-tree identification using automatic object-based crown detection algorithms. The algorithms were used to calculate mean temperature and hyperspectral reflectance for pure crowns. We used image segmentation procedures as described in Calderón *et al.*[29]. In this study, we applied four image segmentation methods to the thermal and hyperspectral images to extract temperature, radiance, and reflectance spectra from each pure tree crown. The very high-resolution imagery acquired enabled the identification and delineation of each tree crown independently in the thermal and hyperspectral datasets, minimizing background and within-crown shadow effects at the border pixels of each tree crown. The object-based image segmentation methods selected for the results reported here were Niblack's thresholding method[30] and Sauvola's binarization techniques[31] to separate tree crowns from the background. Next, we applied a binary watershed analysis using the Euclidean distance map for each object[32] to automatically separate trees with overlapping crowns. We calculated narrow-band spectral indices for each tree crown from the 260 spectral bands extracted by image segmentation. The spectral index-based traits explored in this study are closely related to specific features of leaf physiology, and therefore potentially sensitive predictors of the disease[13]. Thus, according to the effects of *Xf* infection in olive trees, we selected spectral indices from the plant-trait functional groups related to chlorophyll, carotene and xanthophyll pigments.

Model inversion methods. The derivation of canopy structural parameters and leaf biochemical constituents from each individual tree was performed by inversion of the radiative transfer model PROSAIL for the pure-vegetation pixels extracted from each tree crown. The model couples the leaf reflectance PROSPECT model, accounting for leaf properties such as pigment concentrations, and the canopy reflectance model SAIL, which accounts for canopy structural properties, such as leaf inclination and the sun-observer geometry. The versions used in the present study were PROSPECT-D[33] and 4SAIL[34], respectively. The inversion of PROSAIL was performed using a Look-up-Table (LUT) approach, in which randomized input parameters (Supplementary Table 2) are used to simulate canopy reflectance data, which was then compared to the acquired airborne spectra. To reduce the complexity and thus alleviate the ill-posed problem of the LUT inversion, we fixed several parameters by assuming that their variation is relatively low for the canopies under investigation or that the spectral range considered (400-885 nm) is not affected by these parameters. The variable parameters considered comprised chlorophyll content, carotenoid content, anthocyanin content, mesophyll structure, leaf area index and the average leaf angle. For the LUT generation, the values for these parameters were sampled from a uniform distribution within a range that is plausible for the assessed plant canopies (Supplementary Table 2). Previous studies demonstrated that wavelet analysis improved radiative transfer model inversions[35–37]. It decomposes the reflectance spectra into frequency components of different scales and thus spectral characteristics, such as absorption features of plant pigments. Accordingly, the correspondence in terms of RMSE between simulated spectra and airborne spectra was measured using a transformation of the reflectance spectra into 6 continuous wavelets derived by a Gaussian kernel. The estimates for each trait were derived by selecting the 1% of the LUT entries and respective spectra that resulted in the smallest RMSE. The parameter values of these LUT entries were subsequently weighted by their RMSE and averaged. A summary of the traits retrieved for each severity level is given in Supplementary Fig. 1.

We retrieved sun-induced chlorophyll fluorescence (SIF) emission throughout the leaf and canopy using the 3-D model FluorFLIGHT[38]. The model is based on existing theory of radiative transfer by coupling the leaf fluorescence model FLUSPECT[39] and the 3-D ray-tracing model FLIGHT[40,41] to account for the canopy components. Input data required to run the models are described in Supplementary Table 3. FluorFlight was used to i) estimate Fi independently from other confounding factors (LAI, C_{a+b}), and ii) to evaluate the Fi estimation from the O_2-A *in-filling* FLD method with a 6.4 nm FWHM sensor. We used FluorFLIGHT in a multi-step LUT-based inversion scheme[38] to retrieve full crown SIF and Fi from a complex scene accounting for the influence of scene structure and composition. Fi was quantified based on the FLD2 calculation from the airborne image using the LUT derived from FluorFLIGHT. As a prior step, we quantified the optimal parameter combination of N, C_{a+b}, C_{x+c} and LAI using PROSAIL[42,43]. The model was originally developed at 1 nm FWHM. For comparisons with the airborne hyperspectral imagery, we used model simulations convolved to 6.5 nm FWHM to match the spectral resolution of the radiance imagery acquired by the hyperspectral airborne sensor, evaluating the effects of the bandwidth on the Fi vs. SIF relationship (Supplementary Fig. 2).

Statistical analysis. We used multivariate analyses based on classification and machine learning algorithms to classify disease incidence and severity. We assessed the ability of various selections of spectral indices to estimate disease severity (DS) using support vector machine (SVM), neural networks (NN) and linear discriminant analysis (LDA). We tested these modelling approaches for three different objectives, assessing the separation between: (i) Case A: asymptomatic (AS) vs. symptomatic trees (AF; affected), and (ii) Case B: Initial *Xf*-symptoms (IN, DS=1) vs. advanced *Xf*-symptoms (AD, DS = 2, 3, and 4) severity levels. We validated the selected models by partitioning the data set into two samples: the training sample (TR), containing 80% of the data collected over two years (2016 and 2017) for each disease severity class selected at random, and the testing or validation sample (TS), with the remaining 20%. We fitted each model using the training sample and validated it by using the testing sample to assess its classification accuracy. In a first step, we performed a variable reduction based on variance inflation factor (VIF) analysis for each of the two objectives described (Cases A and B) on the training set. This was done to avoid multicollinearity among predictor variables (i.e. plant traits). The variables with a VIF lower than 10 were retained for model development. Variables used to build the different models evaluated were i) single reflectance bands, for operational purposes we assessed the 10 most sensitive wavelengths related to the disease; ii) spectral indices listed in Supplementary Table 1, with which we found the indices most sensitive to the disease to be NPQI, CWSI, PRI·CI, PRIn SIF, BF1, PRIM1, CRI700m, BF2, PRIM4, DCabxc,VOG2, and TCARI/OSAVI; and iii) plant traits estimated by model inversion (Fig. 2) using the radiative transfer models indicated above. Wilks' lambda method[44] was used to identify the variables with the greatest contribution. Then, we used the data retained through VIF analysis in the three classification methods (SVM, NN and LDA). We performed the SVM analysis using R software (version 3.4.0; R Development Core Team, Vienna, Austria) with the "e1071" package[45]. We applied a non-linear SVM classification method using the radial basis function kernel. We built the NN using the "nnet" package[46] in R, based on feed-forward networks with a single hidden layer. To reach the best performance of the NN, guaranteeing the maximization of its algorithm, we trained 500 NNs for each objective and selected the one with the highest classification accuracy. In addition, we set the NN parameter size, the number of units in the hidden layer, and the weight decay for the quantification of the penalty of misclassification errors using a cross-validation approach within the "caret" package[47] in R. We also conducted LDA using the "caret" package in R to generate a discriminant function capable of determining the classification accuracy of the dataset, based on the pooled covariance matrix and the prior probabilities of the classification groups[44]. We assessed the classification accuracies of three different sets of plant traits: 1) Pigment-, Structure-, Fluorescence and Temperature-based Functional Traits (PSFT); 2) Pigment- and Structure-based Functional Traits (PS); and 3) Standard RGB-NIR bandset (SVI) by calculating the overall accuracy (OA, in %) and the kappa coefficient (κ), which provides an overall accuracy assessment for the classification based on commission and omission errors for all classes[48].

We applied non-linear SVM classification models using the radial basis function with a leave one out cross validation (LOOCV) and a stochastic gradient boosting machine to test the remote sensing-based PSFT model with qPCR assay data obtained in: i) one field with trees affected by *Xf* and asymptomatic trees (n = 67 trees tested; total number of trees in the orchard = 157); and ii) trees tested with qPCR (n=100) located within eight olive orchards throughout the study area. Training of the SVM model was performed using an iterative procedure implemented with the "caret" package[47] in R. In a first step, balance techniques were performed to minimize unbalanced data effects; then, we conducted 50 iterations of non-linear SVM classification methods to predict the quantitative PCR data using the using the remote sensing-based PSFT model. In the next step, a sequential stochastic gradient boosting was trained using an ensemble model obtained from 50 SVM predictions. We fitted each non-linear SVM model and ensemble model to assess its classification accuracy. We assessed the classification accuracies of the proposed remote sensing SVM-PSFT disease detection model and the visual evaluation performed by plant pathologists against qPCR assay data obtained at the orchard level.

In October 2016, February, June, and July 2017 we revisited 1,700 out of the 3,328 trees evaluated in June 2016 to assess the potential of the remote sensing-based methods to detect trees affected by *Xf* before symptoms become visible. We selected the revisited plots to cover a wide range of initial disease incidence and severity values. The revisit study focused on calculating the confusion matrix for each model to predict disease severity for the trees evaluated in June and re-evaluated in October 2016. We used this confusion matrix to calculate the percentage of true negatives (TN, i.e. trees classified as asymptomatic by remote sensing and field assessment in June) and false positives (FP, i.e. trees classified as symptomatic by remote sensing but showing no visual symptoms in the field assessment in June) that developed symptoms in October. In total, the 1700 evaluated trees in the revisit consisted of 818 (TN), 412 (True Positives, TP), 178 (FP), and 292 (False Negatives, FN). The results for the studied cases (A, B) and all classification methods (SVM, NN, and LDA) are shown in Supplementary Tables 4 and 5, the results of the qPCR data analysis across eight orchards are shown in Supplementary Table 6, and the revisit study for the SVM method is displayed in Supplementary Table 7.

Acknowledgments. We thank Z.G. Cerovic, J.Flexas, F.Morales, and P.Martín for scientific discussions, QuantaLab-IAS-CSIC for laboratory assistance, and G.Altamura, A.Ceglie, and D.Tavano for field support. The study was funded by the European Union's Horizon 2020 research and innovation programme through grant agreements POnTE (635646) and XF-ACTORS (727987). The views expressed are purely those of the writers and may not in any circumstance be regarded as stating an official position of the European Commission.

Author contributions. P.J.Z.-T., C.C., P.S.A.B., B.B.L., D.B., M.S. and J.A.N.-C. designed research; P.J.Z.-T., C.C., P.S.A.B., R.C., A.H., R.H.-C., T.K., M.M.B., L.S., M.M., V.G.-D., P.R.J.N., B.B.L., D.B., M.S., and J.A.N.-C. performed research; P.J.Z.-T., C.C., P.S.A.B., R.C., A.H., R.H.-C., T.K., V.G.-D., and J.A.N.-C. analyzed data; and P.J.Z.-T., C.C., P.S.A.B. and J.A.N.-C. wrote the paper. All authors provided comments, read, and approved the final submission.

Data and code availability. The data and the custom code required for the analysis conducted in this study are available at the GitHub repository, address: https://github.com/Quantalab/Xf-NPlants-2018

References

1. Paini, D. R. *et al.* Global threat to agriculture from invasive species. *Proc. Natl. Acad. Sci.* **113,** 7575–7579 (2016).
2. Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E. & Leung, H. Plant-pathogen interactions: disease resistance in modern agriculture. *Trends Genet. TIG* **29,** 233–240 (2013).
3. Fisher, M. C. *et al.* Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484,** 186–194 (2012).
4. Flood, J. The importance of plant health to food security. *Food Secur.* **2,** 215–231 (2010).
5. Strange, R. N. & Scott, P. R. Plant disease: a threat to global food security. *Annu. Rev. Phytopathol.* **43,** 83–116 (2005).

6. Purcell, A. H. *Xylella fastidiosa*, a regional problem or a global threat ? *J. Plant Pathol.* **79,** 99–105 (1997).
7. Stokstad, E. Italy’s olives under siege. *Science* **348,** 620–620 (2015).
8. Almeida, R. P. P. Can Apulia’s olive trees be saved? *Science* **353,** 346–348 (2016).
9. EFSA. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2015. *EFSA J.* **14,** (2016).
10. Hopkins, D. L. & Purcell, A. H. *Xylella fastidiosa*: Cause of Pierce’s Disease of Grapevine and Other Emergent Diseases. *Plant Dis.* **86,** 1056–1066 (2002).
11. *Xylella fastidiosa* World distribution. EPPO Global Database. Available at: https://gd.eppo.int/taxon/XYLEFA/distribution. (Accessed: 5th March 2018)
12. Saponari, M. *et al.* Pilot project on Xylella fastidiosa to reduce risk assessment uncertainties. *EFSA Support. Publ.* **13,** n/a-n/a (2016).
13. Calderón, R., Navas-Cortés, J. A., Lucena, C. & Zarco-Tejada, P. J. High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. *Remote Sens. Environ.* **139,** 231–245 (2013).
14. Asner, G. P. *et al.* Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. *Science* **355,** 385–389 (2017).
15. 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* **100,** 1282–1288 (2010).
16. Gamon, J. A., Peñuelas, J. & Field, C. B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* **41,** 35–44 (1992).
17. Barnes, J. D., Balaguer, L., Manrique, E., Elvira, S. & Davison, A. W. A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants. *Environ. Exp. Bot.* **32,** 85–100 (1992).
18. Peñuelas, J., Filella, I., LLoret, P., Muñoz, F., Vilajeliu, M., Reflectance assessment of mite effects on apple trees. *Int.J.Remote Sen.* **16**, 2727-2733 (1995).
19. Lev-Yadun, S. & Gould, K. S. Role of Anthocyanins in Plant Defence. in *Anthocyanins* 22–28 (Springer, New York, NY, 2008). doi:10.1007/978-0-387-77335-3_2
20. De La Fuente L, Parker JK, Oliver JE, Granger S, Brannen PM, van Santen E, et al. (2013) The Bacterial Pathogen Xylella fastidiosa Affects the Leaf Ionome of Plant Hosts during Infection. PLoS ONE 8(5): e62945. https://doi.org/10.1371/journal.pone.0062945
21. Zeng, W., Melotto, M. & He, S. Y. Plant stomata: a checkpoint of host immunity and pathogen virulence. *Curr. Opin. Biotechnol.* **21,** 599–603 (2010).
22. Berger, S., Sinha, A. K. & Roitsch, T. Plant physiology meets phytopathology: plant primary metabolism and plant-pathogen interactions. *J. Exp. Bot.***58,**4019–4026 (2007).
23. Tung, J., Goodwin, P. H. & Hsiang, T. Chlorophyll fluorescence for quantification of fungal foliar infection and assessment of the effectiveness of an induced systemic resistance activator. *Eur. J. Plant Pathol.* **136,** 301–315 (2013).
24. Chaerle, L., Hagenbeek, D., De Bruyne, E., Valcke, R. & Van Der Straeten, D. Thermal and chlorophyll-fluorescence imaging distinguish plant-pathogen interactions at an early stage. *Plant Cell Physiol.* **45,** 887–896 (2004).
25. Barón, M., Pineda, M. & Pérez-Bueno, M. L. Picturing pathogen infection in plants. *Z. Naturforschung C J. Biosci.* **71,** 355–368 (2016).

26. Giampetruzzi, A. *et al.* Transcriptome profiling of two olive cultivars in response to infection by the CoDiRO strain of Xylella fastidiosa subsp. pauca. *BMC Genomics* **17,** 475 (2016).
27. Gueymard, C. *SMARTS2: a simple model of the atmospheric radiative transfer of sunshine: algorithms and performance assessment*. (Florida Solar Energy Center Cocoa, FL, 1995).
28. Gueymard, C. A. Parameterized transmittance model for direct beam and circumsolar spectral irradiance. *Sol. Energy* **71,** 325–346 (2001).
29. Calderón, R., Navas-Cortés, J.A., Zarco-Tejada, P.J. Early Detection and Quantification of Verticillium Wilt in Olive Using Hyperspectral and Thermal Imagery over Large Areas. *Remote Sens.* **7,** 5584–5610 (2015).
30. Niblack, W. *An Introduction to Digital Image Processing*. (Prentice-Hall, 1986).
31. Sauvola, J. & Pietikäinen, M. Adaptive document image binarization. *Pattern Recognit.* **33,** 225–236 (2000).
32. Hartig, S. M. Basic Image Analysis and Manipulation in ImageJ. in *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc., 2001). doi:10.1002/0471142727.mb1415s102
33. Féret, J.-B., Gitelson, A. A., Noble, S. D. & Jacquemoud, S. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sens. Environ.* **193,** 204–215 (2017).
34. Verhoef, W., Jia, L., Xiao, Q. & Su, Z. Unified Optical-Thermal Four-Stream Radiative Transfer Theory for Homogeneous Vegetation Canopies. *IEEE Trans. Geosci. Remote Sens.* **45,** 1808–1822 (2007).
35. Blackburn, G. A. Wavelet decomposition of hyperspectral data: a novel approach to quantifying pigment concentrations in vegetation. *Int. J. Remote Sens.* **28,** 2831–2855 (2007).
36. Blackburn, G. A. & Ferwerda, J. G. Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sens. Environ.* **112,** 1614–1632 (2008).
37. Banskota, A. *et al.* Investigating the Utility of Wavelet Transforms for Inverting a 3-D Radiative Transfer Model Using Hyperspectral Data to Retrieve Forest LAI. *Remote Sens.* **5,** 2639–2659 (2013).
38. Hernández-Clemente, R., North, P. R. J., Hornero, A. & Zarco-Tejada, P. J. Assessing the effects of forest health on sun-induced chlorophyll fluorescence using the FluorFLIGHT 3-D radiative transfer model to account for forest structure. *Remote Sens. Environ.* **193,** 165–179 (2017).
39. Vilfan, N., Tol, C., Muller, O., Rascher, U. & Verhoef, W. *Fluspect-B: A model for leaf fluorescence, reflectance and transmittance spectra*. **186,** (2016).
40. North, P. R. J. Three-dimensional forest light interaction model using a Monte Carlo method. *IEEE Trans. Geosci. Remote Sens.* **34,** 946–956 (1996).
41. North, P. R. J., Rosette, J. A. B., Suárez, J. C. & Los, S. O. A Monte Carlo radiative transfer model of satellite waveform LiDAR. *Int. J. Remote Sens.* **31,** 1343–1358 (2010).
42. Jacquemoud, S. & Baret, F. PROSPECT: A model of leaf optical properties spectra. *Remote Sens. Environ.* **34,** 75–91 (1990).
43. Verhoef, W. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sens. Environ.* **16,** 125–141 (1984).

44. Khattree, R. & Naik, D. N. *Multivariate Data Reduction and Discrimination with SAS Software*. (Wiley-SAS, 2000).
45. Meyer, D. *et al. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. (2017).
46. Ripley, B. & Venables, W. *nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models*. (2016).
47. Wing, M. K. C. from J. *et al. caret: Classification and Regression Training*. (2017).
48. Richards, J. A. *Remote Sensing Digital Image Analysis*. (1999).