scientific reports

OPEN

Environmental factors driving the abundance of *Philaenus spumarius* in mesomediterranean habitats of Corsica (France)

Marguerite Chartois[1,2✉], Xavier Mesmin[1,2], Ileana Quiquerez[3], Sabrina Borgomano[3], Pauline Farigoule[1,4], Éric Pierre[1], Jean-Marc Thuillier[1], Jean-Claude Streito[1], François Casabianca[5], Laetitia Hugot[3], Jean-Pierre Rossi[1,6], Jean-Yves Rasplus[1,6] & Astrid Cruaud[1,6]

Philaenus spumarius (*Ps*) is considered the main insect vector of the bacterium *Xylella fastidiosa* (*Xf*) in Europe. As such, it is a key actor of the *Xf* pathosystem on which surveillance and management strategies could be implemented. Although research effort has increased in the past years, the ecological factors shaping *Ps* abundance and distribution across landscapes are still poorly known in most regions of Europe. We selected 64 plots of 500m² in Corsican semi-natural habitats in which we sampled nymphs and adults of *Ps* during three years. While local or surrounding vegetation structure (low or high scrubland) had little effect on *Ps* abundance, we highlighted a positive relationship between *Ps* abundance and the density of *Cistus monspeliensis* in the plots. We also found larger populations of *Ps* in cooler and moister plots. The pattern of host association highlighted here is unique, which calls for more studies on the ecology of *Ps* in Europe, to help designing surveillance and management strategy for *Xf*.

The meadow spittlebug *Philaenus spumarius* (Linnaeus, 1758) (Hemiptera: Aphrophoridae) (*Ps*) is considered the main vector of the plant pathogen *Xylella fastidiosa (Xf)* in Europe[1,2]. Research effort on its biology and ecology has recently increased, generally with the aim of proposing management strategies. However, effort is unequal among regions and habitats and it is still unknown whether environmental drivers of abundance and distribution of *Ps* (e.g., climate and host plants) are similar between different ecosystems and/or geographical regions.

Humidity and temperature are important drivers of the development and the geographic distribution of *Ps*[2,3]. Overall, studies have either addressed physiological or phenological responses of *Ps* to temperature and humidity at local spatial scales, or climate preferences at larger scales[4], notably using species distribution models[5–7]. Regional studies investigating the effects of climatic conditions on populations of the meadow spittlebug are still scarce, with the exception of works by Drosopoulos and Asche[8] in Greece, Godefroid and Durán[9] in southern Spain, and Karban and Strauss[10] in California. Authors agree that humidity and temperature affect the development time of immature stages of *Ps*[3,11,12]. While exact values vary, cool and moist conditions appear more favorable, with minimum temperatures for egg hatching and larval development ranging from 4 to 10 °C[3,11,13]. No data are available regarding maximum temperatures but nymphal development is still observed at 27 °C[3,11]. Humidity promotes hatching and survival of nymphs that need sufficient sap flow to feed. Nymphs produce a foam that covers them and to which they owe the name "spittlebugs"[3]. This foam likely protects them from predators and dehydration[14]. In this regard, extreme or prolonged summer droughts drastically reduce *Ps* populations[10,15]. Moreover, shift in host plants from dry vegetation to less water-stressed shrubs and trees or migration toward coolest places (e.g., near rivers) have been documented, showing that humidity and temperature play a major role in the ecology of this insect[3,15–19].

Trophic interactions between insect vectors and their host plants can shape the epidemiology of plant diseases caused by *Xf*. In this regard, with more than 1000 host plant species in 75 families[3,20,21], the highly

[1]CBGP, INRAE, CIRAD, IRD, Institut Agro, Univ Montpellier, Montpellier, France. [2]AGAP, INRAE, CIRAD, Institut Agro, Univ Montpellier, San Giuliano, France. [3]Conservatoire Botanique National de Corse, Office de l'Environnement de la Corse, Corte, France. [4]AgroParisTech, Paris, France. [5]Centre INRAE de Corse, San Giuliano, France. [6]These authors jointly supervised this work: Jean-Pierre Rossi, Jean-Yves Rasplus and Astrid Cruaud. ✉email: marguerite.chartois@inrae.fr

polyphagous[2,3,22] *Ps* is a key actor of the *Xf*-pathosystem. Several studies have explored local host associations of *Ps* in olive groves in Italy[23,24], Portugal[25], Spain[19] and Greece[26], where nymphs mostly feed on Asteraceae, Fabaceae and Apiaceae. However, no study has explored the link between the local abundance of host plants and the degree of polyphagy of *Ps*. Summer migration of adults from ground cover to olive foliage, linked with drying of herbaceous plants has been observed in southern Italy[2,21]. In particular, adults of *Ps* were observed in the vicinity of olive groves, from May to August, on *Quercus ilex*, *Quercus x crenata*, *Pistacia lentiscus*, *P. terebinthus*, *Hedera helix* and *Myrtus communis*[17,21,24]. In the Apulia region, *Ps* adults have also been found on *Pinus halepensis* from May to September and have been reported once on *Ulmus minor* in October[17]. On the other hand, in Belgium, Hasbroucq et al.[27] reported adults of *Ps* on Ranunculaceae, *Prunus* sp., *Rubus* sp., *Crataegus* sp., *Alnus* sp., *Picea* sp., *Quercus* sp., *Salix* sp., *Carex* sp. and *Urtica* sp.. Recent studies performed in Corsica, France[5,28,29] suggest that *Ps* has affinity with *Cistus monspeliensis*, at least locally. Provided this holds on a larger scale, *Ps* populations are expected to grow with *C. monspeliensis* cover[30]. However, it should be evaluated whether insect load (i.e., insect population size per plant unit) will increase, decrease, or be uncorrelated with host plant cover[30,31]. The degree of preference for a plant species over its neighbors seems to play a key role in this relationship and is often negative (i.e., insect load decreases with plant density) when the focal plant species is preferred over its neighbors[31]. This phenomenon is known as "resource dilution" and given that it is usually observed in specialist species[32], it potentially does not contribute to determining the preference in a generalist insect such as *Ps*.

Another important aspect to consider within a vector-borne pathosystem is the dispersal ability of vectors. In flight mills, the average distance flown by *Ps* adults vary from one to five hundred meters[33,34]. In olive groves and meadow landscapes, the median distance covered during *Ps* adult life was assessed at 374 m and 507 m, respectively[35]. Although limited, this dispersal ability creates connection between habitats (e.g., grassland, meadow, forest, crops etc.) that are within a radius of some hundreds of meters. Consequently, *Ps* population size at a given site is likely to be influenced by the suitability of the surrounding habitat. Accordingly, population size of *Ps* has been shown to correlate positively with the surface of olive groves[36,37] in a 125 m-radius buffer surrounding studied plots, while it negatively correlated with the surface of vineyards[36]. *Ps* density was also found to be lower in sites with high forest cover[38]. Considering landscape composition around study sites to investigate drivers of *Ps* population size seems therefore crucial.

In the present study, we aimed at comparing *Ps* abundance in plots distributed along a gradient of climate conditions and characterized by contrasted vegetation structure, and various landscape contexts in mesomediterranean areas of Corsica. We addressed five research questions stemming from our review of the literature: (1) Are populations of *Ps* larger in areas densely covered with *C. monspeliensis*?; (2) Does *Ps* load on *C. monspeliensis* plant increase, decrease or remain steady with *C. monspeliensis* cover?; (3) Are populations of *Ps* larger in cooler and moister plots, especially during the hot and dry seasons?; (4) Does vegetation structure (tree cover versus open areas) impact the abundance of *Ps*, especially during the hot season? and (5) Does vegetation structure of the surrounding environment impact the abundance of *Ps*?.

Material and methods

Sampling plots. Sixty-four plots of 500 m² were distributed in different regions of Corsica (Fig. 1; Supplementary map) along two gradients: one gradient of *C. monspeliensis* cover (from 0 to 95%) and one elevation gradient (from 6 to 595 m above sea level). These two uncorrelated gradients (Spearman's rank correlation test p-value: 0.7586) allowed us to independently assess the effects of vegetation and climate on *Ps* abundance.

Sampling period and frequency. First, a phenological survey of populations of *Ps* was performed biweekly on two plots (CAN-01 and CAN-02, Table S1) to identify periods during which densities of nymphs and adults were the highest. Based on this phenological survey, field work was conducted on all sampling plots in April (highest density of nymphs); June (highest density of young adults) and October (breeding season). Conducting a field survey for a fortnight at peak abundance indeed contributed to minimize differences in abundance between plots. These three sampling sessions were repeated during three consecutive years (2018, 2019 and 2020).

Within-plot vegetation structure. All studied plots were located in the scrubland vegetation, a common vegetation type at the mesomediterranean level[39,40] (see botanic inventory in Table S1 for details). The percent cover of live *C. monspeliensis* was visually estimated on each plot at each sampling session. The median values over all sessions were used in subsequent analyses for all but two plots (VENT-03 and VENT-04) on which vegetation was cut by municipal services to prevent bushfires. For these two plots, *C. monspeliensis* covers measured at each session were used instead.

Vegetation structure within plot was retrieved from the OCS GE database (© IGN–2022, https://geoservices.ign.fr/ocsge) provided by the French National Institute of Geographic and Forestry Information. This database includes land cover and land use, photo-interpreted from 50 cm-pixels orthophotographs, in the form of a polygon database whose contours are positioned with a precision of *ca.* 1 m. Vegetation structure within plot was coded as follows: "high scrubland" (when tree cover exceeded 25%); number of plots n = 33, and "low scrubland" (shrub and herbaceous formations with tree cover below 25%; n = 31; Fig. 1; Table S1; Supplementary map).

Vegetation structure around plots. Vegetation structure around plots was also characterized through the OCS GE database. Based on *Ps* dispersal ability reported by Bodino et al.[35] in natural conditions, we calculated the proportion of each type of vegetation structure within a radius buffer zone of 374 m, which was drawn around the plot but excluding the plot itself (see Supplementary map). It is noteworthy that results were unchanged with a radius buffer zone of 500 m (data not shown). Four vegetation structure were identified: "high
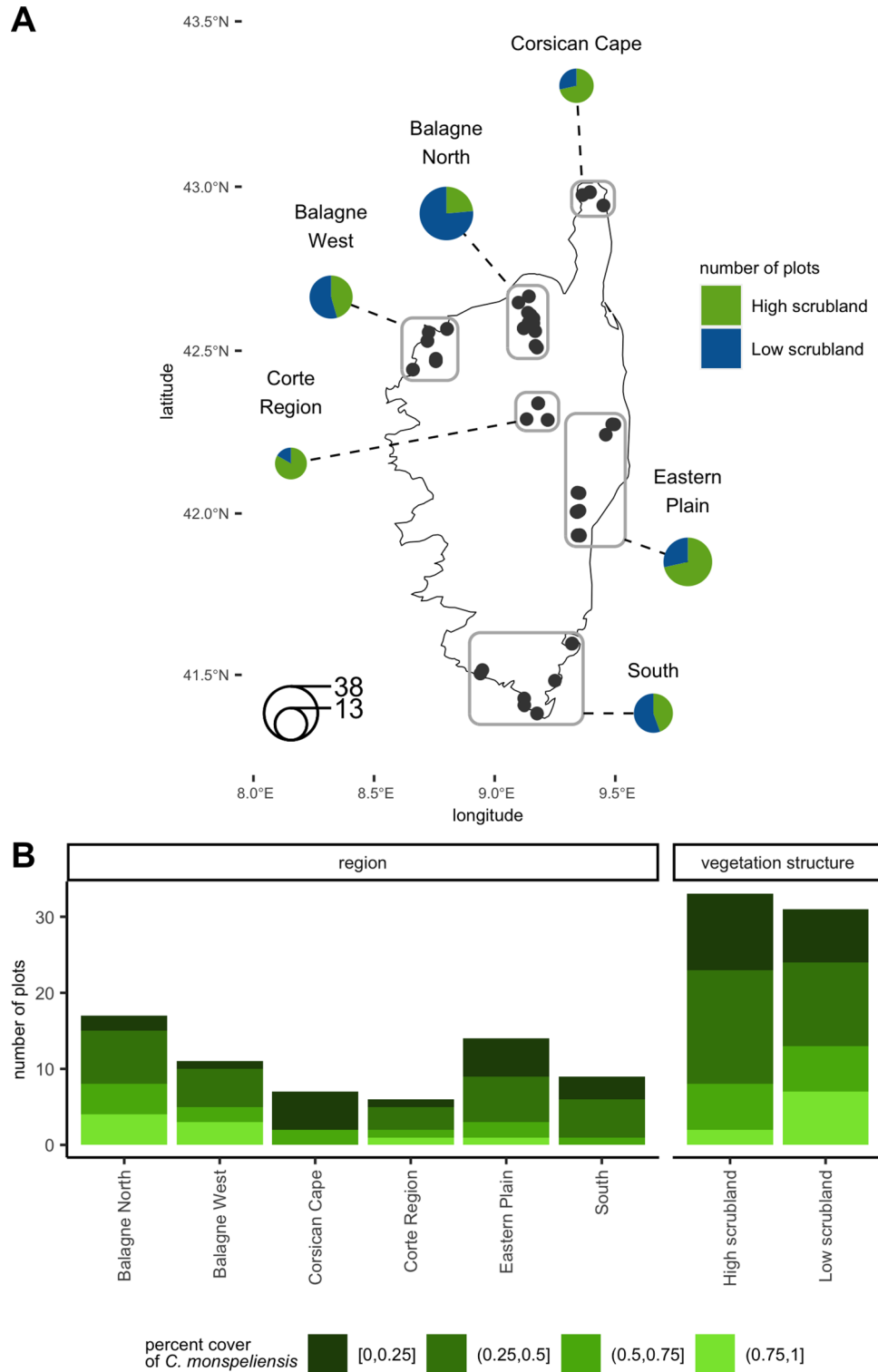
Figure 1. Sampling scheme. (**A**) Distribution of plots and vegetation structure. The size of the circles is proportional to the number of plots. (**B**) Percent cover of *C. monspeliensis* in each region and for each type of vegetation structure: high scrubland (tree cover exceeding 25%) and low scrubland (shrub and herbaceous formations with tree cover below 25%).

scrubland", "low scrubland" (see previous section), "unvegetated" (artificialized areas, bare ground, water and sea surfaces), and "orchards" (tree or woody vines cover exceeding 25% with agricultural use, i.e., mainly olive groves, *Citrus* groves and vineyards).

Vegetation structure around plots was then analyzed by a Principal Component Analysis (PCA). The proportion of each of the four vegetation structures within each buffer was used as input for the PCA (Table S1). The PCA was performed using the R package ade4[41]. Scores of each plot on the first (PC1$_{\text{landscape}}$) axis and second (PC2$_{\text{landscape}}$) axis of the PCA were used as input variables in all generalized linear mixed models (GLMMs).

Abundance of *P. spumarius*. In a preliminary experiment, we counted the number of nymphs in 20 randomly selected foams on each plot. An average of 1.24 nymphs per foam was found (± 0.57 standard deviation, Fig. S1) and this number did not differ significantly among plots (minimum p-value of Dunn test for multiple comparisons of means = 0.09, Fig. S1). Thus, we considered that the number of foams could be used as a good proxy for the number of nymphs and nymph's abundance was defined as the number of foams counted in 10 min in the vegetation of the plot (herbaceous, shrub, tree foliage of lowest branches, undergrowth) by the same operator.

Adult abundance was defined as the number of adults collected by two operators sweep netting the vegetation for 10 min, a time window sufficient to sample almost all the vegetation present in the plot (i.e., herbaceous, shrub, lowest branches of trees, and undergrowth vegetation). Sweep netting consisted in periods of 30 s of alternate backhand and forehand strokes with the net followed by 10–30 s of collection of adults with mouth aspirators. Sweeping the vegetation over a longer period would have increased risks of missing the target species in the net among plant debris and non-target arthropods. At any rate, after 10-min, adults collected in the mouth aspirators were counted and released afterward to preserve local insect populations.

To ascertain that foams/adults were correctly identified on sight, nymphs used to estimate the average number of nymphs per foam in the preliminary experiment and about 30 adults collected in the vicinity of each plot every June and October were killed using ethyl acetate, stored in 75° ethanol, and brought to the laboratory. We collected a total of 11,018 adults and 1911 nymphs. These specimens were identified to species level in the lab using a binocular microscope, the identification key of Biedermann and Niedringhaus[42]; illustrations from Stöckmann et al.[43] and the Arthemis database (https://arthemisdb.supagro.inrae.fr[44]). All nymphs (1911) and 11,003 adults (99.9%) were confirmed as *Ps*. The remaining 0.1% of adults (N = 15), were re-identified as *Neophilaenus campestris*. Given this negligible error rate, counts based on identification on sight in the field were considered as reliable.

Ecological drivers of *P. spumarius* abundance using bioclimatic data (GLMM1 models). A first set of GLMMs (hereafter referred to as GLMM1 models) was used to assess the effects of percent cover of *C. monspeliensis*, climate, plot vegetation structure and surrounding vegetation structure on *Ps* abundance. Three GLMMs were built, one for each sampling month: GLMM1-April (nymphs), GLMM1-June (young adults) and GLMM1-October (breeding adults).

In this set of GLMMs, climate was described by a PCA performed on 19 bioclimatic variables (Table S1)[45] retrieved from the SAFRAN model from Météo France[46], which interpolates data measured several times a day by a network of over 1000 meteorological stations spread over the French territory. SAFRAN provides daily data of temperature (2 m above ground) and precipitation interpolated at a resolution of 8 km. Bioclimatic variables were derived from the monthly minimum, maximum, mean temperature and mean precipitation values, and were computed over the whole sampling period (2018–2020). They represented annual trends, seasonality, and extreme values of annual climate conditions. As no strong message emerged from the literature about the bioclimatic factors driving *Ps* abundance, we used all 19 bioclimatic variables. Scores of each sampling plot on the first axis (PC1$_{\text{bioclim}}$) and second axis (PC2$_{\text{bioclim}}$) of the PCA (Table S1) were used as input variables for all three GLMM1 models. The 19 bioclimatic variables were computed for all plots using the function biovars of the R package dismo[47] to the three-year mean values of monthly temperatures (minimum and maximum) and precipitations (sum).

Structure of GLMM1 models, following glmmTMB notation was:

$$Ps \text{ abundance on month } m \sim \text{poly}(\%\text{cover of Cm}, \text{degree3}) + \text{plot scores on PC1}_{\text{bioclim}}$$
$$+ \text{ plot scores on PC2}_{\text{bioclim}} + \text{local vegetation structure} + \text{plot scores on PC1}_{\text{landscape}}$$
$$+ \text{ plot scores on PC2}_{\text{landscape}} + \text{year} + (1|\text{id}) + (1|\text{observation}), \text{family} = \text{nbinom1(link = "log")}$$

where Cm, (1|id) and (1|observation) respectively stand for *C. monspeliensis*, random effect on the identifier of the plot and observation-level random effect, and $m \in \{April, June, October\}$.

P. spumarius load per host plant unit *C. monspeliensis* cover (GLMM2 models). A second set of models (hereafter referred to as GLMM2 models; one model per sampling month) was built with the same structure as GLMM1 models but the response analyzed was the *Ps* load per unit cover of *C. monspeliensis*. To do that we used rate models[48], adding an offset—*offset*(log(%*cover of Cm*))—in the glmmTMB formula reported above. It should be noted that the response analyzed was only a proxy of *Ps* density on *C. monspeliensis*: we did not record the plant host of every *Ps* sampled. This proxy is thus slightly overestimating of the actual density, as some (but not much) individuals were collected on other host plants. Only significant variables of each GLMM1 models were included in corresponding GLMM2 models and only the 61 plots where *C. monspeliensis* was present were analyzed.

Ecological drivers of *P. spumarius* abundance using seasonal climate data (GLMM1bis models). A last set of GLMMs (hereafter referred to as GLMM1bis models) was used to check the robustness of the trends observed with GLMM1 models. They only differed from GLMM1 models in the way the effect of climate was tested. In GLMM1bis models (one for each sampling month), climate was described by a set of seasonal variables: average day temperatures (minimum, mean and maximum), sum and maximum day precipitations over the 1 or 2-month period preceding field surveys, all averaged over the 3 years of the study (10 climate descriptors, Table S1). A PCA was performed for each sampling month: April, June and October and scores of each plot on the first ($PC1_{seasonal}$) and second ($PC2_{seasonal}$) axis of the PCA were used for each corresponding GLMMs (GLMM1bis-April, GLMM1bis-June, GLMM1bis-October). This analysis is complementary to that of GLMM1 models because it relies on raw climate data (instead of elaborated bioclimatic variables) retrieved on a relevant period with respect to *Ps* phenology. The structure of GLMM1bis models was the same as for GLMM1 models. No significant differences between the results of GLMM1 and GLMM1bis models were found. Therefore, we did not use seasonal climate data to analyze *Ps* load per unit *C. monspeliensis* cover.

Modelling framework. All GLMMs were built with the R package glmmTMB[49,50] using linearly parameterized negative binomial distributions ("nbinom1"[51]). To account for repeated measures on the same plot, we added a random effect on the identifier of the plot[48]. An observation-level random effect was also added to correct overdispersion[52]. Year was included as an experimental design fixed effect because the number of factor levels was below 5[53]. As preliminary data analyses suggested a nonlinear relationship between *Ps* abundance and percent cover of *C. monspeliensis*, quadratic and cubic components of this variable were added to GLMMs. We used the DHARMa[54] and performance[55] packages to detect possible significant deviations from model assumptions (normality of residuals, homoscedasticity, the absence of collinearity between factors and the absence of highly influential data points by calculating Cook's distance). To explore the results of the fitted GLMMs, type II analyses of deviance (R package car[56]) and post-hoc pairwise comparisons of factor levels were performed (R package emmeans[57] and multcomp[58]).

Results

For all correlations mentioned in the text, p-values are below 0.005. Details on χ^2, df and exact p-values are available in Table S2.

Spatiotemporal variability of the abundance of *P. spumarius*. A total of 19,808 foams (used as a proxy for nymphs, see methods) and 18,645 adults of *Ps* were counted during our three-year survey. Spatial and temporal variability of *Ps* abundance is illustrated in Fig. 2 (raw data are provided in Table S1). No data was collected in April 2020 because of restrictions due to the Covid-19 pandemic.

Structure of vegetation surrounding sampling plots. PC1 and PC2 of the PCA performed on surrounding vegetation structure accounted for 50.8% and 30.7% of the variability, respectively (Fig. 3). $PC1_{landscape}$ represented a gradient of vegetation density and opposed plots located in high scrubland landscapes (tree cover above 25%; correlation to $PC1_{landscape}$: 0.997) to plots located in low scrubland landscapes (correlation to $PC1_{landscape}$: −0.945). The area covered by high scrubland in the landscape was therefore strongly negatively correlated with the area covered by low scrubland. $PC2_{landscape}$ opposed plots located in landscapes with large unvegetated areas (correlation to $PC2_{landscape}$: 0.794) to plots located in the vicinity of orchards (correlation to $PC2_{landscape}$: −0.756). Plots were first separated according to the scrubland structure (high or low) in their vicinity. Then, rarer elements of the landscape, unvegetated soils and orchards, were captured on $PC2_{landscape}$ which distinguished plots with more rocky soil and distributed on the coast (e.g., LUM-07, CALV-04, BONI-02) from plots located in the Balagne region where the landscape is historically rich in olive orchards (e.g., LAG-01, URT-13).

Ecological drivers of *P. spumarius* abundance using bioclimatic data (GLMM1 models). The transformation of the 19 bioclimatic variables through the PCA yielded two synthetic variables, PC1 and PC2, which supported 61.9% and 17.8% of the climate variability, respectively (Fig. 4). $PC1_{bioclim}$ opposed plots with high temperatures and low precipitations to plots with high precipitations and low temperatures. The variables exhibiting the largest scores upon $PC1_{bioclim}$ were the temperature of the warmest month and of the wettest quarter (right side of the axis, Fig. 4) and the temperature seasonality and the precipitation of the driest month (left side of the axis, Fig. 4). $PC2_{bioclim}$ opposed plots with high seasonal climate variations to plots with low seasonal climate variations. The bioclim variables that displayed the largest scores on $PC2_{bioclim}$ were the diurnal and annual temperature ranges (the "high seasonal variation" end of the gradient), and many equally projected precipitation-related variables in the positive direction (the "low seasonal variation" end of the gradient). Examples of raw climate values of the most contrasted plots according to the PCA are given in Table S3.

Models showed that *Ps* abundance was positively correlated with the percent cover of *C. monspeliensis* for each sampling month (Table S2). This correlation was characterized by a strong rise in *Ps* abundance for *C. monspeliensis* cover ranging from 0 to 25%, then abundance levelled off and finally rose again for *C. monspeliensis* cover exceeding 80% (Fig. 5A). Climate had a significant effect only on the abundance of adults (June and October). The abundance of adults decreased with sample plot scores on $PC1_{bioclim}$ (Fig. 5B). Therefore, *Ps* abundance increased with decreasing temperatures and increasing precipitations. Abundance of adults also increased with the scores of sample plots on $PC2_{bioclim}$, that is with decreasing seasonal contrast in climatic conditions (Fig. 5C). Local (i.e., in plot) and surrounding vegetation structure correlated with *Ps* abundance only in June. Plots with high scrubland vegetation hosted higher abundances of young adults than plots with low scrubland vegetation (Fig. 5D), but
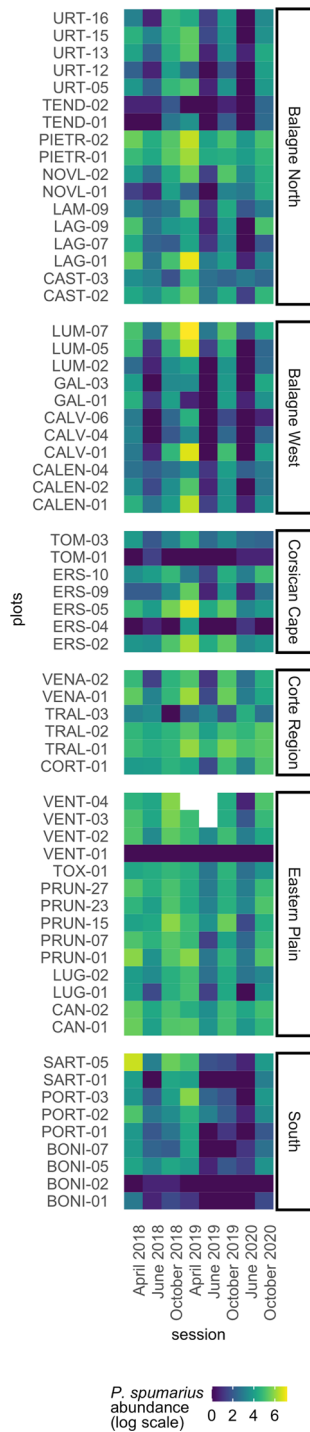
Figure 2. Abundance of *P. spumarius* and climate data. Spatiotemporal variability of *Ps* abundance (log scale). Rows show plots and columns show sampling sessions. White cells show NAs (vegetation cut).

vegetation structure of the plot had no effect on the abundance of nymphs or adults during the breeding season (Table S2). Abundance of young adults significantly increased with sample plot scores on PC1$_{landscape}$ (Fig. 5E), which means that abundances of *Ps* was higher when tree cover in surrounding vegetation was higher. Contrastingly, structure of landscape vegetation had no effect on the abundance of nymphs or adults during the breeding season (Table S2). In all years, adult abundance was higher in October than in June (Figs. 4, S2).

***P. spumarius* load per unit *C. monspeliensis* cover (GLMM2 models).** The effects of climate, local and landscape vegetation structures observed in GLMM1 models were identical to those observed with GLMM2
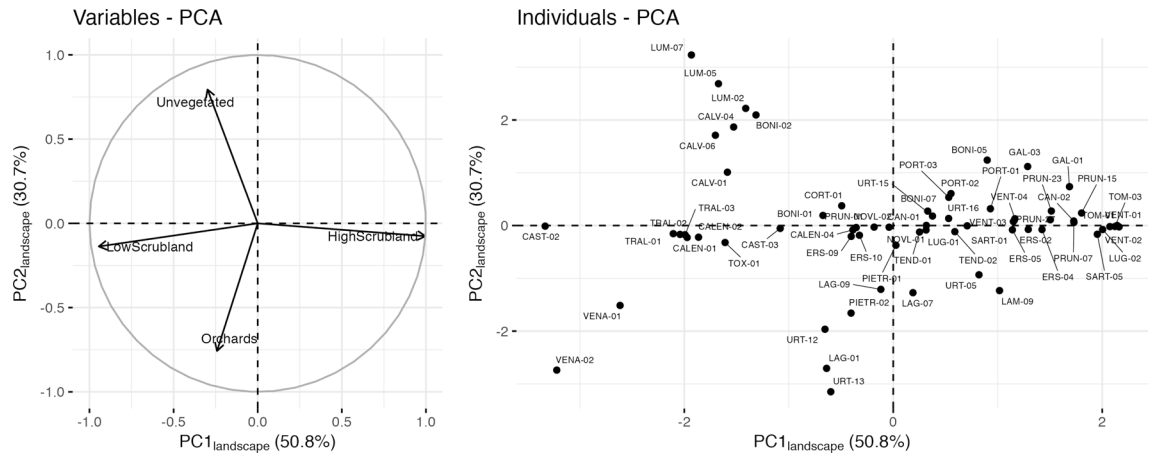
Figure 3. Principal component analysis of the proportion of each land cover type. (**A**) Correlation circle. (**B**) Projection of sampling plots in the first factorial plan.
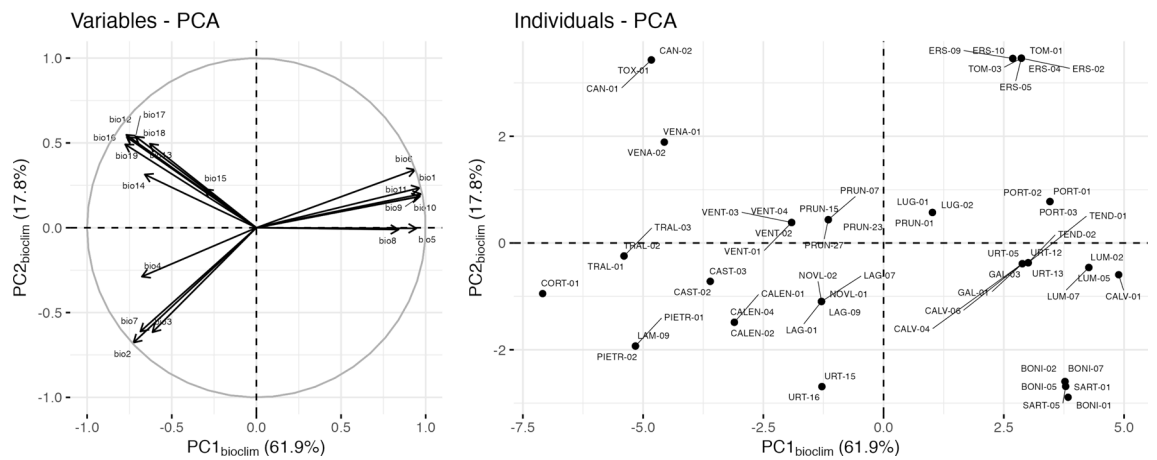

Figure 4. Principal component analysis of the 19 bioclimatic variables computed over the 2018–2020 period. (**A**) Correlation circle. Variable abbreviations are as follows: bio1 = annual mean temperature, bio2 = mean diurnal range, bio3 = isothermality (bio2/bio7) (×100), bio4 = temperature seasonality, bio5 = maximum temperature of warmest month, bio6 = minimum temperature of coldest month, bio7 = temperature annual range (bio5-bio6), bio8 = mean temperature of wettest quarter, bio9 = mean temperature of driest quarter, bio10 = mean temperature of warmest quarter, bio11 = mean temperature of coldest quarter, bio12 = annual precipitation, bio13 = precipitation of wettest month, bio14 = precipitation of driest month, bio15 = precipitation seasonality (coefficient of variation), bio16 = precipitation of wettest quarter, bio17 = precipitation of driest quarter, bio18 = precipitation of warmest quarter, bio19 = precipitation of coldest quarter. (**B**) Projection of sampling plots on the first factorial plan. Raw data are provided in Table S1.

models (Table S2). In addition, for all sampling months, *Ps* load per unit *C. monspeliensis* cover was negatively correlated with *C. monspeliensis* percent cover (Fig. 5F).

Ecological drivers of *P. spumarius* abundance using seasonal climate data (GLMM1bis models).
The PCA performed on the datasets associated to each date yielded two principal axes (PC1 and PC2) that accounted for 72.8% and 14.9% (April), 71.2% and 15.4% (June), and 62.3% and 31.8% (October) of the climate variability, respectively (Fig. S3). For the three PCA, PC1 opposed plots with high temperatures to plots with high precipitations and PC2 opposed plots with high temperatures and precipitations to plots with low temperatures and precipitations.

As in GLMM1 models, there was a positive correlation between *Ps* abundance and *C. monspeliensis* cover, with similar sigmoid pattern (Table S2; Fig. S4A). The abundance of adults increased with plot scores on PC1$_{seasonal}$ in June and October (Fig. S4B), that is with increasing precipitations and decreasing temperatures. Only in June, abundance of *Ps* decreased with plot scores on PC2$_{seasonal}$ (Fig. S4C), that is with decreasing precipitations and temperatures. Finally, neither local nor landscape vegetation structure had a significant effect on *Ps* abundance in GLMM1bis models.
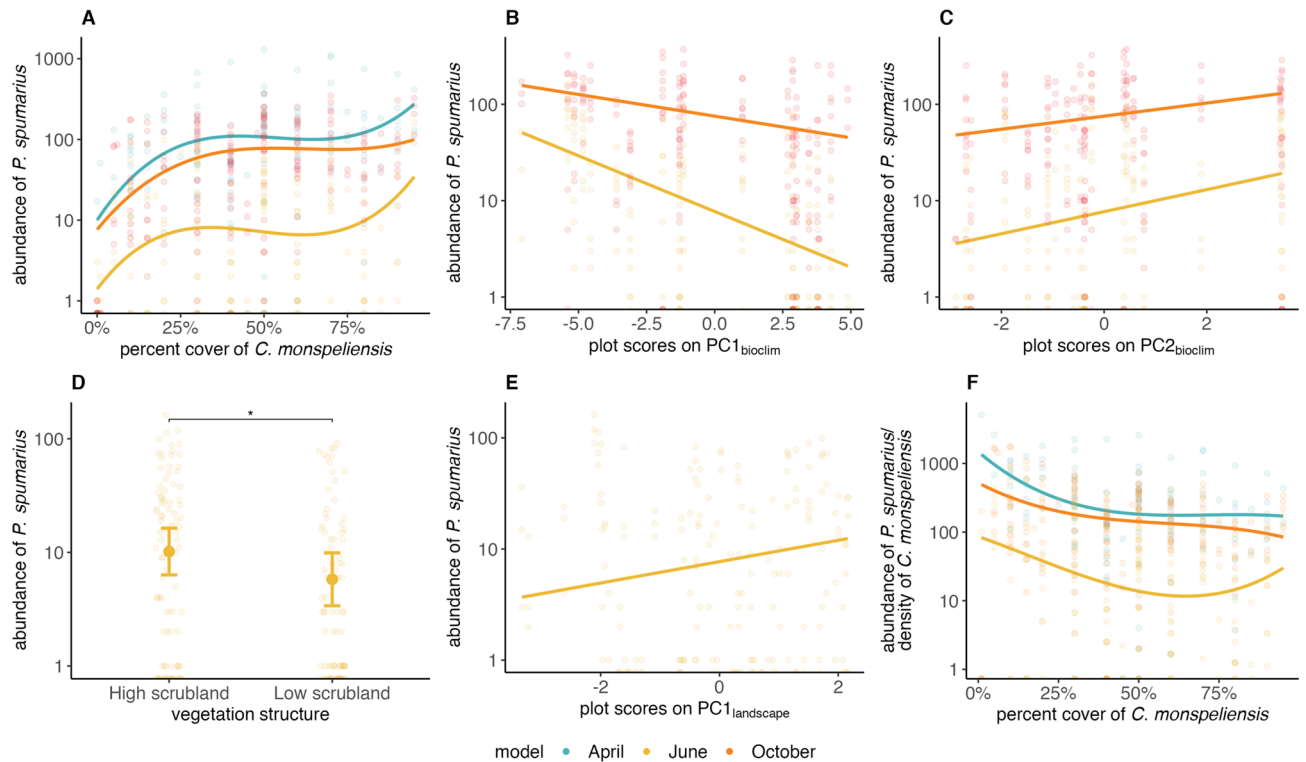
Figure 5. Scatterplots of raw data, regression lines and model prediction for pairwise associations between *P. spumarius* (*Ps*) abundance and fixed effects (GLMM1: panels (**A**) to (**E**); GLMM2: panel (**F**)). (**A**) *Ps* abundance and percent cover of *Cistus monspeliensis*. (**B**) *Ps* abundance and plot scores on PC1$_{bioclim}$ (which rises with increasing temperature and decreasing precipitation). (**C**) *Ps* abundance and plot scores on PC2$_{bioclim}$ (which rises with decreasing climate seasonality). (**D**) *Ps* abundance as a function of vegetation structure observed in the plot: high scrubland (tree cover exceeding 25%); low scrubland (shrub and herbaceous formations with tree cover below 25%). (**E**) *Ps* abundance and plot scores on PC1$_{landscape}$ (which opposed landscapes with tree cover above 25% (high scrublands) at positive scores to landscapes with open areas and tree cover below 25% (low scrubland) at negative scores). (**F**) *Ps* abundance per host plant and percent cover of *C. monspeliensis*. Colors: sampling month. Error bars on scatterplots: 95% confidence intervals. Points: raw data; lines: regression curves from models. All correlations have p-values < 0.005. All details on significance tests and R2 of the regression lines are given in Table S2.

Discussion

As for our first question, we found that *Ps* population size increased (although not linearly) with *C. monspeliensis* cover in mesomediterranean semi-natural habitats of Corsica. In addition, this relationship holds at both nymphal and adult stages, demonstrating that the affinity of *Ps* for *C. monspeliensis* goes beyond previously described local trends[28,29]. To answer the second question, we did not use the number of *Ps* found on each *C. monspeliensis* plant but added instead an offset to the model, which was equivalent to using the number of *Ps* divided by the area covered by *C. monspeliensis* in each plot as the output variable. Using this proxy, we found that *Ps* load on *C. monspeliensis* decreased with *C. monspeliensis* cover, especially from 0 to 50% plot cover. This pattern is expected when an insect has preference for one plant species over all its neighbors[31,32]. Insect load per *C. monspeliensis* plant was maximal when *C. monspeliensis* plants were scarce and disseminated in a matrix of other plant species. Then, when *C. monspeliensis* cover increases, *Ps* populations tend to dilute over the resource.

To our knowledge, Corsica is the only region where *Ps* population size shows such a pattern of strong association with *C. monspeliensis* density, although this plant is widely distributed across the Mediterranean region[59]. The availability of *C. monspeliensis* in mesomediterranean habitats of Corsica might primarily explain this unusual relationship given that *Ps* is highly polyphagous. The observed preference may be driven by retained chemical signal in generations of *P. spumarius* succeeding on *C. monspeliensis* as already observed in other taxa[60], and/or by specific communities of endosymbionts[61]. The question remains opened and more investigations are required to understand the determinisms of the association observed in Corsica (e.g., tests of host preferences in controlled conditions). In the light of this study and a previous one[28], assessing the consequences of pulling away *C. monspeliensis* from the direct vicinity of agricultural areas as a risk mitigation strategy against *Xf* appears relevant and so seems the release of biocontrol agents[62] near *C. monspeliensis* bushes.

C. monspeliensis-dense areas favoring larger populations of *Ps* and dilution over the resource may have contributed to the spread of *Xf* over Corsica[5,7]. Indeed, *C. monspeliensis* can be infected by *Xf*[63,64] and is very abundant in early stages of mesomediterranean vegetation series in Corsica[39]. It is noteworthy that *C. monspeliensis* falls well within the description of the "hidden reservoir" suggested by epidemiological modelling[65]. *C. monspeliensis*

being a pyrophyte[66] and an early colonizer[39], fire breaks and land clearings may have contributed to the spread of *Xf* in Corsica by increasing cover of *C. monspeliensis*. The strong association between *Ps* and *C. monspeliensis* in mesomediterranean habitats might contribute to explain why we did not observe any major outbreaks on plants sensitive to the strains of *Xf* ssp. *multiplex* present in Corsica so far[63].

Populations of adults of *Ps* were larger in cooler and moister habitats with lower seasonal fluctuations, whatever the sampling month, and this pattern was similar whatever the type of climate data used (i.e., bioclimatic variables or seasonal raw data). This result was expected from the physiological requirements of *Ps*[2,11,12] and from observations made in Italian grasslands, where larger populations of *Ps* are found in cooler sites[37]. This also fits with species modelling approaches, which predict low suitability for hot and dry areas in Mediterranean regions[5,6] and is consistent with results of Avosani et al.[38], who highlighted that the probability of observing *Ps* decreased when the average number of sunlight hours increased. It is noteworthy that the correlation between nymph abundance and either $PC1_{bioclim}$ or $PC1_{seasonal}$ was not significant, which could be the result of non-restrictive climatic conditions at sampling sites in April. The number of overwintering eggs is not well predicted by the number of gravid females, but depends primarily on weather conditions during the oviposition period[67]. Oviposition and winter survival of eggs add variability to population dynamics and may blur the climate effect found at adult stages.

Interestingly, the slope of the correlation with $PC1_{bioclim}$ and $PC1_{seasonal}$ was steeper in June than in October. Furthermore, local or surrounding vegetation structure had effect only on the abundance of young adults when bioclimatic data were used, with larger populations observed in plots with high scrubland vegetation and surrounded by high scrubland landscapes (tree cover > 25%). This pattern could be explained by the migration to summer hosts[17,29] outside of the sampling plot in hotter and drier low scrubland places and a microclimatic effect benefiting *Ps* in higher scrubland vegetation[9]. Apart from this situation, local and landscape vegetation structures had no effect on *Ps* populations and *C. monspeliensis* cover was much more important to explain *Ps* abundance than broad vegetation structure.

We believe that our results can help designing surveillance plans for *Xf* that are adapted to the specificities of Corsica. Surveillance may indeed target *C. monspeliensis*-rich habitats and use *Ps* sampled there as sentinel insects to track *Xf* in the environment[5,7]. As they highlight an unexpected and unique association pattern between *Ps* and *C. monspeliensis*, our results call for more studies on the ecology of *Xf* vectors in Europe, including common ones, to better understand, anticipate and control the spread of *Xf* in a climatically changing world. So far, only a few studies have investigated spatial distribution of vectors across European semi-natural landscapes[9,37,68]. In Mediterranean regions, studies principally focused on agroecosystems and targeted only a few species of potential vectors of *Xf*. They were primarily conducted in Italian olive groves, which is explained by the dramatic economic and socio-cultural consequences of *Xf* in this area[69]. A few were also performed in almond orchards, citrus orchards or vineyards in other Mediterranean areas[16,28,70]. Yet, only studies that attempt to decipher plant-vector trophic networks or investigate the main drivers of local abundance of vectors will help to better anticipate future outbreaks and provide clues on how to manage this bacterium with sustainable practices.

Data availability

The data sets generated and analyzed during the current study are available in Table S1.

References

1. Saponari, M. *et al.* Infectivity and transmission of *Xylella fastidiosa* by *Philaenus spumarius* (Hemiptera: Aphrophoridae) in Apulia, Italy. *J. Econ. Entomol.* **107**, 1316–1319. https://doi.org/10.1603/EC14142 (2014).
2. Cornara, D., Bosco, D. & Fereres, A. *Philaenus spumarius*: When an old acquaintance becomes a new threat to European agriculture. *J. Pest Sci.* **91**, 957–972. https://doi.org/10.1007/s10340-018-0966-0 (2018).
3. Weaver, C. R. & King, D. Meadow spittlebug, *Philaenus leucophthalmus* (L.). *Ohio Agric. Exp. Stn. Res. Bull.* **741**, 258 (1954).
4. Halkka, A., Halkka, L., Halkka, O., Roukka, K. & Pokki, J. Lagged effects of North Atlantic Oscillation on spittlebug *Philaenus spumarius* (Homoptera) abundance and survival. *Glob. Change Biol.* **12**, 2250–2262. https://doi.org/10.1111/j.1365-2486.2006.01266.x (2006).
5. Cruaud, A. *et al.* Using insects to detect, monitor and predict the distribution of *Xylella fastidiosa*: A case study in Corsica. *Sci. Rep.* **8**, 15628. https://doi.org/10.1038/s41598-018-33957-z (2018).
6. Godefroid, M. *et al.* Climate tolerances of *Philaenus spumarius* should be considered in risk assessment of disease outbreaks related to *Xylella fastidiosa*. *J. Pest Sci.* **2021**, 1–14. https://doi.org/10.1007/s10340-021-01413-z (2021).
7. Farigoule, P. *et al.* Vectors as sentinels: Rising temperatures increase the risk of *Xylella fastidiosa* outbreaks. *Biology* **11**, 1299. https://doi.org/10.3390/biology11091299 (2022).
8. Drosopoulos, S. & Asche, M. Biosystematic studies on the spittlebug genus *Philaenus* with the description of a new species. *Zool. J. Linn. Soc.* **101**, 169–177. https://doi.org/10.1111/j.1096-3642.1991.tb00891.x (1991).
9. Godefroid, M. & Durán, J. M. Composition of landscape impacts the distribution of the main vectors of *Xylella fastidiosa* in southern Spain. *J. Appl. Entomol.* **146**, 666–675. https://doi.org/10.1111/jen.13003 (2022).
10. Karban, R. & Strauss, S. Y. Physiological tolerance, climate change, and a northward range shift in the spittlebug, *Philaenus spumarius*. *Ecol. Entomol.* **29**, 251–254. https://doi.org/10.1111/j.1365-2311.2004.00576.x (2004).
11. Chmiel, S. M. & Wilson, M. C. Estimation of the lower and upper developmental threshold temperatures and duration of the nymphal stages of the meadow Spittlebug, *Philaenus spumarius*. *Environ. Entomol.* **8**, 682–685. https://doi.org/10.1093/ee/8.4.682 (1979).
12. Yurtsever, S. On the polymorphic meadow spittlebug, *Philaenus spumarius* (L.) (Homoptera: Cercopidae). *Turk. J. Zool.* **24**, 447–460 (2000).
13. Ahmed, D. D. & Davidson, R. H. Life history of the meadow spittlebug in Ohio. *J. Econ. Entomol.* **43**, 905–908. https://doi.org/10.1093/jee/43.6.905 (1950).
14. Whittaker, J. B. Cercopid spittle as a microhabitat. *Oikos* **21**, 59–64. https://doi.org/10.2307/3543839 (1970).

15. Drosopoulos, S. New data on the nature and origin of colour polymorphism in the spittlebug genus *Philaenus* (Hemiptera: Aphrophoridae). *Ann. Soc. Entomol. Fr. NS* **39**, 31–42. https://doi.org/10.1080/00379271.2003.10697360 (2003).

16. Bodino, N. *et al.* Phenology, seasonal abundance, and host-plant association of spittlebugs (Hemiptera: Aphrophoridae) in vineyards of Northwestern Italy. *Insects* **12**, 1012. https://doi.org/10.3390/insects12111012 (2021).

17. Cornara, D. *et al.* Natural areas as reservoir of candidate vectors of *Xylella fastidiosa*. *Bull. Insectol.* **74**, 173–180 (2021).

18. Gargani, E. *et al.* A five-year survey in Tuscany (Italy) and detection of *Xylella fastidiosa* subspecies multiplex in potential insect vectors, collected in Monte Argentario. *Redia* **104**, 75–88. https://doi.org/10.19263/REDIA-104.21.09 (2021).

19. Morente, M. *et al.* Distribution and relative abundance of insect vectors of *Xylella fastidiosa* in olive groves of the iberian peninsula. *Insects* **9**, 175. https://doi.org/10.3390/insects9040175 (2018).

20. Delong, D. *et al.* Spittle-insect vectors of Pierce's disease virus. I. Characters, distribution, and food plants. *Hilgardia* **19**, 339–356 (1950).

21. Bodino, N. *et al.* Phenology, seasonal abundance and stage-structure of spittlebug (Hemiptera: Aphrophoridae) populations in olive groves in Italy. *Sci. Rep.* **9**, 1–17. https://doi.org/10.1038/s41598-019-54279-8 (2019).

22. Wiegert, R. G. Population energetics of meadow spittlebugs (*Philaenus spumarius* L.) as affected by migration and habitat. *Ecol. Monogr.* **34**, 217–241. https://doi.org/10.2307/1948501 (1964).

23. Dongiovanni, C. *et al.* Plant selection and population trend of spittlebug immatures (Hemiptera: Aphrophoridae) in olive groves of the Apulia region of Italy. *J. Econ. Entomol.* **112**, 67–74. https://doi.org/10.1093/jee/toy289 (2019).

24. Bodino, N. *et al.* Spittlebugs of mediterranean olive groves: Host-plant exploitation throughout the year. *Insects* **11**, 130. https://doi.org/10.3390/insects11020130 (2020).

25. Villa, M., Rodrigues, I., Baptista, P., Fereres, A. & Pereira, J. A. Populations and host/non-host plants of spittlebugs nymphs in olive orchards from northeastern Portugal. *Insects* **11**, 720. https://doi.org/10.3390/insects11100720 (2020).

26. Antonatos, S. *et al.* Seasonal appearance, abundance, and host preference of *Philaenus spumarius* and *Neophilaenus campestris* (Hemiptera: Aphrophoridae) in olive groves in Greece. *Environ. Entomol.* **50**, 1474–1482. https://doi.org/10.1093/ee/nvab093 (2021).

27. Hasbroucq, S., Casarin, N., Ewelina, C., Bragard, C. & Grégoire, J.-C. Distribution, adult phenology and life history traits of potential insect vectors of *Xylella fastidiosa* in Belgium. *Belg. J. Entomol.* **92**, 2569 (2020).

28. Mesmin, X. *et al.* Interaction networks between spittlebugs and vegetation types in and around olive and clementine groves of Corsica; implications for the spread of *Xylella fastidiosa*. *Agric. Ecosyst. Environ.* **334**, 107979. https://doi.org/10.1016/j.agee.2022.107979 (2022).

29. Albre, J., García-Carrasco, J. M. & Gibernau, M. Ecology of the meadow spittlebug *Philaenus spumarius* in the Ajaccio region (Corsica)—I: Spring. *Bull. Entomol. Res.* **111**, 246–256. https://doi.org/10.1017/S0007485320000711 (2021).

30. Andersson, P., Löfstedt, C. & Hambäck, P. A. Insect density–plant density relationships: A modified view of insect responses to resource concentrations. *Oecologia* **173**, 1333–1344. https://doi.org/10.1007/s00442-013-2737-1 (2013).

31. Hambäck, P. A., Inouye, B. D., Andersson, P. & Underwood, N. Effects of plant neighborhoods on plant–herbivore interactions: Resource dilution and associational effects. *Ecology* **95**, 1370–1383. https://doi.org/10.1890/13-0793.1 (2014).

32. Otway, S. J., Hector, A. & Lawton, J. H. Resource dilution effects on specialist insect herbivores in a grassland biodiversity experiment. *J. Anim. Ecol.* **74**, 234–240 (2005).

33. Lago, C. *et al.* Flight performance and the factors affecting the flight behaviour of *Philaenus spumarius* the main vector of *Xylella fastidiosa* in Europe. *Sci. Rep.* **11**, 17608. https://doi.org/10.1038/s41598-021-96904-5 (2021).

34. Casarin, N. *et al.* Investigating dispersal abilities of Aphrophoridae in European temperate regions to assess the threat of potential *Xylella fastidiosa*-based pathosystems. *J. Pest Sci.* https://doi.org/10.1007/s10340-022-01562-9 (2022).

35. Bodino, N. *et al.* Dispersal of *Philaenus spumarius* (Hemiptera: Aphrophoridae), a vector of *Xylella fastidiosa*, in olive grove and meadow agroecosystems. *Environ. Entomol.* **50**, 267–279. https://doi.org/10.1093/ee/nvaa140 (2020).

36. Santoiemma, G., Tamburini, G., Sanna, F., Mori, N. & Marini, L. Landscape composition predicts the distribution of *Philaenus spumarius*, vector of *Xylella fastidiosa*, in olive groves. *J. Pest. Sci.* **92**, 1101–1109. https://doi.org/10.1007/s10340-019-01095-8 (2019).

37. Cappellari, A. *et al.* Spatio-temporal dynamics of vectors of *Xylella fastidiosa* subsp. pauca across heterogeneous landscapes. *Entomol. Gen.* **42**, 515–521. https://doi.org/10.1127/entomologia/2022/1427 (2022).

38. Avosani, S., Tattoni, C., Mazzoni, V. & Ciolli, M. Occupancy and detection of agricultural threats: The case of *Philaenus spumarius*, European vector of *Xylella fastidiosa*. *Agric. Ecosyst. Environ.* **324**, 107707. https://doi.org/10.1016/j.agee.2021.107707 (2022).

39. Allier, C. & Lacoste, A. Processus dynamiques de reconstitution dans la série du *Quercus ilex* en Corse. In *Vegetation Dynamics in Grasslans, Healthlands and Mediterranean Ligneous Formations* 83–91 (Springer, 1981).

40. Delbosc, P., Bioret, F. & Panaïotis, C. *Plant landscape of Corsica: Typology and mapping plant landscape of Cap Corse region and Biguglia Pond* (Springer Nature, 2020).

41. Chessel, D., Dufour, A.-B. & Thioulouse, J. The ade4 package—I: One-table methods. *R. News* **4**, 5–10 (2004).

42. Biedermann, R. & Niedringhaus, R. *The Plant-and Leafhoppers of Germany: Identification Key to All Species* (Wabv Fründ, 2009).

43. Stöckmann, M., Biedermann, R., Nickel, H. & Niedringhaus, R. *The Nymphs of the Planthoppers and Leafhoppers of Germany* (WABV, 2013).

44. INRAE-CBGP. *Arthemis DB@se - ARTHropod Ecology, Molecular Identification and Systematics*. https://arthemisdb.supagro.inrae.fr https://doi.org/10.15454/TBGRIB. Accessed 2021.

45. Xu, T. & Hutchinson, M. ANUCLIM version 6.1 user guide. *Aust. Natl. Univ. Fenner Sch. Environ. Soc. Canberra* **2011**, 256 (2011).

46. Quintana-Seguí, P. *et al.* Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France. *J. Appl. Meteorol. Climatol.* **47**, 92–107. https://doi.org/10.1175/2007JAMC1636.1 (2008).

47. Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. *Dismo: Species Distribution Modeling* https://CRAN.R-project.org/package=dismo (2017).

48. Faraway, J. J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models* (Chapman and Hall/CRC, 2006).

49. Brooks, M. E. *et al.* glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R. J.* **9**, 378–400. https://doi.org/10.3929/ethz-b-000240890 (2017).

50. R Core Team. *R: A Language and Environment for Statistical Computing* https://www.R-project.org/ (2019).

51. Hardin, J. W. & Hilbe, J. M. *Generalized Linear Models and Extensions* 4th edn. (Stata Press, 2018).

52. Harrison, X. A. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* **2**, e616. https://doi.org/10.7717/peerj.616 (2014).

53. Bolker, B. M. *et al.* Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**, 127–135. https://doi.org/10.1016/j.tree.2008.10.008 (2009).

54. Hartig, F. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models* https://CRAN.R-project.org/package=DHARMa (2020).

55. Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P. & Makowski, D. performance: An R package for assessment, comparison and testing of statistical models. *J. Open Sourc. Softw.* **6**, 3139. https://doi.org/10.21105/joss.03139 (2021).

56. Fox, J. & Weisberg, S. *An {R} Companion to Applied Regression* Third edn, https://socialsciences.mcmaster.ca/jfox/Books/Companion/ (Sage, Thousand Oaks CA, 2019).

57. Lenth, R. V. *Emmeans: Estimated Marginal Means, Aka Least-Squares Means* https://CRAN.R-project.org/package=emmeans (2021).
58. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363. https://doi.org/10.1002/bimj.200810425 (2008).
59. Fernández-Mazuecos, M. & Vargas, P. Ecological rather than geographical isolation dominates Quaternary formation of Mediterranean *Cistus* species. *Mol. Ecol.* **19**, 1381–1395. https://doi.org/10.1111/j.1365-294X.2010.04549.x (2010).
60. Berenbaum, M. R. & Feeny, P. P. 1. Chemical mediation of host-plant specialization: The papilionid paradigm. In *Specialization, Speciation, and Radiation* (ed. Tilmon, K.) 3–19 (University of California Press, 2008). https://doi.org/10.1525/california/9780520251328.003.0001.
61. Kapantaidaki, D. E., Antonatos, S., Evangelou, V., Papachristos, D. P. & Milonas, P. Genetic and endosymbiotic diversity of Greek populations of *Philaenus spumarius*, *Philaenus signatus* and *Neophilaenus campestris*, vectors of *Xylella fastidiosa*. *Sci. Rep.* **11**, 3752. https://doi.org/10.1038/s41598-021-83109-z (2021).
62. Mesmin, X. *et al.* *Ooctonus vulgatus* (Hymenoptera, Mymaridae), a potential biocontrol agent to reduce populations of *Philaenus spumarius* (Hemiptera, Aphrophoridae) the main vector of *Xylella fastidiosa* in Europe. *PeerJ* **8**, e8591. https://doi.org/10.7717/peerj.8591 (2020).
63. Denancé, N. *et al.* Several subspecies and sequence types are associated with the emergence of *Xylella fastidiosa* in natural settings in France. *Plant Pathol.* **66**, 1054–1064. https://doi.org/10.1111/ppa.12695 (2017).
64. EFSA, Delbianco, A., Gibin, D., Pasinato, L. & Morelli, M. Update of the Xylella spp host plant database—systematic literature search up to 31 December 2020. *EFSA J.* **19**, 6. https://doi.org/10.2903/j.efsa.2021.6674 (2021).
65. 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).
66. Roy, J. & Sonié, L. Germination and population dynamics of *Cistus* species in relation to fire. *J. Appl. Ecol.* **29**, 647–655. https://doi.org/10.2307/2404472 (1992).
67. Whittaker, J. B. Density regulation in a population of *Philaenus spumarius* (L.) (Homoptera: Cercopidae). *J. Anim. Ecol.* **42**, 163–172. https://doi.org/10.2307/3410 (1973).
68. Chapman, D. *et al.* Improving knowledge of *Xylella fastidiosa* vector ecology: modelling vector occurrence and abundance in the wider landscape in Scotland. Project Final Report. PHC2020/04, Scotland's Centre of Expertise for Plant Health (PHC) https://doi.org/10.5281/zenodo.6523478 (2022).
69. Saponari, M., Giampetruzzi, A., Loconsole, G., Boscia, D. & Saldarelli, P. *Xylella fastidiosa* in olive in Apulia: Where we stand. *Phytopathology* **109**, 175–186. https://doi.org/10.1094/PHYTO-08-18-0319-FI (2019).
70. López-Mercadal, J. *et al.* Collection of data and information in Balearic Islands on biology of vectors and potential vectors of *Xylella fastidiosa* (GP/EFSA/ALPHA/017/01). *EFSA Supp. Publ.* **18**, 10. https://doi.org/10.2903/sp.efsa.2021.EN-6925 (2021).

Acknowledgements

Author contributions

A.C. and J.Y.R. designed research. A.C., J.Y.R., L.H. and F.C. provided funding. M.C., X.M., I.Q., S.B., P.F., E.P., J.M.T., J.S.C., J.Y.R. and A.C. conducted experiments. M.C., X.M. and J.P.R. analyzed data. M.C., X.M., J.P.R. and A.C. drafted the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-28601-4.

Correspondence and requests for materials should be addressed to M.C.

Reprints and permissions information is available at www.nature.com/reprints.

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