scientific reports

OPEN

Assessment of potential invasion for six phytophagous quarantine pests in Taiwan

Hsin-Ting Yeh[1], Harn-Yeu Cheah[2,4], Ming-Chih Chiu[3✉], Jhih-Rong Liao[4✉] & Chiun-Cheng Ko[4,5]

Pest risk assessment is typically performed by expert taxonomists using a pest's biological data. However, the biological data or expert taxonomists may be difficult to find. Here, we used species distribution modelling to predict potential invasion in which phytophagous quarantine pests survive in Taiwan; the pests (unrecorded yet in Taiwan) included were three notorious quarantine whiteflies (*Crenidorsum aroidephagus*, *Aleurothrixus trachoides*, and *Paraleyrodes minei*) and three aphids (*Nasonovia ribisnigri*, *Macrosiphum euphorbiae*, and *Viteus vitifoliae*). In brief, maximum entropy modelling (MaxEnt) was used to predict the suitability of the pests' habitats under certain climatic conditions, and then receiver operating characteristic curve analysis was performed (to verify the prediction result). We then analysed environmental variables affecting the habitat suitability and matched them with Taiwan's crop cultivation areas for the assessment of potential invasion. We observed that the habitat suitability of the cultivation areas of host plants was low for *C. aroidephagus*, *A. trachoides*, and *N. ribisnigri* but was high for the remaining three species. Moreover, precipitation of coldest quarter negatively affected habitat suitability for *C. aroidephagus*, *P. minei*, *N. ribisnigri*, and *M. euphorbiae.* Seasonal temperature changes also negatively affected the habitat suitability for *A. trachoides*. This is the first study to demonstrate the use of species distribution modelling as the preliminary step for the pest risk assessment of these emerging pests with limited biological data before their invasion.

Biological invasion is a serious global economic and ecological threat[1,2]. Invasive organisms negatively influence native fauna and flora through multiple interactions, such as predation or competition[3,4]. Multiple human activities (e.g., international trade, travel) have promoted the movement of alien species across geographical barriers[5]. Therefore, interruption of pest invasion by plant quarantine is crucial topic related to biological invasion. For the management of alien species invasion, pest risk assessment can be used to evaluate the probability of invasion and spread of a pest. It should entail the use of scientific methods to determine whether government intervention is necessary in the prevention of pest invasion, and can also be used to assess the potential effects of an invasive species. Pest risk assessment includes identification of transmission route and biological characteristics and review of previous risk analysis results, and records of invasion of other countries[6]. It is typically performed by expert taxonomists. However, due to the wide abundance of insects, not every insect's biological data is available. Even experienced experts sometimes have difficulty obtaining sufficient data for risk assessment. Therefore, this study used species distribution modelling (SDM) to assess potential invasions; this modelling approach is useful under the condition of limited biological data and provides greater insight into the potential distribution (or habitat suitability) than specialist knowledge alone can.

SDM, also known as ecological niche modelling refers to mathematical statistics or prediction modules based on the ecological niche theory[7]. SDM helps to understand the species niche or environmental space on the basis of known species' distribution data and environmental characteristics[8]. It can also be used to predict habitat suitability before invasion by alien species[9], and also aids in early monitoring, warning generation, and emergency control measure adoption. Maximum entropy modelling (MaxEnt) is a machine learning method based on presence-only data, devised by Phillips et al.[8]. This method is widely used because it requires a small sample size and provides high prediction accuracy and its software easy to use[10,11]. The model's predictive ability and the use

---

[1]The Experimental Forest, College of Bio-Resources and Agriculture, National Taiwan University, Zhushan Township, Nantou County 557004, Taiwan. [2]Master Program for Plant Medicine, College of Bio-Resources and Agriculture, National Taiwan University, Taipei City 106332, Taiwan. [3]Center for Marine Environmental Studies (CMES), Ehime University, Matsuyama, Ehime 7908577, Japan. [4]Department of Entomology, National Taiwan University, Taipei City 106332, Taiwan. [5]Chiun-Cheng Ko is deceased. ✉email: mingchih.chiu@gmail.com; k1107053@hotmail.com

of normalisation prevent over-fitting during predictive simulation[8,12]. Zhang et al.[13] analysed potential areas for invasion by the pest *Viteus vitifoliae* (Fitch) in China using MaxEnt and geographic information system (GIS) and proposed immediate preventive measures. Prabhulinga et al.[14] predicted the habitat suitability of *Bemisia tabaci* (Gennadius) in North India using MaxEnt. MaxEnt is a crucial tool not only for understanding relationships between biological habitat and the related environmental factors but also for managing invasive alien species[9].

Whiteflies (superfamily Aleyrodoidea) and aphids (superfamily Aphidoidea) both belong to the order Hemiptera, suborder Sternorrhyncha. Thus far, more than 1,660 known whiteflies[15,16] and nearly 5,000 known aphids have been recorded worldwide[17–19]. Both small phytophagous insects with piercing and sucking mouthparts; certain species can transmit plant disease and become notorious agricultural and forestry pests. International trade, tourism, and business activities have increased year by year. The invasion by alien pests is greatly raised by man-made activities (e.g., importing agricultural products, mail packages, passenger luggage) and even affects agricultural production and the ecological environment[5]. For example, the sweet potato whitefly, *Bemisia tabaci* (MAMAI, b-biotype) invaded Taiwan in the 1990s and brought huge economic losses[20].

The present study aimed to establish a model for predicting the habitat suitability of different areas for quarantine pests before their invasion. The study results can be used by the government as a reference to formulate strategies for pest quarantine and monitoring. According to plant quarantine records in Taiwan (data obtained from the Bureau of Animal and Plant Health Inspection and Quarantine, Council of Agriculture, Executive Yuan) and the pest list of Centre for Agriculture and Bioscience International (CABI) and Invasive Species Compendium, six prominent quarantine pests have not yet been recorded in Taiwan, namely three whiteflies [*Crenidorsum aroidephagus* Martin & Aguiar, *Aleurothrixus trachoides* (Back), and *Paraleyrodes minei* Iaccarino] and three aphid species [*Nasonovia ribisnigri* (Mosley), *Macrosiphum euphorbiae* (Thomas) and *V. vitifoliae*]. MaxEnt was used to predict habitat suitability under certain climatic conditions for these six quarantine pests based on distribution points (Fig. 1, Supplementary Tables S1–S6), which were then matched with the information on host plant cultivation areas. The result is a combination of climatic factors and host plant information for improving the credibility of risk assessment for potential invasions.

Results

Six quarantine pests predicted by MaxEnt in potential distribution areas.
Figure 2 shows the omission rate of the predictions for the six species. The training set omission rate (blue line) should be close to the predicted omission rate (black line). The average training set omission rates for *C. aroidephagus*, *A. trachoides*, *P. minei*, *N. ribisnigri*, *M. euphorbiae*, and *V. vitifoliae* were 0.000, 0.000, 0.000, 0.015, 0.000, 0.000, respectively (all < 0.10). The areas under the receiver operating characteristic (ROC) curve (AUCs; output from the MaxEnt) of the training set for *C. aroidephagus*, *A. trachoides*, *P. minei*, *N. ribisnigri*, *M. euphorbiae*, and *V. vitifoliae* were 0.913, 0.921, 0.934, 0.896, 0.853, 0.841, respectively (Fig. 3). According to the AUC evaluation, this prediction approach achieved "moderate to high accuracy"[13,21–23].

Comparisons of habitat suitability of important quarantine pests and crop cultivation areas.
Comparisons of habitat suitability of six important quarantine pests in Taiwan and their host plant cultivation areas are shown in Figs. 4, 5, 6, 7, 8, and 9. Green represents high habitat suitability, red represents moderate habitat suitability, and white represents low habitat suitability. The potential distribution areas of *C. aroidephagus* are in the mid- and high-altitude mountains of Taiwan and in northern and eastern Taiwan but not in western and southwestern Taiwan. In addition, two outlying islands (Ludao and Lanyu) are also considered as highly suitable areas. The host plant *Anthurium andraeanum* of *C. aroidephagus* is mainly cultivated in western Taiwan from most plain to shallow mountainous areas (Fig. 4). The potential distribution areas of *A. trachoides* are in high-altitude mountains of Taiwan, southeastern Taiwan and in the two outlying islands of Ludao and Lanyu. The host plants potato (*Solanum tuberosum*), tomato (*S. lycopersicum*), and sweet potato (*Ipomoea batatas*) of *A. trachoides* are cultivated in most plain areas in Taiwan, particularly in western Taiwan. In addition, these crops are also cultivated in the outlying islands Penghu, Ludao, and Lanyu (Fig. 5). The potential distribution areas of *P. minei* are in most parts of Taiwan and in Ludao and Lanyu, with the exceptions being high-altitude mountainous areas and southwestern Taiwan. The host plants guava (*Psidium guajava*), avocado (*Persea americana*), and citrus (*Citrus* spp.) of *P. minei* are cultivated in plain areas in Taiwan, especially western Taiwan, and in Penghu (Fig. 6). The potential distribution areas of *N. ribisnigri* include the areas around high-altitude mountain in Taiwan. In addition, southwestern Taiwan and the mountainous region of eastern Taiwan had low habitat suitability. The host plant lettuce (*Lactuca sativa*) of *N. ribisnigri* is cultivated in most plain to shallow mountainous areas in Taiwan and Penghu (Fig. 7). The potential distribution areas of *M. euphorbiae* are in most parts of Taiwan, particularly in low- and mid-altitude mountainous areas, but not in high-altitude mountainous areas. In addition, the plain areas of northern and southwestern Taiwan are considered to have low suitability. The host plants of *M. euphorbiae* are potato, tomato and sweet potato, and the main cultivation areas are in the eastern and western plains to the shallow mountains of Taiwan (Fig. 8). The potential distribution for *V. vitifoliae* are mainly in the mid- to high-altitude mountainous areas of Taiwan and northern Taiwan, but not in southwestern Taiwan. The host plant of *V. vitifoliae* is mainly cultivated in the shallow mountainous areas to mid-altitude mountainous areas of western Taiwan, and a small part is in the southeastern region of Taiwan (Fig. 9). In addition, a histogram of habitat suitability of crop locations in six important quarantine species were shown in Supplementary Fig. S2. For example, less than 50% crop locations have a habitat suitability of > 0.5 for *C. aroidephagus*, *A. trachoides*, and *N. ribisnigri*; whereas more than 50% of crop locations have a habitat suitability of > 0.5 for *P. minei*, *M. euphrobiae*, and *V. vitifoliae*.
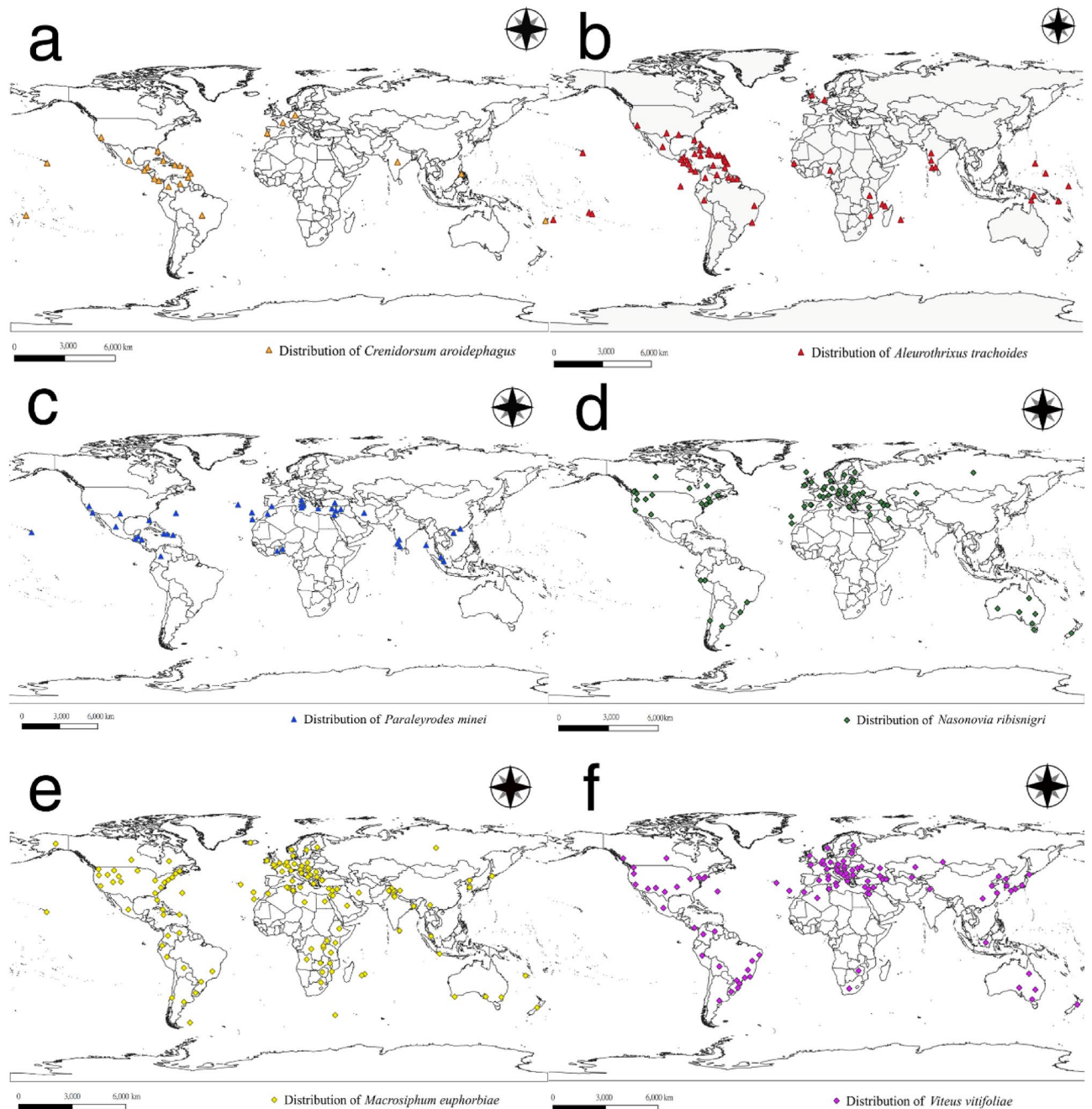
Figure 1. Distribution data of six important quarantine species: (**a**) *Crenidorsum aroidephagus*, (**b**) *Aleurothrixus trachoides*, (**c**) *Paraleyrodes minei*, (**d**) *Nasonovia ribisnigri*, (**e**) *Macrosiphum euphorbiae*, (**f**) *Viteus vitifoliae*. Triangle represent three whiteflies' distribution points; diamond shapes represent three aphids' distribution points. Maps were created using QGIS version 3.12.1 (www.qgis.org).

Environmental variables influencing the habitat suitability of six important quarantine species.

Figure 10 and Table 1 present the environmental variables and their permutation importance, as determined in a jackknife test. In particular, precipitation of coldest quarter (Bio19), temperature seasonality (Bio04), and isothermality (Bio03) were the most influential factors on the prediction of habitat suitability for *C. aroidephagus*, and their total contribution was 74.6%. By contrast, precipitation of driest month (Bio14) had little influence on the results, with a total contribution of 3%. Bio04, Bio14 and Bio03 were the three most influential factors on the prediction of habitat suitability for *A. trachoides*, and their total contribution was 74.4%. By contrast, temperature annual range (Bio07) and Bio19 had little influence on the results, with a total contribution of 9.2%. Bio19 and mean temperature of driest quarter (Bio09) were the most influential factors on the prediction of habitat suitability for *P. minei*, and their total contribution was 73.7%. By contrast, mean diurnal range (Bio02), and precipitation seasonality (Bio15) had little influence on the results, with a total contribution of 9.5%. Bio19, mean temperature of coldest quarter (Bio11), and annual mean temperature (Bio01) were the
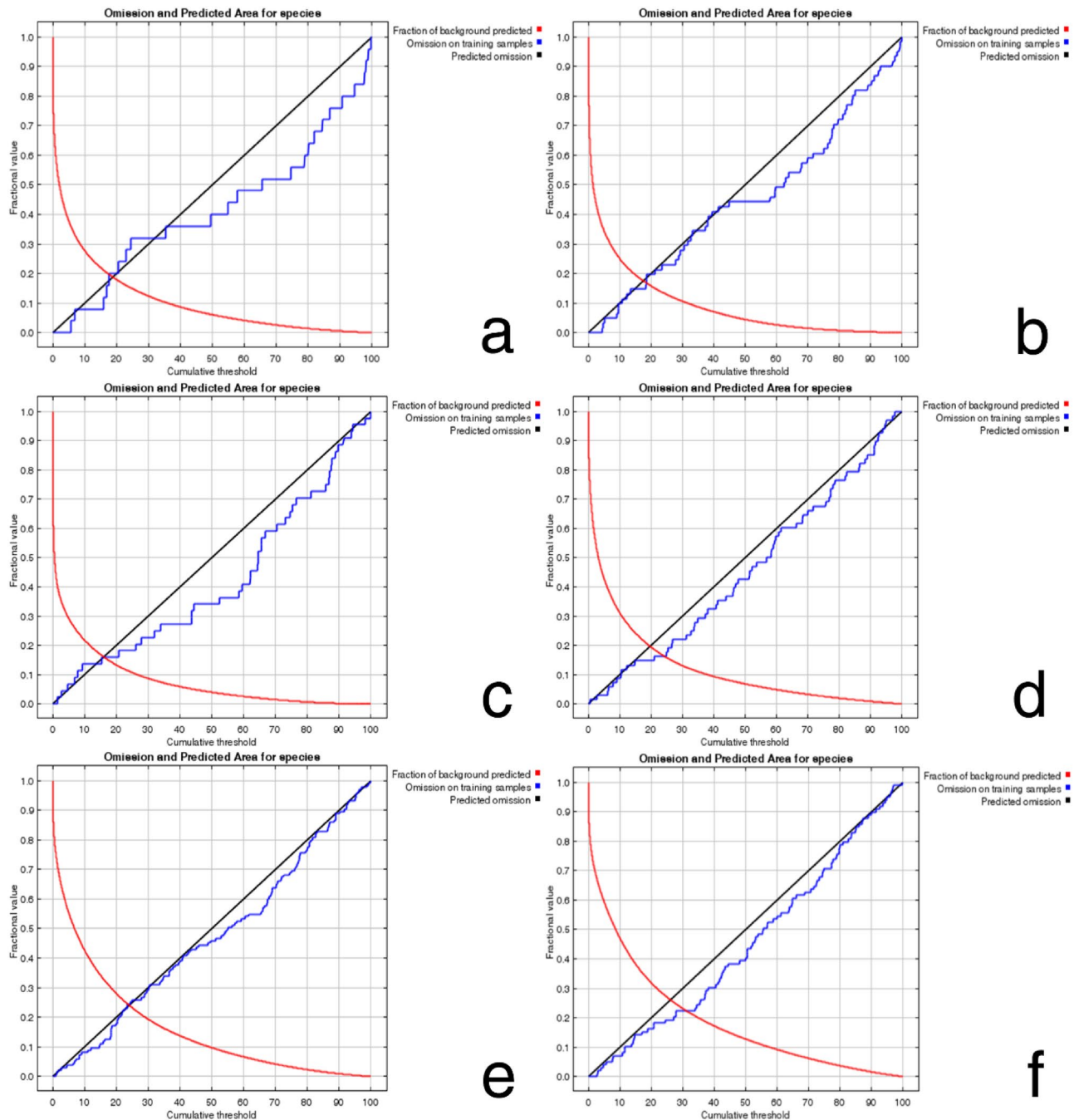
Figure 2. Omission rate of six important quarantine species: (**a**) *Crenidorsum aroidephagus*, (**b**) *Aleurothrixus trachoides*, (**c**) *Paraleyrodes minei*, (**d**) *Nasonovia ribisnigri*, (**e**) *Macrosiphum euphorbiae*, (**f**) *Viteus vitifoliae*. Fraction of background predicted (red line), omission rate for training samples (blue line), and predicted omission rate (black line). Figures were created using R version 4.04 (https://www.r-project.org/).

most influential factors on the prediction of habitat suitability for *N. ribisnigri*, with a total contribution of 80.4%. By contrast, min temperature of coldest month (Bio06) and annual precipitation (Bio12) had little influence on the results, with a total contribution of 10%. Bio19, Bio03, Bio11, and Bio01 were the most influential factors on the prediction of habitat suitability for *M. euphorbiae*, with a total contribution of 83.6%. By contrast, Bio12 and mean temperature of wettest quarter (Bio08) had little influence on the results, with a total contribution of 4.5%. Bio09, Bio11, Bio19 and Bio01 were the most influential factors on the prediction of habitat suitability for *V. vitifoliae*, with a total contribution of 93.1%. By contrast, Bio09, Bio03, and max temperature of warmest month (Bio05) had little influence on the results, with a total contribution of 6.8%.

The response curves for the selected variables are shown in Fig. 11. These response curves indicated several patterns of variation. They were generally unimodal, monotonically increasing or decreasing, and were not overly complex or suggestive of overfitting. For example, Bio19, a key variable for *C. aroidephagus*, exhibited a
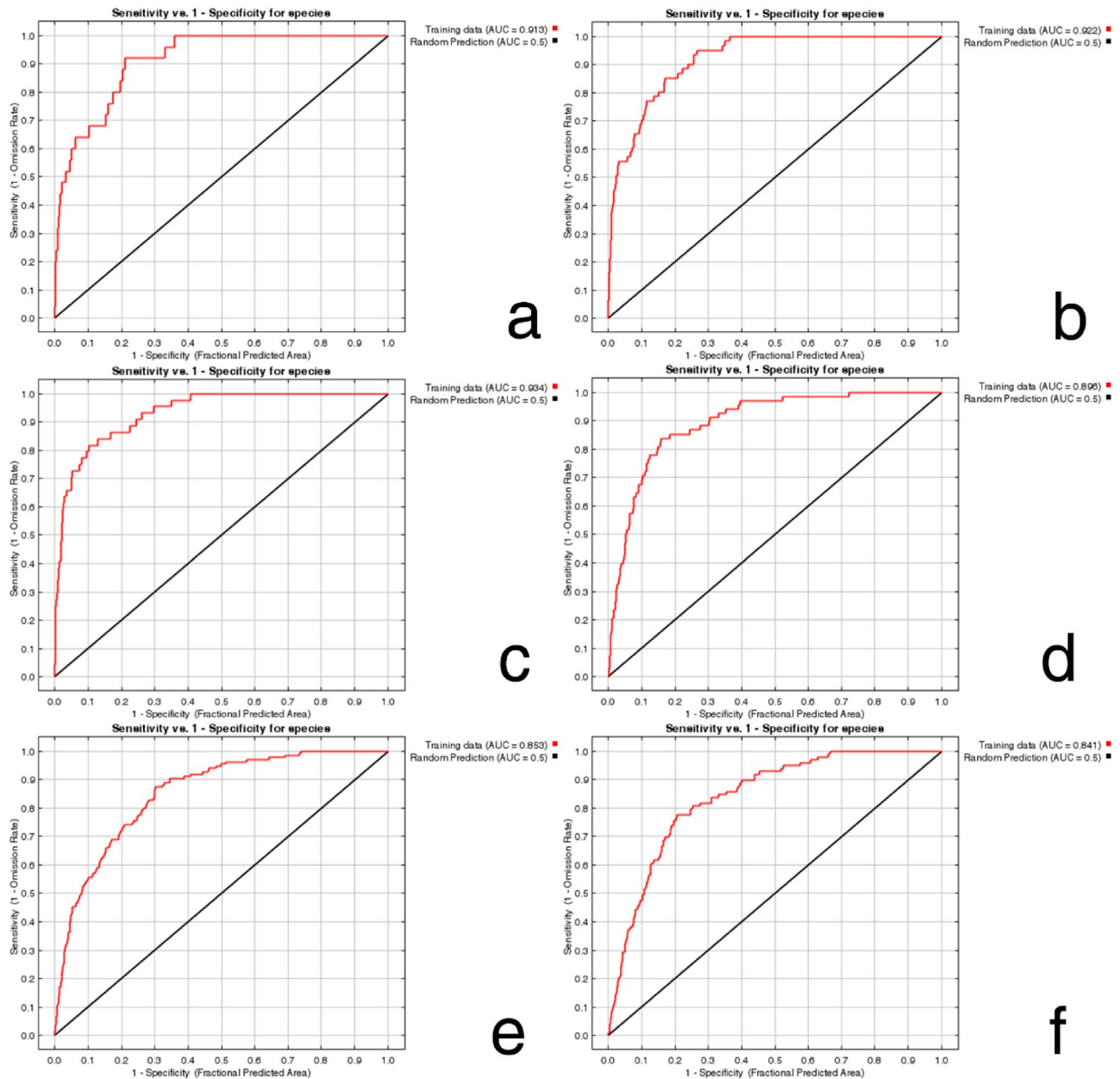
Figure 3. The receiver operating characteristic (ROC) curve and area under the curve (AUC) of six important quarantine species: (**a**) *Crenidorsum aroidephagus*, (**b**) *Aleurothrixus tracheoides*, (**c**) *Paraleyrodes minei*, (**d**) *Nasonovia ribisnigri*, (**e**) *Macrosiphum euphorbiae*, (**f**) *Viteus vitifoliae*. Training data (red line), random prediction (black line). Figures were created using R version 4.04 (https://www.r-project.org/).

unimodal pattern and Bio04 exhibited a unimodal pattern in the case of *A. tracheoides*. The probability of occurrence monotonically increased with Bio19, a key factor for *P. minei*. Bio11 exhibited a unimodal pattern in the cases of *N. ribisnigri* and *M. euphorbiae*.

Discussion

In this study, six important quarantine pests (three whitefly species and three aphid species not yet recorded in Taiwan) were selected for habitat suitability analysis prior to their invasion. The results revealed good modelling performance when potential distribution areas matched with the cultivation areas of host plants. This is the first study to successfully use SDM for the preliminary step of risk assessment of these emerging pests with limited biological data. Future work will focus on performing plausible spread simulations that take into account predicted suitability as well as potential dispersal paths. These paths depend on the existence of physical barriers such as water bodies and roads, biological constraints such as natural enemies (e.g. parasitoids, predators, and entomopathogenic fungi), and other drivers including human population density and trade nodes.
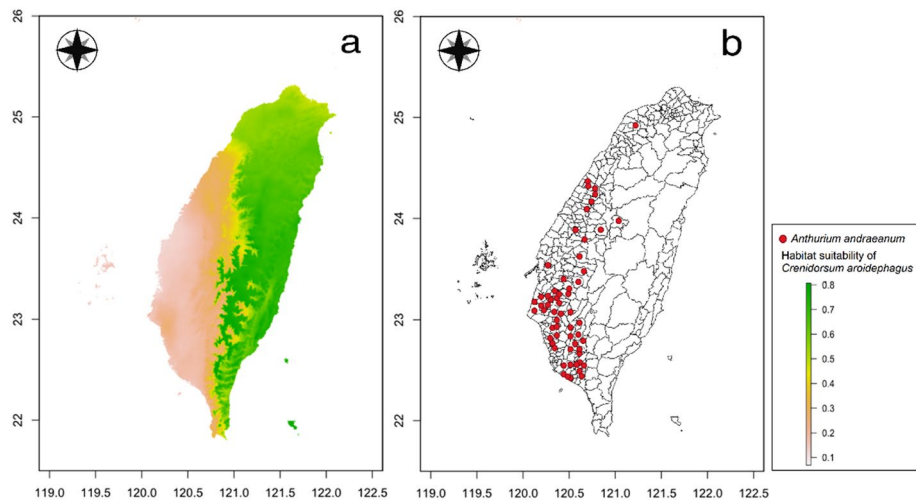
Figure 4. Comparison of potential distribution area of *Crenidorsum aroidephagus* and cultivation areas of *Anthurium andraeanum* in Taiwan: (**a**) potential distribution (**b**) cultivation areas. Maps were created using R version 4.04 (https://www.r-project.org/) and QGIS version 3.12.1 (www.qgis.org).
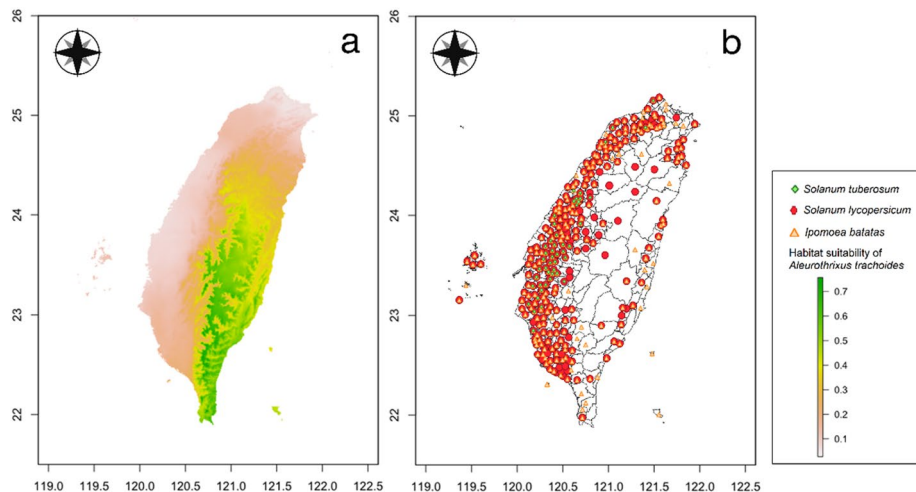

Figure 5. Comparison of potential distribution area of *Aleurothrixus trachoides* and cultivation area of *Solanum tuberosum*, *S. lycopersicum*, and *Ipomoea batatas* in Taiwan: (**a**) potential distribution (**b**) cultivation areas. Maps were created using R version 4.04 (https://www.r-project.org/) and QGIS version 3.12.1 (www.qgis.org).

Feasibility of SDM for evaluating potential pest invasion. The island of Taiwan covers an area of approximately 36,000 km², and the highest altitude is 3952 m. Approximately 70% of the land is mountainous and hilly. Plain areas in Taiwan are mainly concentrated at the western coast. The considerable differences in terrain and altitude create a diverse ecological environment. Because of restrictions in the terrain, crops are also mainly cultivated in small areas. Therefore, if exotic phytophagous pests invade Taiwan, it is highly probable that they will find sufficient food resources and growth conditions to establish and spread the population[5]. The pest risk assessment of important international quarantine pests requires the participation of experienced insect taxonomists. However, because insect species are so diverse, conducting risk assessment based on expert experience alone can sometimes be difficult. Therefore, in this study, MaxEnt was used to predict the habitat suitability of quarantine pests, and the potential distribution areas of the host plants of pests were compared with each other.

In this study, omission rates and ROC curves were used to verify the credibility and accuracy of the habitat suitability predicted by MaxEnt[24,25]. The verification results indicated that the potential distribution areas were predicted with moderate to high accuracy (Figs. 2, 3). Jiang et al.[26] proposed that prediction differs due to factors such as species, sample size, operation method, explanatory variables, and research scale. As the sample size (species distribution localities) increases, the accuracy of the MaxEnt prediction also increases until reaching the maximum accuracy and then eases[27]. However, large samples (such as those of 2,008 and 8,014 localities) will slightly reduce the prediction accuracy. By contrast, some models constructed with small sample sizes (< 10
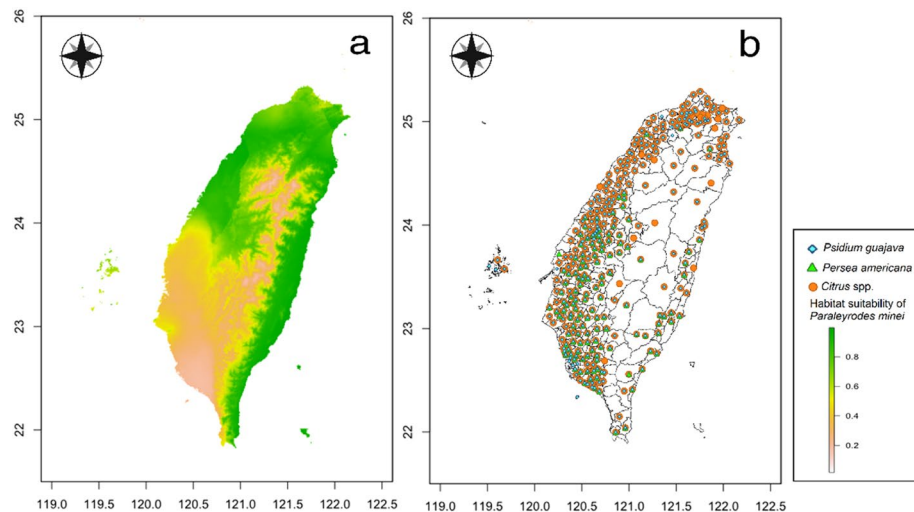
Figure 6. Comparison of potential distribution area of *Paraleyrodes minei* and cultivation areas of *Psidium guajava*, *Persea americana*, in Taiwan: (**a**) potential distribution (**b**) cultivation areas. Maps were created using R version 4.04 (https://www.r-project.org/) and QGIS version 3.12.1 (www.qgis.org).
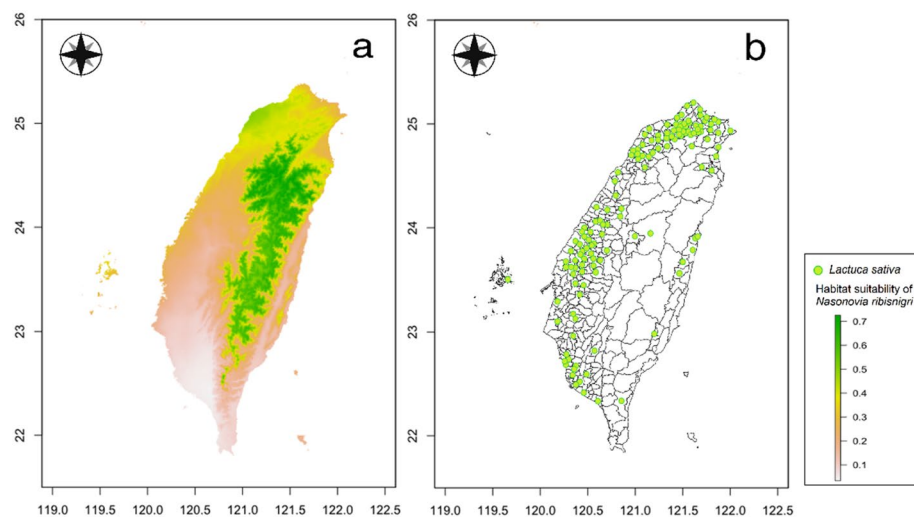

Figure 7. Comparison of potential distribution area of *Nasonovia ribisnigri* and cultivation areas of *Lactuca sativa* in Taiwan: (**a**) potential distribution (**b**) cultivation areas. Maps were created using R version 4.04 (https://www.r-project.org/) and QGIS version 3.12.1 (www.qgis.org).

localities) may also perform well. Although this study included a relatively few samples (25–136 localities), the accuracy rate was high.

A further comparison of the MaxEnt prediction areas with the crop cultivation areas revealed that except for *C. aroidephagus*, *A. trachoides*, and *N. ribisnigri* which pose low risk to crops in the their cultivation areas, the other three species pose a high risk to the host plant cultivated in a given areas (or some part thereof Figs. 4, 5, 6, 7, 8, and 9). *P. minei* poses a high risk to citrus, guava, and avocado cultivation in Taiwan, except in southwestern Taiwan; *M. euphorbiae* poses a high risk to potato, tomato and sweet potato cultivation in northern, central, and eastern Taiwan, but not in southwestern Taiwan; and *V. vitifoliae* poses a high risk to grape cultivation areas, except in southwestern Taiwan. The results of the risk assessment can also be applied to pest management. A pest monitoring mechanism can be established for areas with a high probability of invasion. In addition, our future research will aim to simulate potential spread paths as part of pest risk assessment.

Environmental variables influencing the habitat suitability of pests. Our results revealed that the six important quarantine pests were influenced by different climatic factors (Figs. 10, 11). First, precipitations of the coldest quarter (Bio19) significantly affected habitat suitability for *C. aroidephagus*, *P. minei*, *N. ribisnigri*, *M. euphorbiae*, and *V. vitifoliae*. The response curves of Bio19 exhibited a unimodal pattern for *C. aroidephagus*, *P.*
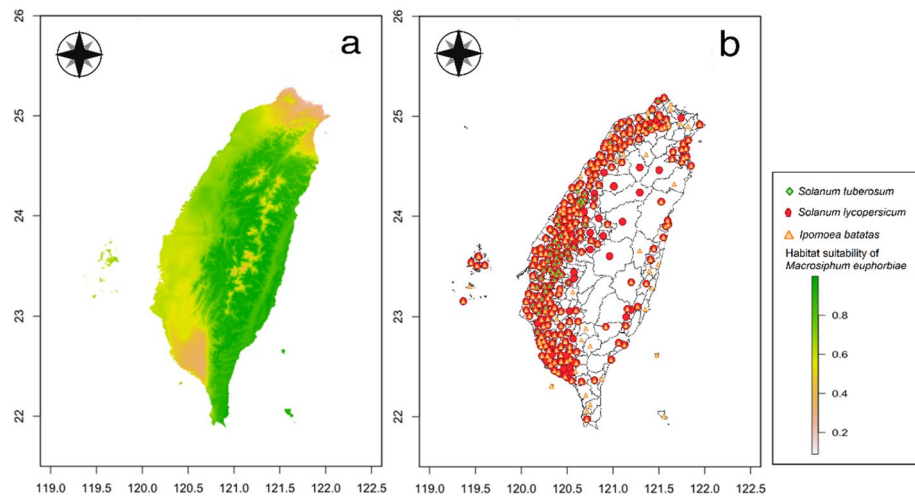
Figure 8. Comparison of potential distribution area of *Macrosiphum euphorbiae* and cultivation areas of *Solanum tuberosum*, *S. lycopersicum*, and *Ipomoea batatas* in Taiwan: (**a**) potential distribution (**b**) cultivation areas. Maps were created using R version 4.04 (https://www.r-project.org/) and QGIS version 3.12.1 (www.qgis.org).
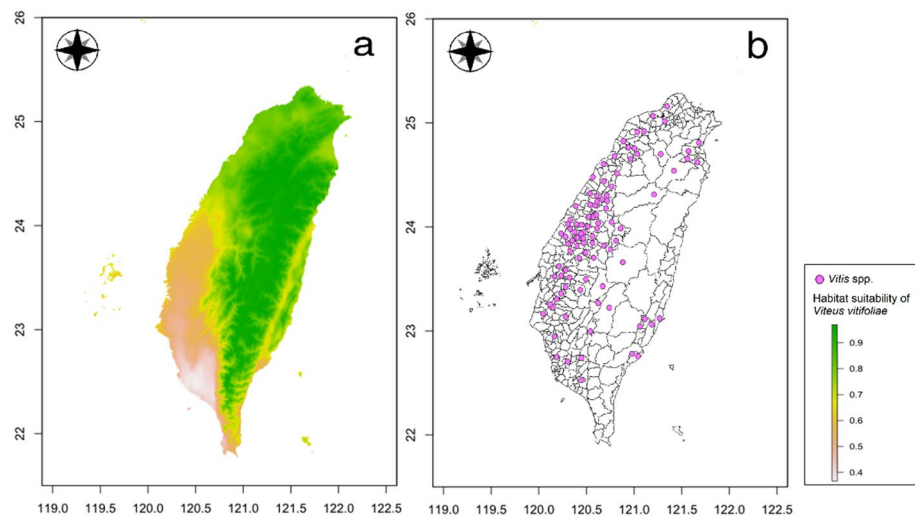

Figure 9. Comparison of potential distribution area of *Viteus vitifoliae* and cultivation areas of *Vitis* spp. in Taiwan: (**a**) potential distribution (**b**) cultivation areas. Maps were created using R version 4.04 (https://www.r-project.org/) and QGIS version 3.12.1 (www.qgis.org).

minei, *N. ribisnigri*, and *M. euphorbiae*, which suggested that the precipitations in winter limits habitat suitability for these species. By contrast, Bio19 exhibited a monotonically increasing pattern for *V. vitifoliae*, which suggests that higher precipitations in winter positively affects habitat suitability for this species. Second, seasonal temperature change is a key factor that negatively affected habitat suitability for *C. aroidephagus* and *A. trachoides*, and the results are consistent with those of previous biological studies[28–31] and occurrence data (Fig. 1). Third, the suitable temperature in driest quarter for *P. minei* was approximately 20 °C which is consistent with biological data[32]. Fourth, *N. ribisnigri* was also influenced by mean temperature of coldest quarter. Previous studies have revealed that *N. ribisnigri* grows at a temperature of 18 °C–24 °C. *M. euphorbiae* preferred areas with low temperature change, which is consistent with previous biological studies: the habitat suitability for *M. euphorbiae* decreased with temperature[33–35]. The response curve results for *P. minei* and *V. vitifoliae* are consistent with those of Zhang et al.[13] and Liu et al.[21].

Uncertainties and limitations of the study based on MaxEnt. Although the study model demonstrated moderate to high accuracy, improvements are necessary to overcome the uncertainties and limitations associated with MaxEnt prediction. The first limitation of this study is the number of distribution data of species. The reliability and stability of MaxEnt increase as the sample size increases[27]. The uncertainties of distribution
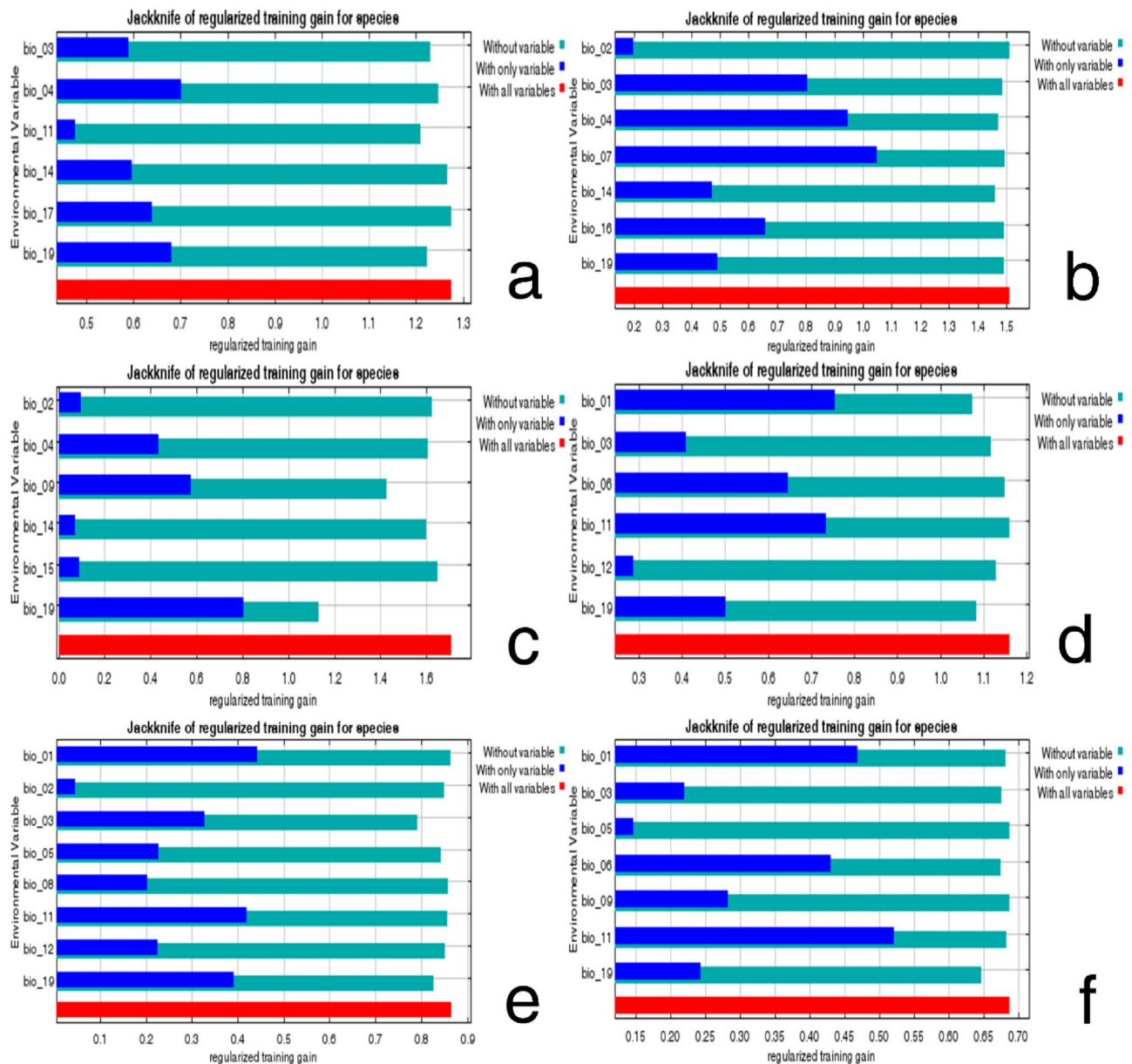
Figure 10. Relative importance of environmental variables of six important quarantine species based on the Jackknife test: (**a**) *Crenidorsum aroidephagus*, (**b**) *Aleurothrixus trachoides*, (**c**) *Paraleyrodes minei*, (**d**) *Nasonovia ribisnigri*, (**e**) *Macrosiphum euphorbiae*, (**f**) *Viteus vitifoliae.* Figures were created using R version 4.04 (https://www.r-project.org/).

data are mainly due to sample size, location, background data (pseudo-absence), and spatial deviation in species sampling[36]. The error deviation of the species distribution locality will be used to select environmental variable information in the MaxEnt calculation process, resulting in the incorrect derivation of the relationship between species and environment, leading to incorrect prediction results[37–40]. Here, we applied distribution data from previous literature and global databases; thus, the accuracy of the related spatial locality could not be guaranteed. Moreover, increasing background points should be considered that does not improve model performance[12]. Next limitation was the availability and selection of environmental prediction variables, biological factors, and host plant data. Araujo and Guisan[41] reported that predictive variables used in the prediction model should focus on the explanatory and ecological basis of environmental variables. However, in reality, predictive variables used are limited by data availability. Typically, many of the predictive variables are related to insect physiology, and the relevant variables are difficult to obtain or are lacking. For example, *C. aroidephagus* and *P. minei* analysed in this study are newly emerging pests and currently, available biological is very scarce. In addition, the default settings of MaxEnt could lead to excessive complexity and overfitting of the data; therefore, the adoption of a user-defined model is highly recommended[42]. Finally, the application of different SDMs may result in different prediction results. Many studies have compared the prediction results of different SDMs[43,44]. Zhang et al.[13] also used MaxEnt to analyse the potential distribution areas of *V. vitifoliae* in China and compared them with

		Crenidorsum aroidephagus		Aleurothrixus trachoides		Paraleyrodes minei		Nasonovia ribisnigri		Macrosiphum euphorbiae		Viteus vitifoliae	
		Percent contribution	Permutation importance	Percent contribution	Permutation importance	Percent contribution	Permutation importance	Percent contribution	Permutation importance	Percent contribution	Permutation importance	Percent contribution	Permutation importance
bio01	Annual mean temperature (°C)							20.2	37.4	12.6	0.4	18.2	31.9
bio02	Mean diurnal range (°C)			6.3	1.2	5.3	4.2			3.8	8.8		
bio03	Isothermality	15.7	14.4	14.5	47.3			9.6	5	21.5	40.3	2.5	3.9
bio04	Temperature seasonality (°C)	26.7	64.3	44.7	7.1	7.5	4.8						
bio05	Max temperature of warmest month (°C)									8.2	17.2	0.5	0
bio06	Min temperature of coldest month (°C)							6.3	15.2			21.6	26.7
bio07	Temperature annual range (°C)			4.6	24.6								
bio08	Mean temperature of wettest quarter (°C)									1	3.3		
bio09	Mean temperature of driest quarter (°C)					27.5	40.7					3.8	0
bio10	Mean temperature of warmest quarter (°C)												
bio11	Mean temperature of coldest quarter (°C)	8.4	3.7					22	15.3	15.6	8.4	24.5	30.3
bio12	Annual precipitation (mm)							3.7	7.1	3.5	4.1		
bio13	Precipitation of wettest month (mm)												
bio14	Precipitation of driest month (mm)	3	3	15.2	0.8	9.3	7.7						
bio15	Precipitation seasonality (mm)					4.2	1.7						
bio16	Precipitation of wettest quarter (mm)			10	17.9								
bio17	Precipitation of driest quarter (mm)	14.1	0										
bio18	Precipitation of warmest quarter (mm)												
bio19	Precipitation of coldest quarter (mm)	32.2	14.6	4.6	1.2	46.2	41	38.2	20.1	33.9	17.5	28.8	7.2

Table 1. Contribution rate and permutation importance of each environment variables.
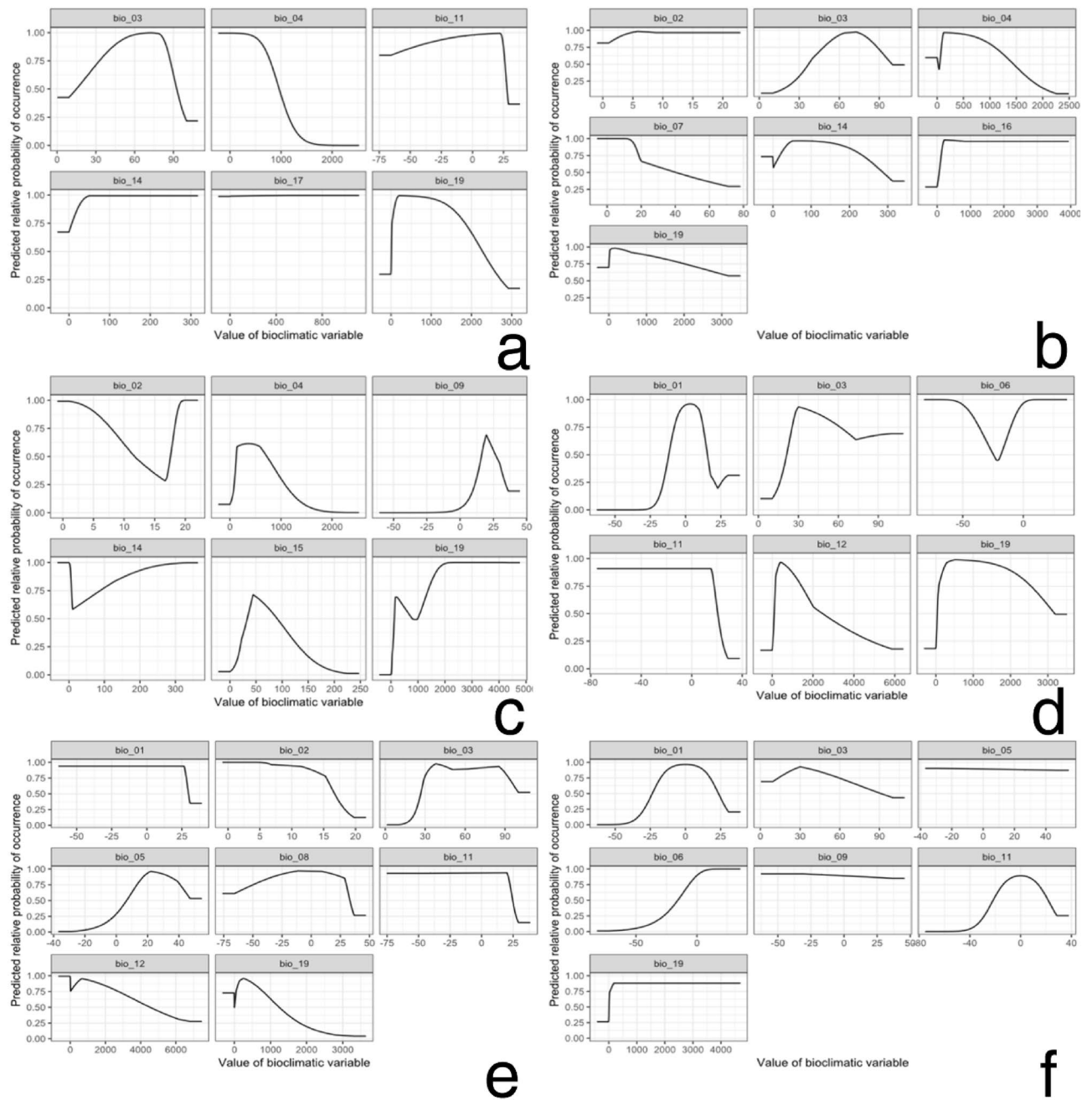
10

Figure 11. Response curves for selected variables of six important quarantine species: (**a**) *Crenidorsum aroidephagus*, (**b**) *Aleurothrixus trachoides*, (**c**) *Paraleyrodes minei*, (**d**) *Nasonovia ribisnigri*, (**e**) *Macrosiphum euphorbiae*, (**f**) *Viteus vitifoliae*. Figures were created using R version 4.04 (https://www.r-project.org/).

those reported by Ma et al.[45], who used CLIMEX to predict the distribution areas of *V. vitifoliae*. Although Zhang et al.[13] pointed out the difference between the prediction results of Ma et al.[45] and theirs, the authors believe that the prediction accuracy of CLIMEX requires a large amount of biological development data to analyse species niche. Further, certain biological development data are difficult to obtain. By contrast, MaxEnt uses the known species distribution data and environmental data as the driving force and the maximum entropy theory as the algorithm basis, which is very suitable for simulating species with limited distribution data.

Conclusions
Invasive alien species cause major economic and ecological losses[1,2]. Therefore, prevention strategies for the invasion of alien species are considered important enough to require the attention of many countries worldwide. In fact, both whiteflies and aphids are well-known tiny agricultural pests that damage a wide range of arbor trees, vegetables and ornamental crops worldwide. They are often spread to parts of the world through trade or human activities. In this study, we selected six important quarantine pests for analysis. To our knowledge, this is the first

study to predict the habitat suitability of these emerging pests. We input their current global distribution data into MaxEnt to predict the potential distribution areas of these six pests in Taiwan and compared the data with the host plant cultivation areas of the six pests to analyse their invasion risk in Taiwan. We believe that this is the first step for pest risk assessment, which can predict the potential distribution of these pests even without basic biological data. In the future, simulation of potential spread paths will be conducted for pest risk assessment.

Methods

Dataset of species occurrences and environmental variables.
Worldwide localities of six important quarantine species were collected from CABI (https://www.cabi.org/) and Global Biodiversity Information Facility (https://www.gbif.org/), and the latitude and longitude coordinates of the location were obtained using Google Earth Pro. The occurrence of repeated data due to the different sources of the collected distribution data may cause research bias and have other influences, leading to increased density of the species distribution data in a certain area; this may lead to deviations in the prediction results[10,11,21]. To avoid affecting the prediction results, we excluded duplicate and unclear distribution localities by using Quantum GIS (QGIS)[46] and ensuring the presence of only one distribution point in each raster. In total, 25 localities of *C. aroidephagus*, 62 localities of *A. trachoides*, 44 localities of *P. minei*, 68 localities of *N. ribisnigri*, 136 localities of *M. euphorbiae*, and 99 localities of *V. vitifoliae* were included for analysis (Fig. 1, Supplementary Tables S1–S6). All distribution data were organised using Microsoft Excel and in accordance with the MaxEnt software format requirements and were saved in .csv file format for further analysis[47].

We used a set of 19 environmental variables from WorldClim v.2.1 (https://www.worldclim.org)[49] with the 'current period' of 1970–2000 at a spatial resolution of 30 arc-seconds (1 km^2). These data contain 19 variables of these years derived from monthly temperature and precipitation values, representing annual trends, seasonality, and extreme or restrictive environmental factors. This bioclimatic variable factor is related to the distribution and survival of small arthropods[49] and has been widely used in global studies of species habitat suitability prediction[50]. We extracted 19 bioclimatic variables from the WorldClim database by R studio to use the R packages 'raster'[51] and 'rgdal'[52]. These variables were imported into MaxEnt for initial model to calculate the contribution rate using jackknife test to avoid multicollinearity. The selected variables are shown in Table 1.

Distribution modelling.
MaxEnt v.3.1.4[53] was used to predict the habitat suitability of the aforementioned six important quarantine pests in Taiwan (i.e., if they are introduced) and outlying islands on the basis of their global distribution data (Fig. 1, Supplementary Table S1–S6) and environmental variables (Table 1) through the use of the R package 'dismo'[54]. The R package 'ENMeval'[55] was applied to evaluate the performance of various models and parameters in relation to the best fitting model in ten-fold replicates. The logistic output was chosen as an estimate of the probability of the outcome being present conditional on the variables, per grid cell. This study only used presence data to generate pseudo-absences, and 10,000 background points were randomly selected by the MaxEnt model. The MaxEnt model ran either 500 iterations of these processes or continued until a convergence threshold of 0.00001 was met. Parameters were set based on best models for each species; the five feature types (i.e., linear, quadratic, product, threshold, and hinge) were allowed in the modelling and the regularisation multiplier was set to 0.5. A jackknife test was conducted to screen for dominant climatic factors and to calculate the contribution of various bioclimatic variables to pest habitat suitability (Fig. 10)[56]. In addition, response curves for selected variables were also generated (Fig. 11).

The prediction results from Maxent modelling were evaluated according to threshold-independent AUC values, calculated in R software. ROC curves were used to plot the true-positive rate against the false-positive rate, and the AUC was used to assess the predictive accuracy of models[8,24]. We selected a test sensitivity of 0% and an omission rate (OR) of 10%[57,58]. The AUC value ranged from 0 to 1, where higher values indicate higher predictive accuracy of models[22,23]. In the case of the default rate (OR), the value at 10% was 0.10, and the sensitivity test value at 0% was 0; poor performance is indicated by a value exceeding the predicted rate[59]. For each species, the final models were generated for probable distributions in Taiwan (Figs. 4, 5, 6, 7, 8, and 9) and the globe (Supplementary Fig. S1).

Host crop mapping.
We mainly selected economically important host crops for these six important quarantine pests. One host plant was chosen for oligophagous pests, *A. andraeanum* for *C. aroidephagus*[31], *L. sativa* for *N. ribisnigri*[60], and *Vitis vinifera* for *V. vitifoliae*[61]. For polyphagous pests, three crops were selected: tomato, sweet potato, and potato for *A. trachoides* and *M. euphorbia*[30,62]. The hosts of *P. minei* are citrus, avocado, and guava[63]. We then obtained data on the 2019 Taiwan crop cultivation areas listed by the Agriculture and Food Administration of the Executive Yuan from the agricultural report resource network database by using towns (or district) as a point. Subsequently, we obtained the coordinates of the locations using Google Earth Pro and located the crop areas using QGIS[46]. The crop cultivation maps were compared with the prediction results of MaxEnt (Figs. 4, 5, 6, 7, 8, and 9) and a histogram of the habitat suitability of crop locations for six important quarantine species was generated (Supplementary Fig. S2).

References

1. Pimentel, D., Zuniga, R. & Morrison, D. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econom.* **52**, 273–288. https://doi.org/10.1016/j.ecolecon.2004.10.002 (2005).
2. Bellard, C., Cassey, P. & Blackburn, T. M. Alien species as a driver of recent extinctions. *Biol. Lett.* **12**, 20150623. https://doi.org/10.1098/rsbl.2015.0623 (2016).
3. Hulmes, P. E. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.* **46**, 10–18. https://doi.org/10.1111/j.1365-2664.2008.01600.x (2009).
4. Lockwood, J. L., Hoopes, M. F. & Marchetti, M. P. *Invasion Ecology, 2nd ed.* (Wiley-Blackwell, 2013)
5. Wan, F. H. & Yang, N. W. Invasion and management of agricultural alien insects in China. *Annu. Rev. Entomol.* **61**, 77–98. https://doi.org/10.1146/annurev-ento-010715-023916 (2016).
6. Wittenberg, R. & Cock, M. J. W. *Invasive Alien Species: A Toolkit of Best Prevention and Management Practices.* (CAB International, 2001)
7. Elith, J. *et al.* A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**, 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x (2011).
8. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026 (2006).
9. Barbet-Massin, M., Rome, Q., Villemant, C. & Courchamp, F. Can species distribution models really predict the expansion of invasive species?. *PLoS ONE* **13**, 0193085. https://doi.org/10.1371/journal.pone.0193085 (2018).
10. Peterson, A. T., Papes, M. & Eaton, M. Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography* **30**, 550–560. https://doi.org/10.1111/j.0906-7590.2007.05102.x (2007).
11. Fourcade, Y., Engler, J. O., Rödder, D. & Secondi, J. Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS ONE* **9**, 97122. https://doi.org/10.1371/journal.pone.0097122 (2014).
12. Phillips, S. J. & Dudík, M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x (2008).
13. Zhang, Y., Zhang, C. B., Hao, J. H. & Gu, H. T. Prediction of potential suitable distribution area of invasive alien pest, *Viteus vitifoliae* Fitch in China. *Chinese J. Ecol.* **34**, 1986–1993 (2015) **(in Chinese)**.
14. Prabhulinga, T. *et al.* Maximum entropy modelling for predicting the potential distribution of cotton whitefly *Bemisia tabaci* (Gennadius) in North India. *J. Entomol. Zool. Stud.* **5**, 1002–1006 (2017).
15. Anjum, H. & Ahmed, S. Y. An updated and consolidated review on Indian Aleyrodids fauna (Hemiptera: Aleyrodidae: Insecta) along with their host plant families and distributional records. *Rec. Zool. Surv. India* **119**, 381–417 (2019).
16. Ouvrard, D., Martin, J. *The White-files - Taxonomic checklist of the world's whiteflies (Insecta: Hemiptera: Aleyrodidae).* http://www.hemiptera-databases.org/whiteflies/ (2021)
17. Blackman, R. L. & Eastop, V. F. *Aphids on the World's Trees: An Identification and Information Guide.* (CAB International, 1994)
18. Blackman, R. L., Eastop, V. F. *Aphids on the World's Crops: An Identification and Information Guide. Second Edition.* (John Wiley & Sons, 2020)
19. Blackman, R. L., Eastop, V. F. *Aphids on the World's Herbaceous Plants and Shrubs.* (John Wiley & Sons Ltd, 2006)
20. Hsieh, C. H. *et al.* Identification of a new invasive Q biotype of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Taiwan by molecular markers. *Formosan Entomol.* **28**, 211–224. https://doi.org/10.6661/TESFE.2008016 (2008) **(in Chinese)**.
21. Liu, B., Qin, W. Q. & Yan, W. Potential geographical distribution of *Paraleyrodes minei* (Hemiptera: Aleyrodidae) in China based on maxent model. *J. Environ. Entomol.* **41**, 1276–1286 (2019) **(in Chinese)**.
22. Franklin, J. *Mapping Species Distributions: Spatial Inference and Prediction.* (Cambridge University Press, 2010)
23. Peterson, A. T. *et al. Ecological Niches and Geographical Distributions* (Princeton, 2011)
24. Peterson, A. T., Papeş, M. & Soberón, J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol. Model.* **213**, 63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008 (2008).
25. Velasco, J. A. & Gonzalez-Salazar, C. Akaike information criterion should not be a "test" of geographical prediction accuracy in ecological niche modelling. *Ecol. Inform.* **51**, 25–32. https://doi.org/10.1016/j.ecoinf.2019.02.005 (2019).
26. Jiang, H. Y. *et al.* Effects of sample size on accuracy of MaxEnt: A case study of *Fagus hayatae*. *Q. J. For. Res.* **36**, 101–114 (2014) **(in Chinese)**.
27. Peterson, A. T. *et al. Ecological Niches and Geographic Distributions.* (Princeton University Press, 2011)
28. Guan, H. Y. Fruit tree quarantine pests-*Viteus vitifoliae*. *J. Hebei Forest Sci. Technol.* **2**, 38–40 (1987) **(in Chinese)**.
29. Lu, J., Wang, Z. W., Wang, Z. Y. & Liao, H. L. Research progress on biological characteristics and control of *Daktulosphaira vitifoliae*. *Jiangxi Plant Prot.* **2**, 51–56 (2008) **(in Chinese)**.
30. Sundararaj, R., Amuthavalli, T. & Vimala, D. Invasion and establishment of the solanum whitefly *Aleurothrixus trachoides* (Back) (Hemiptera: Aleyrodidae) in South India. *Curr. Sci.* **115**, 29–31 (2018).
31. Martin, J. H., Agular, A. M. F. & Baufeld, P. *Crenidorsum aroidephagus* Martin & Aguiar sp. nov. (Sternorrhyncha: Aleyrodidae), a New World whitefly species now colonising cultivated Araceae in Europe, Macaronesia and The Pacific Region. *Zootaxa* **4**, 1–8n. https://doi.org/10.11646/zootaxa.4.1.1 (2001).
32. Iaccarino, F. M. Descrizione di. *Paraleyrodes minei* n. sp. (Homoptera: Aleyrodidae), nuovo aleirode degli agrumi, in Siria. *Boll. Lab. Entomol. Agrar. Portici* **46**, 131–149 (1990) **(in Italian)**.
33. Palumbo, J. C. *Seasonal abundance and control of the lettuce aphid,* Nasonovia ribisnigri*, on head lettuce in Arizona.* https://repository.arizona.edu/handle/10150/220018 (2000)
34. Palumbo, J. C. *Population growth of lettuce aphid,* Nasonovia ribisnigri*, on resistant butter and head lettuce cultivars.* https://repository.arizona.edu/handle/10150/214949 (2002)
35. Diaz, B. M. & Fereres, A. life table and population parameters of *Nasonovia ribisnigri* (Homoptera: Aphididae) at different constant temperatures. *Environ. Entomol.* **34**, 527–534. https://doi.org/10.1603/0046-225X-34.3.527 (2005).
36. Feeley, K. J. & Silman, M. R. Keep collecting: Accurate species distribution modelling requires more collections than previously thought. *Divers. Distrib.* **17**, 1132–1140. https://doi.org/10.1111/j.1472-4642.2011.00813.x (2011).
37. Phillips, S. J. *et al.* Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* **19**, 181–197. https://doi.org/10.1890/07-2153.1 (2009).
38. Naimi, B., Skidmore, A. K., Thomas, A. G. & Nicholas, A. S. H. Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *J. Biogeogr.* **38**, 1497–1509. https://doi.org/10.1111/j.1365-2699.2011.02523.x (2011).
39. Fitzpatrick, M. C., Gotelli, N. J. & Ellison, A. M. MaxEnt versus MaxLike: empirical comparisons with ant species distribution. *Ecosphere* **4**, 1–15. https://doi.org/10.1890/ES13-00066.1 (2013).
40. Yackulic, B. C. *et al.* Presence-only modelling using MAXENT: when can we trust the inferences. *Methods Ecol. Evol.* **4**, 236–243. https://doi.org/10.1111/2041-210x.12004 (2013).
41. Araujo, M. B. & Guisan, A. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* **33**, 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x (2006).

42. Halvorsen, R., Mazzoni, S., Bryn, A. & Bakkestuen, V. Opportunities for improved distribution modelling practice via a strict maximum likelihood interpretation of MaxEnt. *Ecography* **38**, 172–183. https://doi.org/10.1111/ecog.00565 (2015).
43. Elith, J. & Leathwick, J. R. Species distribution models: Ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. S.* **40**, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159 (2009).
44. Byeon, D.-A., Jung, S. & Lee, W.-H. Review of CLIMEX and MaxEnt for studying species distribution in South Korea. *J. Asia Pac. Biodivers.* **11**, 325–333. https://doi.org/10.1016/j.japb.2018.06.002 (2018).
45. Ma, F. *et al.* The potential geographical distribution of *Daktulosphaira vitifoliae* based on CLIMEX in China. *J. Environ. Entomol.* **36**, 293–297 (2014) **(in Chinese)**.
46. QGIS Development Team. *QGIS Geographic Information System. Open Source Geospatial Foundation Project.* http://qgis.osgeo.org (2019)
47. Phillips, S. J. *A brief tutorial on Maxent.* http://biodiversityinformatics.amnh.org/open_source/maxent/ (2017)
48. Fick, S. E. & Hijmans, R. J. WorldClim 2, New 1 km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315. https://doi.org/10.1002/joc.5086 (2017).
49. De Meyer, M. *et al.* Ecological niche and potential geographic distribution of the invasive fruit fly *Bactrocera invadens* (Diptera, Tephritidae). *Bull. Entomol. Res.* **100**, 35–48. https://doi.org/10.1017/S0007485309006713 (2010).
50. Bidinger, K., Loetters, S. & Roedder, D. Species distribution models for the alien invasive asian harlequin ladybird (*Harmonia axyridis*). *J. Appl. Entomol.* **136**, 109–123. https://doi.org/10.1111/j.1439-0418.2010.01598 (2012).
51. Hijmans, R. *Package "raster".* http://raster.r-forge.r-project.org/ (2011)
52. Bivand, R., Keitt, T. & Rowlingson, B. *Package "rgdal"* https://CRAN.R-project.org/package=rgdal (2015)
53. Phillips, S. J., Dudik, M. & Schapire, R.E. *Maxent software for modeling species niches and distribution (version 3.4.1)* https://biodiversityinformatics.amnh.org/open_source/maxent/
54. Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. *Package 'dismo'* http://cran.r-project.org/web/packages/dismo/index.html (2011)
55. Muscarella, R. *et al.* ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol. Evol.* **5**, 1198–1205. https://doi.org/10.1111/2041-210X.12261 (2014).
56. Hernandez, P. A., Graham, C. H., Master, L. L. & Albert, D. L. The effect of sample size and species characteristics on performance of different species distribution modelling methods. *Ecography* **29**, 773–785. https://doi.org/10.1111/j.0906-7590.2006.04700.x (2006).
57. Liu, C., White, M. & Newell, G. Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.* **40**, 778–789. https://doi.org/10.1111/jbi.12058 (2013).
58. Kumar, S., Neven, L. G., Zhu, H. & Zhang, R. Assessing the global risk of establishment of *Cydia pomonella* (Lepidoptera: Tortricidae) using CLIMEX and MaxEnt niche models. *J. Econ. Entomol.* **108**, 1708–1719. https://doi.org/10.1093/jee/tov166 (2015).
59. Boria, R. A., Olson, L. E., Goodman, S. M. & Anderson, R. P. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol. Model.* **275**, 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012 (2014).
60. Yeh, H. T., Ko, C. C. & Hsu, T. C. Diagnostic techniques for the quarantine pest, *Nasonovia ribisnigri* (Mosley) (Hemiptera: Aphididae). *Formosan Entomol.* **26**, 283–293 (2006) **(in Chinese)**.
61. Benheim, D. *et al.* Grape phylloxera (*Daktulospharia vitifoliae*)-a review of potential detection and alternative management options. *Ann. Appl. Biol.* **161**, 91–115 (2012).
62. CABI. *Invasive Species Compendium.* https://www.cabi.org/ISC (2018)
63. Sujithra, M., Rajkumar, P. V. H., Hegde, V. & Poorani, J. Occurrence of nesting whitefly *Paraleyrodes minei* Iaccarino (Hemiptera: Aleyrodidae) in India. *Indian J. Entomol.* **81**, 1–4 (2019).

Acknowledgements

Author contributions

H. T. Yeh, H. Y. Cheah, and C. C. Ko proposed the concept of the study. H. Y. Cheah drafted the first version of manuscript and performed calculations and statistically analysed the results. M. C. Chiu provided the instructions for MaxEnt. J. R. Liao drafted the manuscript. C. C. Ko was the head of the present research group and provided whitefly knowledge and funding. All coauthors participated in the scientific discussions and commented on 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/s41598-021-89914-w.

Correspondence and requests for materials should be addressed to M.-C.C. or J.-R.L.

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.