Modeling and mapping the current and future distribution of Pseudomonas syringae pv. actinidiae under climate change in China

Qing Li

Abstract

Citation: Qing Li Modeling and mapping the current and future distribution of Pseudomonas syringae pv. actinidiae under climate change in China. **protocols.io**
dx.doi.org/10.17504/protocols.io.mdic24e
Published: 21 Dec 2017

Protocol

Occurrence records of Psa

Step 1.

In this study, the occurrence points of Psa were obtained from field data collected by the authors in the Chinese provinces of Sichuan and Shaanxi, from the published literature, and from the online databases GBIF and EPPO (S1 Table). When coordinates were published, we used the records directly. If there were only localities, Google Earth was used to collect coordinates of the records. All occurrence records were checked for accuracy in ArcGIS prior to use. Records with obvious geocoding errors were discarded, and duplicate records were removed manually. All records were imported into Microsoft Excel and saved as “*.CSV” format.

Environmental variables

Step 2.

From the WorldClim database (http://www.worldclim.org), we obtained 67 environmental variables (19 bioclimatic variables and 48 monthly averages of temperature and precipitation) for the current period. In the Worldclim database, ‘current period’ was defined from 1950 to 2000, and these data have been widely used in creating species distribution models. In 2013, the Fifth Assessment Report was released by the UN’s Intergovernmental Panel on Climate Change (IPCC), and four representative concentration pathways (RCPs, including RCP2.6, RCP4.5, RCP6.0 and RCP8.5) were published in the report. The impacts of climate change strategies on greenhouse gas emissions are considered more in the RCPs scenarios, and the projection of future climate change is more scientifically described. RCP4.5 and RCP6.0 are medium greenhouse gas emission scenarios, and RCP4.5 is of higher priority than RCP6.0. Therefore, RCP2.6 (the minimum greenhouse gas emission scenario), RCP4.5 (the medium greenhouse gas emission scenario) and RCP8.5 (the maximum greenhouse gas emission scenario) for the 2030s (2021-2040), 2050s (2041-2060), 2070s (2061-2080) and 2080s (2071-2090) were selected for the future model prediction of Psa in China. The future environmental variables were downloaded from the Climate Change, Agriculture and Food Security (CCAFS) website. All environmental variables were in raster format with a 2.5-arc minute resolution (4.5 km^2).

Environmental variables derived from WorldClim and CCAFS, which has been widely used in the prediction of the potential distribution of species, can reflect the characteristics of temperature and precipitation as well as their seasonal variation characteristics. The 19 bioclimatic variables with strong biological significance explained the adaptation of species with extreme environmental factors. These variables were also suitable for describing the distribution of species across large scales such as the intercontinental scale [63, 64]. Due to the various reasons mentioned above, the environmental variables provided above were chosen as the initial variables to be used in the modeling in this article. Based on Worthington's method on how to filter available variables for modeling, the jackknife test was used to evaluate each variable's contribution to the simulation, and 25 variables were removed due to their lack of contribution (percent contribution=0). Next, the highly correlated variables were eliminated, and variables with a Pearson's $|r| \leq 0.8$ were retained. After this process, 22 variables (S2 Table) were retained to simulate the current and future distributions of Psa in China.

Distribution modeling

Step 3.

MaxEnt software was utilized to predict the suitable habitat distribution of Psa in China. MaxEnt uses presence-only and small sample size data to model habitat suitability as a function of environmental variables, and it is consistently among the highest performing SDM methods. Response curves indicate the relationships between climatic variables, and the predicted probability of the presence of Psa was determined by MaxEnt. The percent contribution and permutation importance of environmental variables were calculated, and jackknife procedures were executed in MaxEnt. These analysis methods are all useful to measure the importance of the environmental variables. There were 10 replicates, and a random test percentage was chosen for each replicate. The remaining model values were set to default values.

MaxEnt estimates the probability a species will be present based on presence records and randomly generates background points by finding the maximum entropy distribution. An estimate of habitat suitability for a species was exported from MaxEnt, and its range generally varied from 0 (lowest) to 1 (highest). Model predictions were imported into a geographic information system (GIS), and maps were generated using ArcMap. Four arbitrary categories of habitat suitability for Psa were defined as no suitability (0-5), low suitability (5-33), medium suitability (33–66) and high suitability (66-100) based on predicted habitat suitability.

In this study, the ROC curve method was utilized to assess the model's explanatory power. The AUC (area under roc curve) is an effective threshold-independent index that can evaluate a model's ability to discriminate presence from absence (or background). The evaluation criterion of AUC is illustrated in S3 Table.

For reducing the bias of estimation, in 1949, Quenouille proposed an unbiased method of nonparametric estimation, and Tukey renamed it jackknife in 1958. This method can estimate parameters and adjust the deviation without assumptions of distribution probability. In SDM, the jackknife method was used to analyze the effects of environmental variables on model results to choose dominant factors. The specific process involves 1. Calculating the training gain for the model with only variable. Higher training gain indicates that the variable has high prediction power and contributes greatly to species distribution; 2. Calculating the training gain for the model without a specific variable and analyzing the correlation between the removed variable and the omission error. If the removal of an environmental variable leads to a significant increase in the omission error, it indicates that the variable has a significant effect on the model's prediction; 3. Calculating the training gain for the model with all variables.

Models of the mean center of highly suitable areas

Step 4.

The mean centers of highly suitable areas of Psa in China were calculated according to Yue's formula:

In this formula, *t* is the variable of time (i.e., current, 2030s, 2050s, 2070s and 2080s), *I* is the patch number of highly suitable areas, *Si(t)* is the area of *i*th patch of highly suitable areas, *S(t)* is the total area of highly suitable areas, (*Xi(t), Yi(t)*) are the longitudinal and latitudinal coordinate, respectively, of the geometric center of the *i*th patch of highly suitable areas, and (*x(t), y(t)*) are the mean centers of the highly suitable areas. The shift in distance and direction of highly suitable areas in the period from *t* to *t* + 1 are, respectively, formulated as Yue [75],

where *D* is the shift in distance of the highly suitable area during the period of *t* to *t*+1; *θ* is the shift in direction of the highly suitable areas, where east is defined as 0°, north is defined as 90°, west is defined as 180° and south is defined as 270°.