Journal of Plant Protection Research eISSN 1899-007X

ORIGINAL ARTICLE

An approach to the meso-scale epidemiological behavior of *Plasmodiophora brassicae* from cruciferous crops under tropical conditions

Andres F. Quintero-Mercado[1,2], Juan Camilo Garcia-Peña[2,4], Andrea Botero-Ramirez[3], Celsa Garcia[2], Joaquín Guillermo Ramírez-Gil[2,4]*

[1] Universidad del Magdalena, Facultad de Ingeniería, Santa Marta, Colombia

[2] Departamento de Agronomía, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Sede Bogotá, 111321, Bogotá, Colombia

[3] Plant Growth Facility Lead, Faculty of Agricultural, Life, and Environmental Sciences, 5-17 Ag/Forestry Building University of Alberta, Canada

[4] Laboratorio de Agrocomputación y Análisis epidemiológico, Center of Excellence in Scientific Computing, Departamento de Agronomía, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, sede Bogotá, Colombia

DOI: 10.24425/jppr.2025.155057

Received: June 20, 2024
Accepted: August 21, 2024
Online publication: July 08, 2025

*Corresponding address:
jgramireg@unal.edu.co

Responsible Editor:
Mohamed Khan

Abstract

Plasmodiophora brassicae is an obligate parasite and a natural soil inhabitant that causes clubroot, a disease with significant economic impact in plants of the Brassicaceae family. This pathology is conditioned by plant/host interactions, edaphoclimatic variables, and mechanisms of inoculum dispersal. However, the epidemiology of this pathogen is not well understood, thereby limiting its incorporation into integrated disease management strategies (IDM). The objective of this work was to adjust a mesoscale risk and prognostic model of *P. brassicae* based on edaphoclimatic factors and potential dispersal mechanisms in brassica-producing areas in Colombia. The presence and inoculum density of the pathogen were determined by visual inspection of symptoms and quantification by qPCR of soil samples in a total of 127 plots located in regions with the highest production of species from the Brassicaceae family. In addition, an edaphoclimatic characterization was carried out based on field data and secondary information by web scraping using freely available databases. The forecast models were determined by fitting a Generalized Linear Model (GLM) using the logit and inverse link functions for binomial and gamma distributions, respectively. The meso- and macroscale spatial risk model was developed under point pattern approaches (Kernel density model and ecological niche model (ENM). The different epidemiological analysis approaches used suggest that *P. brassicae* presents a high risk in areas with host presence and conducive edaphoclimatic characteristics, indicating the need to carry out epidemiological surveillance, reduce the dispersion of infested soil, and implement *P. brassicae* exclusion methods.

Keywords: clubroot, driving factors, ecological niche model, generalized linear model, risk analysis

Introduction

Clubroot, caused by the obligate soil pathogen *Plasmodiophora brassicae* Woronin, is an economically impactful disease that affects plant species in the Brassicaceae family (Dixon 2009a). Taxonomically, *P. brassicae* is classified within the class Phytomyxea, order Plamodiophorida, and the family Plasmodiophoraceae (Hittorf *et al*. 2020). This disease is widely distributed throughout the world, with reports of its presence in more than 80 countries on all continents except Antarctica (Javed *et al.* 2023). In Colombia, for the

year 2019, the total area planted with vegetables was 23,626 ha. Cabbage, broccoli, cauliflower, and brussels sprout crops represented 10% (2,404.6 ha) of this area and their production was concentrated in the regions of Cundinamarca, Antioquia, Valle del Cauca, Boyacá, Cauca, Nariño and Norte de Santander (DANE 2020; Ministerio de Agricultura y Desarrollo Rural 2023).

Historically, losses due to clubroot have been reported to range from 10–15% (Dixon 2009b). However, more recent studies in multiple crops including canola, Chinese cabbage, cabbage, broccoli, and cauliflower in countries such as Canada, Nepal, China, and Colombia have shown that yield losses are proportional to disease severity levels and can range from 30 to 100% (Tewari *et al.* 2005; Timila *et al.* 2008). Its high economic impact is due to the fact that the causal agent of clubroot induces galls in the roots that prevent the absorption of water and nutrients by the plant, causing stunting, wilting, chlorosis, and in severe cases even the death of the host (Dixon 2009a; Strelkov and Hwang 2014).

The first report of clubroot in Colombia was in 1969 from the municipality of Manizales (Caldas) affecting a cabbage crop (Torres 1972). Since then, the disease has spread to all regions where species of the Brassicaceae family are planted (Botero-Ramirez *et al.* 2022a; b). Infestation of production plots by *P. brassicae* represents a major limitation for the growth of species within the Brassicaceae family in the country due to the longevity of resistance spores. Such spores survive an average of 4 years under temperate conditions but can remain viable in soil in the absence of a host for up to 17 years (Wallenhammar 1996; Hwang *et al.* 2013). However, it should be noted that the longevity of such structures under tropical conditions has not yet been evaluated.

Management of clubroot is difficult due to the longevity of pathogen resistance spores (Botero *et al.* 2019). For this reason, strategies have focused on two fronts: preventing infestation of pathogen-free plots and reducing the incidence and severity of symptoms in infested plots (Gossen *et al.* 2015). Strategies aimed at preventing the infestation of unaffected plots have focused on the sanitization of tools and farm machinery (Donald *et al.* 2009; Diederichsen *et al.* 2014; Hill *et al.* 2022). However, the use of this strategy is limited in Latin America (Botero *et al.* 2019). In this context, the exclusion of pathogens is a basic principle for the protection of areas where host species of *P. brassicae* are produced and where it is essential to know the edaphoclimatic conditions that are conducive to clubroot and to identify areas with potential inoculum movement. This implies the need to develop new strategies based on understanding the epidemiology of the disease for effective management of clubroot.

Clubroot epidemiology on the meso-scale (agroecological zone and region) and macroscale (country) is still not very well known, especially under tropical conditions. There are several possible reasons for this, including the difficulty in quantifying the disease (Bhering *et al.* 2017) and the lack of consistent studies with continuity over time (Botero *et al.* 2019). In temperate zones, severity is associated with a set of favorable conditions related to the inoculum, its density, distribution, and dissemination (Bhering *et al.* 2017).

In this regard, physical, chemical, and biological characteristics of soil are reported as conducive factors, especially texture, moisture, temperature, pH, Ca, N, B, as well as macro and microflora (Webster and Dixon 1991; Gossen *et al.* 2015). Meanwhile, at the climatic level, there is evidence that temperature may be an important factor (Sharma *et al.* 2011; Gossen *et al.* 2015). Additionally, since *P. brassicae* is a soil pathogen, another factor of epidemiological relevance is the mechanism of dispersal and colonization into new areas by resistance spores (RS), which is associated with soil movement, agricultural machinery, crop residues, and contaminated plant material (Cao *et al.* 2009; Strelkov and Hwang 2014).

Detailed knowledge of the epidemiology of plant diseases not only allows identification and modeling of their spatial distribution and temporal dynamics (Nelson *et al.* 1999; Madden *et al.* 2007), but also facilitates the definition of the variables that favor or limit their incidence and severity (Madden *et al.* 2007; Ramirez-Gil *et al.* 2021; Shah *et al.* 2021). Therefore, plant disease epidemiology presents a set of mathematical, statistical, and computational tools that allow better information management as a basis for evidence-based decision making (Shah *et al.* 2021).

In the context of plant disease epidemiology, generalized linear models (GLM) are widely recommended statistical methods for the determination of epidemiologically important variables and the adjustment to discrete-type plant disease intensity measures (Garret *et al.* 2004), especially when observations present a non-normal distribution (Garret *et al.* 2004). These models are considered interpretable and can be used to draw global conclusions about the phenomenon in question, derived from model-specific (significance, weight, and magnitude of predictors) or model-agnostic (partial dependence, feature and permutation feature importance, individual conditional expectations, etc.) methods (Molnar 2022).

On the other hand, ecological niche models (ENM) have been used as meso- and macroscale approaches to assess risk factors based on the relationship between predictor variables and the specific characteristics of the causal agents of disease, their vectors, and the interaction with their hosts (Escobar 2020). From ENMs it is possible to generate maps of spatial distribution and movement dynamics, potential colonization of new areas, potential changes under scenarios

of climate variability and change, etc. (Escobar 2020; Ramírez-Gil *et al.* 2021). ENMs are based on the principle that diseases do not occur randomly, so it is important to know the patterns and results in order to model and predict their behavior (Escobar 2020). In ENM, the use of Maxent stands out as the most widely used algorithm (Phillips *et al.* 2006; Elith *et al.* 2011; Merow *et al.* 2013).

To date, the availability of epidemiological studies of clubroot and its causal agent under tropical conditions is limited (Botero-Ramirez *et al.* 2019; Botero-Ramirez *et al.* 2022a, b). For this reason, it is necessary to explore in more depth the factors that determine the development of *P. brassicae* and how these can be used to adjust prognostic models; and to elucidate the spatial relationships of the pathogen with the potential development of the disease from a risk analysis approach at the meso- and macro- (country) scales. Based on the above, the hypotheses of this work were: (i) there are edaphoclimatic factors that lead to the presence of inoculum of *P. brassicae* and affect the amount of it that can be used for the adjustment of prognostic models, and (ii) spatial models based on the point pattern approach (Kernel density) and ENM are epidemiological tools to estimate the risk of *P. brassicae* in areas producing Brassicaceae in Colombia on meso- and macro spatial scales. Based on these hypotheses, the objectives were: (i) to adjust a predictive model for the presence and density of *P. brassicae inoculum* as a function of edaphoclimatic factors and (ii) to generate a meso- and macroscale risk model of *P. brassicae* in production areas in Colombia using the Kernel density approach and ENM.

Materials and Methods

Data on infested hosts and quantification of inoculum associated with *Plasmodiophora brassicae* in Colombia

To carry out this work, we used data on the presence of hosts infected with *P. brassicae* of the Brassicaceae family (cabbage, broccoli, and cauliflower) confirmed by visual inspection of aerial symptoms (wilting and chlorosis) and root galls (Botero-Ramirez *et al.* 2022a). This quantification measure was conducted in the field using systematic W-shaped sampling, collecting 30 points per plot, and inspecting three plants at each point. The presence of the pathogen in each host was confirmed in the laboratory by qPCR using DR1F and DR1R primers (Rennie *et al.* 2011).

Furthermore, the amount of resistance spores for *P. brassicae* was determined from soil samples using the soil count method, corroborated by qPCR using primers previously reported (Botero-Ramirez *et al.* 2022a). For each evaluated plot, five points or subsamples were taken from the first 20 cm of soil depth in a W-shaped transect. Each point constituted a composite sample of 500 grams of soil, ensuring complete coverage of the plot. The same soil collection procedure was followed in plots where farmers reported the presence of the disease through typical symptoms. The samples were transported in resealable bags to the Laboratorio de Suelos y Aguas de la Facultad de Ciencias Agrarias de la Universidad Nacional de Colombia sede Bogotá, for subsequent physical and chemical analyses.

The data were part of a prevalence study conducted in eight Brassicaceae producing regions of Colombia carried out during the months of February and March 2017, and involved sampling 127 plots (Botero--Ramirez *et al.* 2022a). These geographical zones were located at an altitudinal profile between 1,754 and 3,163 m elevation in areas of Colombia reported as production areas of the hosts of this pathogen (Fig. 1).

Direct and indirect determination of edaphoclimatic variables at the plot level

The set of climatic variables for each of the evaluated plots (Fig. 1) was determined using two approaches. The first was associated with the acquisition of minimum, average and maximum temperature (°C), accumulated precipitation (mm), and average relative humidity (%) for the months in which the sampling was carried out (February and March 2017). The second approach was associated with the calculation of the annual climatological norms (1981–2010) for each set of the above variables, including solar radiation (kJ m^{-2} day^{-1}), wind speed (m $^{s-1}$) and water vapor pressure (kPa).

Data were obtained from two sources depending on availability and quality. One source was the hydroclimatological stations of the IDEAM Institute of Hydrology, Meteorology and Environmental Studies (http://dhime.ideam.gov.co/atencionciudadano/) with influence (buffer zone of 5 km and 10 km over the same altitudinal profile for precipitation and temperature, respectively) on each of the evaluated plots. This information was complemented with data obtained from the free WordClim database (https://www.worldclim.org/data/worldclim21.html) (Fick and Hijmans 2017), at a spatial resolution of 1 km, extracting the values for each of the plot coordinates using the point sampling tool of plugins implemented in the free software qGis (QGIS Geographic Information System, QGIS Association 2021).

The edaphic variables were determined at each sampling point following the methodology of the Soil Laboratory of the Universidad Nacional de Colombia, Bogotá. The concentration of soil nutrients was quantitatively analyzed according to the protocols of the

laboratory. The variables determined and the analytical methods used were: pH in solution 1: 1 (VC-R-004 version 2); organic matter content (OM,%) (Walkey & Black); P using the Bray II method (mg × kg^{-1}); S determined by the monocalcium phosphate method (mg × kg^{-1}); exchangeable acidity (Al+H), using the reagent KCl 1 M [cmol(+) × kg^{-1}]; exchangeable Ca, Mg, K and Na using the Colombian Technical Standard (NTC 5349, 2008) [cmol(+) × kg^{-1}]; cation exchange capacity (CEC), determined as the sum of bases and exchangeable acidity [cmol(+) × kg^{-1}]; available Fe, Mn, Cu and Zn using Olsen's method (NTC 5526, 2007) (mg × kg^{-1}); and available B using the monobasic calcium phosphate method (mg × kg^{-1}). As a complement and using the coordinates of each of the points sampled and for a depth of 0-30 cm, the extraction of soil property values that were not present in laboratory analytical methods was performed using the values in raster format from the free worldwide soil variable database with a special resolution of ~250 m (Hengl *et al.* 2017). Point-to-point extraction was carried out according to the protocol described above for climatic variables. The parameters extracted were cation exchange capacity, nitrogen, organic carbon, bulk density, clay, silt, and sand content.

Epidemiological relationships of *Plasmodiophora brassicae* on a mesoscale and macroscale for Colombian conditions

Phase 1. Models to predict the presence and density of the *Plasmodiophora brassicae* inoculum

To identify prognostic models using edaphoclimatic variables at the meso- (region) level as predictors of the presence/absence of *P. brassicae* and resistance spores to the density of the inoculum [(ID) g · $soil^{-1}$] of *P. brassicae*, the GLM fit approach was used. A detailed and correct process of applying the GLMs was performed with strict adherence to the assumptions to ensure the validity and reproducibility of the results (Dobson and Barnet 2018; McCullagh 2019).

The presence/absence of *P. brassicae* variable, given its binary response to a number of events, was fitted to a GLM using a binomial probability distribution and a logic link function (Nelder and Wedderburn 1972). Meanwhile, the fit to the inoculum density was performed using a Gamma distribution and inverse link function, given that this variable presented positive values with deviation towards the upper level (Nelder and Wedderburn 1972).

An optimization process of the GLMs was carried out by evaluating a set of models, starting with the adjustment of all the edaphic and climatic variables as predictors, following an elimination scheme, where those variables with biological and epidemiological significance were kept. The procedure was performed using the h2o library (LeDell *et al.* 2020) implemented in the free R software. The models were calibrated with 80% of the data and validated by cross-validation using 5 Kfold. The other 20% (test set) was used to evaluate predictive capacity and determine the accuracy and performance of the models. Similarly, within the GLM configuration in the library, variables that presented collinearity were removed and standardized before being used as predictors.

For the presence/absence of *P. brassicae* variable classification model (binomial), the classes were balanced, and the following metrics were used as indicators of model performance: Akaike Information Criterion (AIC), Area under the ROC curve (AUC), Area under Precision Recall Curve (AUCPR), Gini index, Logarithmic Loss (Logloss), Mean Square Error (MSE) and Root Mean Square Error (RMSE). In addition, the confusion matrix was determined on the test set to determine the predictive capacity and accuracy of the model. For the regression model (using the Gamma distribution), the AIC, MSE, RMSE, mean average error (MAE), R^2 and root mean squared logarithmic error (RMSLE) were used as model performance metrics. The estimation and fitting of the GML parameters were performed using iteratively weighted least squares using log-likelihood (Nelder and Wedderburn 1972).

Finally, from the GLMs optimized for each response variable, global interpretations were made by obtaining the coefficients (with their respective magnitude and direction) of the predictors, their standardized importance and statistical significance, together with the partial dependence (on new observations, test set) of those variables with a standardized importance greater than 1%.

Phase 2. Potential distribution of *Plasmodiophora brassicae* in host-producing areas in Colombia as a basis for risk analysis

In the first phase of the spatial analysis and with the objective of having an approximation of the regional situation of *P. brassicae*, a descriptive statistical approach, spatial visualization and potential distribution were used. In this sense, the presence/absence was analyzed by determining the proportion of plots in which it was detected.

Meanwhile, the inoculum density variable and its relationship in geographical space were evaluated using a point pattern analysis process, by means of the Kernel density algorithm. This nonparametric method is used to estimate density from a finite data set using point entities and assigning a kernel function according to the intensity of the phenomenon that is associated with the value of each point (cell) and decreases until no related entities are found at the analysis distance or search radius (Silverman 1998; Skelsey 2021).

This approach can be used to infer the relative risk of a disease or its causal agent in geographic space based on reported cases of the disease and driving parameters (Bithell 1990; Skelsey 2021).

The Kernel density estimation library available in SAGA GIS 7.8.2 (Conrad *et al.* 2020), implemented in the free software QGIS 3.2 (QGIS Geographic Information System, QGIS Association 2021), was used to perform the analysis. The model parameters were as follows: Gaussian kernel, output resolution of ~1km and a search radius of 10 km from each point (Fig. 1B). This buffer zone was considered as the distance where the pathogen is present and may have moved or spread over short distances depending on dissemination factors such as the movement of machinery, operators, and plant material in the presence of infected soil (Cao *et al.* 2009; Rennie *et al.* 2011).

The ENM approach was used to understand the current and potential distribution of *P. brassicae* in the cultivated areas of hosts of this pathogen in Colombia. This method allows an approximation of the risk of the presence of a pathogen and, depending on favorable environmental conditions and susceptible hosts, the risk that disease events can occur (Ramirez-Gil *et al.* 2021). For the implementation of this analysis, a multistage process was followed to ensure correct use of the ENM approach, with emphasis on the model calibration phase (Merow *et al.* 2013; Cobos *et al.* 2019; Ramirez-Gil *et al.* 2021).

The first phase of the process consisted of removing the spatial autocorrelation of the data on the presence of *P. brassicae* and its associated disease (Fig. 1A), avoiding the sampling bias associated with areas having a greater number of reports. For this, a 1 km filter was applied based on the proximity of commercial plots producing host species. From this, duplicate reports of presence per raster cell (1 km) were discarded, obtaining a total of 50 presence points. Subsequently, an influence area of 10 km (area M) was determined for each point of presence (Fig. 1B orange zone), according to the indications previously described in the kernel analysis. Area M is part of the scheme called BAM (Biotic, Abiotic, and Mobility), and represents the area that a species has had access to in a period that

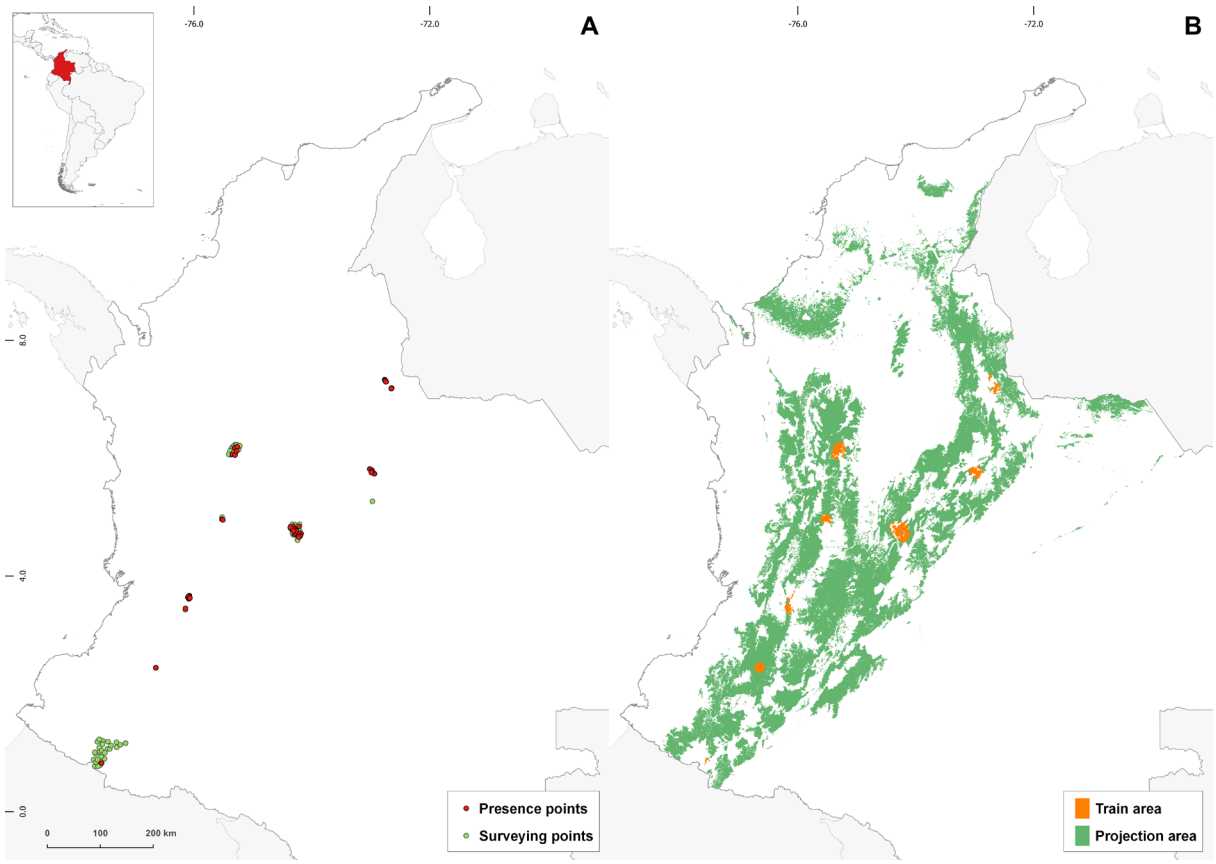

Fig. 1. Geographic location of samples of clubroot and *Plasmodiophora brassicae* in eight crucifer producing regions of Colombia and characteristics associated with the modeling of ecological niches.
A – geographic location of *P. brassicae* in Colombia. B: Basic characteristics associated with the calibration area and projection of ecological niche models for the generation of a mesoscale risk model of *P. brassicae*. Training area or area M was associated with a buffer of 10 km from each point where *P. brassicae* was found. The projection of the model was based on current and potential distributions areas for the cultivation of Brassicaceae species in Colombia (Supplementary file Tab. 1).

allows its establishment, development, and movement process (Barve *et al*. 2011; Saupe *et al.* 2012).

Area M (Fig. 1B orange zone) was used as a calibration area. A set of regionalized variables was selected to use as predictors in the ENMs. For this study, climatic variables and indices from the open access databases WordClim (Fick and Hijmans 2017) and ENVIREM (Title and Bemmels 2018) at 1 km resolution were selected. They were complemented with edaphic parameters obtained from Soilgrids (Hengl *et al*. 2017) at 250 m resolution. To obtain a set of predictors with the same spatial resolution (1 km) and geographic extension, the QGIS 3.2 raster calculator of the QGIS 3.2 software was used and the WordClim variables were used as a basis. Subsequently, a pre-selection of variables was made based on their biological importance related to pathogens and hosts and their potential relationship as epidemiological parameters (Bhering *et al.* 2017). A total of 31 predictors were initially used (Supplementary file 2).

In the next step, highly correlated variables were discarded using a Spearman correlation test, eliminating those variables with less epidemiological significance and with an index greater than or equal to 0.8 (Supplementary file 3). This resulted in a set of 16 variables (listed in Supplementary file 4). The analysis was performed using the Nichetoolbox library (Osorio-Olvera *et al.* 2020), run on open-source R software version 4.1.3 (R Core Team 2022), and based on the values of environmental variables for presence points.

From the previous analysis (selection of nonautocorrelated variables), the MaxEnt model calibration process was performed in region M (Fig. 1B), employing the SDMtune library (Vignali *et al.* 2020), implemented in open-source R software and using the MaxEnt algorithm (Phillips *et al.* 2006). Ten thousand background points (bg) were used, and the presences and background points were partitioned into an 80% training set (40 presences, 8000 bg) and a 20% calibration set (10 presences, 2000 bg). Since MaxEnt requires an internal optimization process to select the best model, a set of models resulting from the combination of predictor variables, different classes of algorithm characteristics, and variation in the hyperparameter β (from 0.5 to 5 in increments of 0.5) were evaluated (Merow *et al*. 2013).

Subsequently, to reduce the set of predictors, a model was calibrated with the Maxent default configuration, employing a combination of lqph feature classes (FC) and a regularization parameter β of 1. Then, a Jackknife test was performed, removing variables with a permutation importance of 5% and evaluating the model's performance based on the area under the curve (AUC), the Akaike information criterion corrected for small sample sizes (AICc), and the true skill statistic (TSS) after each removal (Supplementary file 4). Eventually, a reduced set of seven predictors with the greatest contribution was obtained, ensuring that they also had biological and epidemiological significance for the pathogen studied. Subsequently, the model's hyperparameters were optimized starting from the reduced set of variables, using the SDMtune library (Vignali *et al.* 2020). The seven variables selected previously were used, and with the predefined training and test sets, all possible combinations of FC and values from 0.5 to 5 with an interval of 0.1 were tested. The optimization and selection of the final model were based on an evaluation of 150 models, from which the one with the lowest AICc was selected, while also ensuring that it had TSS values greater than 0.4 and AUC values greater than 0.7 (Allouche *et al.* 2006; Osorio-Olvera *et al*. 2020).

Finally, based on the optimized model, the importance of permutation of environmental variables was calculated and projected onto current and potential planting suitability zones for the pathogen's hosts (Fig. 1B). These projection regions (Fig. 1B) were derived from the main hosts of the disease according to their edaphoclimatic requirements (Supplementary file 4), using layers of precipitation, mean temperature (Fick and Hijmans 2017), and elevation (NASA SRTM 2013) for Colombia. The open-source QGIS 3.2 software (QGIS Geographic Information System. QGIS Association 2021) was used for this purpose. In this process, areas with slopes greater than 60%, obtained from the digital elevation model (NASA EarthData), were discarded, and protected areas in Colombia were similarly excluded, along with areas where commercial agriculture is prohibited.

Results

Basic aspects associated with the regional distribution of *Plasmodiophora brassicae* in departments producing cruciferous crops in Colombia

According to the spatial location of the plots infested with *P. brassicae*, it was determined that they were in adjacent zones, resulting in regional clustering. This implied a local infection process among the plots, possibly caused by dispersion mechanisms such as the movement of plant material, machinery, and workers (Fig. 2). A total of 50 plots showed the presence of *P. brassicae*, the causal agent of clubroot in cruciferous crops, across all sampled departments. The distribution was as follows: Cundinamarca (35 sites; 16 positive), Antioquia (29; 5), Nariño (28; 1), Caldas (3; 2), Boyacá (10; 8), Norte de Santander (9; 8), Valle del Cauca (10; 8), and Cauca (3; 2). Antioquia and Nariño showed the lowest presence-to-sampling

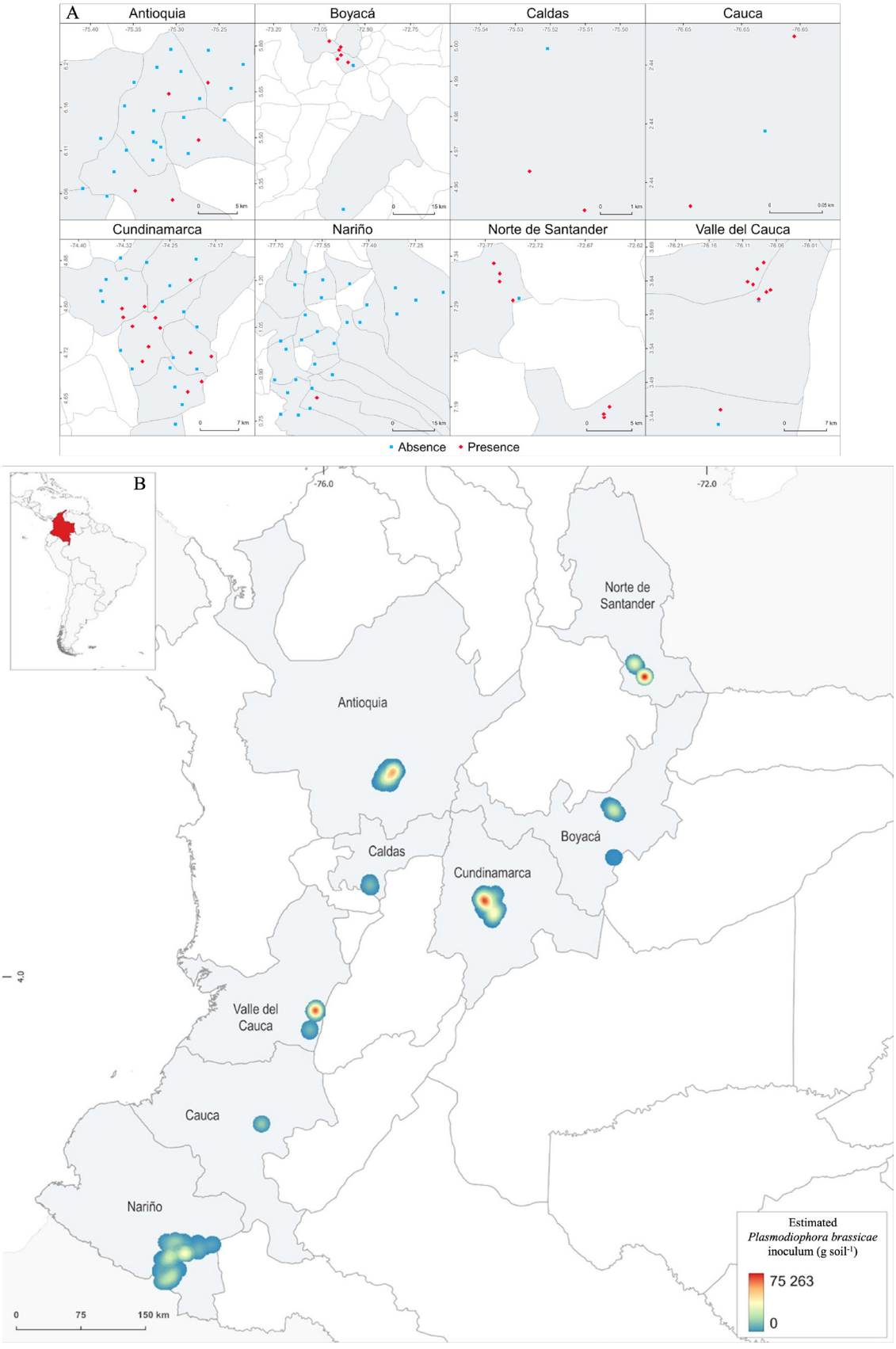

Fig. 2. Descriptive statistics and spatial risk visualization of the regional distribution of *Plasmodiophora brassicae* in the cruciferous producing departments of Colombia

A – spatial location in sampled cruciferous-producing areas of Colombia; B – relative risk of *P. brassicae* based on the quantity of inoculum associated with the cruciferous producing areas sampled in Colombia using the Kernel density method. The radius of the radial function of the method was determined on the basis of the potential dispersion process and the movement of the inoculum at the regional level through the movement of infected plant material and soil (machinery and workers)

ratio, while Boyacá, Cundinamarca, Norte de Santander, and Valle del Cauca had the highest prevalence (Fig. 2A). Regarding the presence of *P. brassicae* inoculum (soil resistance structures g · soil^{-1}), the estimation of inoculum density through the Kernel density analysis (Fig. 2B), indicated that the highest modeling risk values (~ 7.5×10^{-4}) were obtained in Norte de Santander, followed by Cundinamarca and Valle del Cauca, with values close to 7.0×10^{-4}, Antioquia (~ 6.0×10^{-4}), Nariño (~ 4.0×10^{-4}), Boyacá (~ 2.6×10^{-4}), Cauca (~ 1.5×10^{-4}) and finally Caldas (~ 1.0×10^{-4}) (Fig. 2B). This risk value is associated with the

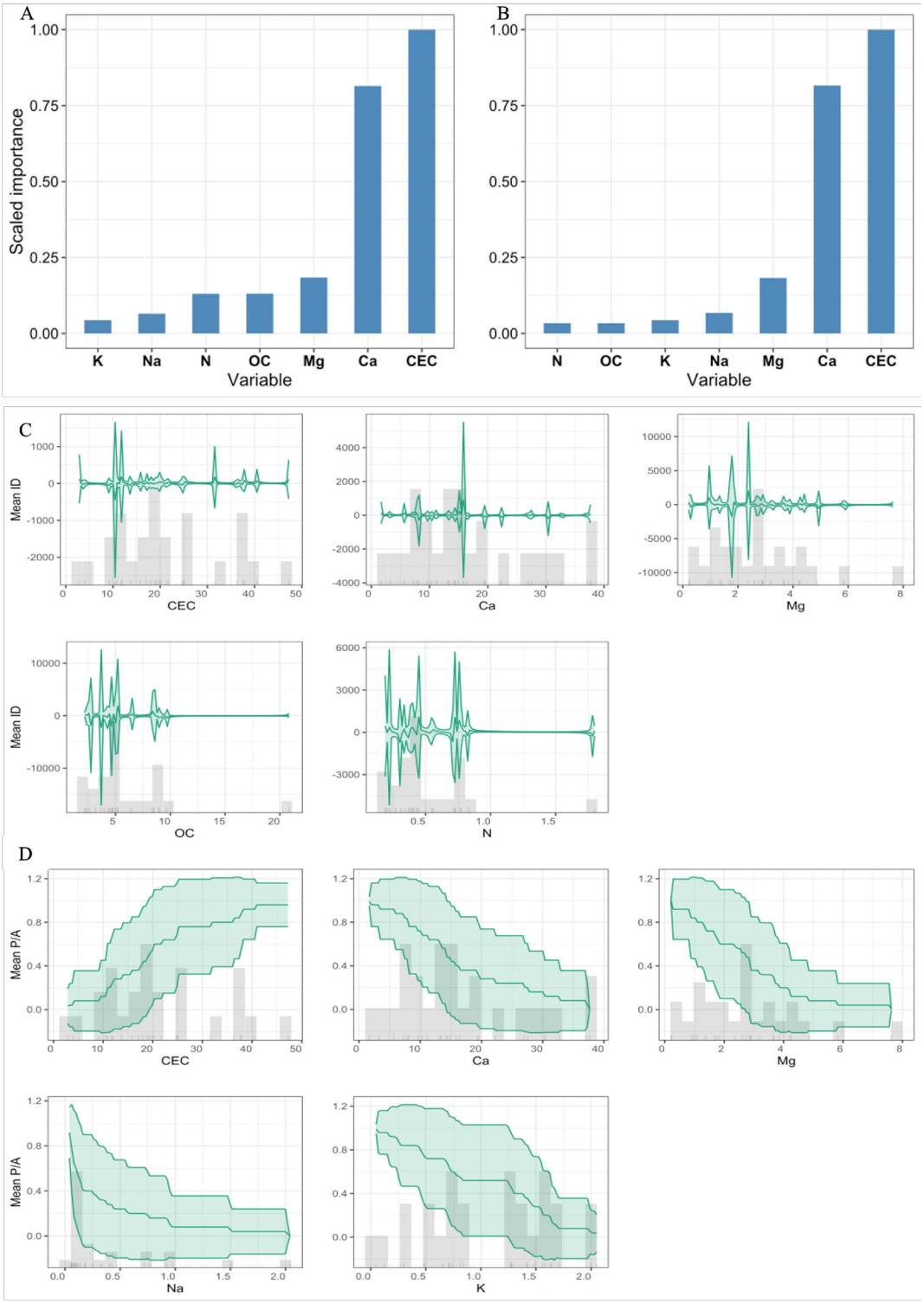

Fig. 3. Graphical representation of the standardized importance and partial dependence of inoculum density of the variables evaluated according to GLM for inoculum density – A–C and for the presence-absence data – B–D. Gray columns indicate the current data distribution for each predictor. The green curve represents the average change in the response variable

probability of finding the amount of inoculum, based on the distribution and value within the radius of the kernel used in relation to the movement of inoculum over short distances (e.g., soil movement by personnel, machinery, plant material, etc.).

These results reveal that the forecast areas generated by the Kernel density method demonstrated a concentration of *P. brassicae* inoculum on close local scales (spatial resolution at lot level and productive unit), particularly within infected plots and neighboring areas. This suggests slow and localized dispersion processes of the pathogen. Furthermore, the analysis highlights that once established in an area, high inoculum values may emerge, potentially impacting the presence of disease, especially when continuous cycles with susceptible hosts are conducted. Using this spatial analysis tool, restricted zones representing areas of high-risk for the occurrence of disease can be identified. By pinpointing these areas, appropriate management strategies can be implemented to mitigate the spread of *P. brassicae* and minimize its impact on susceptible hosts.

Epidemiological relationships of *Plasmodiophora brassicae* on a mesoscale and macroscale for Colombian conditions

Phase 1. Models to predict the presence and density of the *Plasmodiophora brassicae* inoculum

Based on the GLM selection and optimization process and selecting the best model (Tab. 1), it was statistically identified that the edaphoclimatic variables that best explain the quantity of inoculum and the presence/absence of *P. brassicae* (Fig. 3A–B, and Supplementary file 5) were the cation exchange capacity (CEC), calcium (Ca), and magnesium (Mg) content, with an approximate scaled importance of 100, 80, and 20%, respectively, for both response variables (Fig. 3A–B). At a more detailed level and with less importance, a causality relationship could be found for the quantity of inoculum with the content of organic carbon (OC) and nitrogen (N) at 12%, while for sodium (Na) and potassium (K) their importance was 5 and 3%, respectively (Fig. 3A and Supplementary file 5). For the presence/absence variable, the statistical relationships found, in decreasing order of importance, were soil contents of sodium (Na), potassium (K), organic carbon (OC), and nitrogen (N), with values of 6, 4, 3, and 3%, respectively. The remaining predictors of both models, such as the climatic variables, presented contributions below 1% (Fig. 3B and Supplementary file 5).

Furthermore, these models exhibited suitable performance metrics (Tab. 1), indicating accurate fitting and prediction, especially for the binomial model, which is capable of correctly characterizing absences and presences (Tab. 2) in the calibration set and predicting in an independent set of observations, with an accuracy of 93% and 81%, respectively.

By characterizing the magnitude of the coefficients of the models (Supplementary file 5) and the marginal response curves (Fig. 3C–D), it is possible to determine the effect of the predictors on the response variables. Initially, a low dependence of the inoculum density (ID) on CEC, Ca, Mg, and OC was observed, given by a magnitude of coefficients close to zero (Supplementary file 5) and predictions whose trend is constant across the range of values of each predictor. In

Table 1. Generalized Linear Models performance metrics associated with fit for the presence-absence and amount of inoculum of *Plasmodiophora brassicae* based on edaphoclimatic variables used as predicts

Response	Metrics	RMSE	MSE	Logloss	Gini	AUC	AUCPR	AIC
Presence-absence (P/A)[1]	value	0.25	6.3E-02	0.20	0.95	0.97	0.96	88.18
Amunt of inoculum (ID)[2]	metrics	RMSE	MAE	MSE		RMSLE	R^2	AIC
	value	3098.92	2047.3	9.6E+06		2.5	0.82	1571.24

[1]Binomial distributions adjusted; [2]Gamma distributions adjusted. RMSE – root mean squared error; MSE – mean squared error; AUC – area under the ROC curve; AUCPR – area under the precision-recall curve; AIC – akaike information criterion; MAE – mean absolute error; R^2 – coefficient of determination

Table 2. Prediction score on the set of trains and independent observations using the GLM binomial confusion matrix to predict the absence or presence of the pathogen

P/A	A		P		Error		Rate [accuracy]	
A	train	test	train	test	train	test	train	test
	48	12	2	4	0.040	0.25	2/50 (0.9)	4/16 (0.75)
P	3	0	27	5	0.100	0	3/30 (0.9)	0/5 (1)
Total	51	12	29	9	0.625	0.19	5/80 (0.93)	4/21 (0.81)

the case of OC and N, a more interpretable behavior was visualized when these variables have values above 10% and 0.8%, respectively, implying a complete decrease/reduction of the inoculum density to 0 resting spores × g^{-1} of soil.

Furthermore, the most important predictors showed a monotonic relationship with respect to the P/A variable (Fig. 3B, Supplementary file 5). In this sense, the model predicts an increase in pathogen presence as the CEC increases, with values greater than 40 cmol(+) · kg^{-1} representing a 100% probability of presence (Fig. 3B). Similarly, the model predicted a predisposition to absence as the Ca, Mg, K, and Na cation content increased (Fig. 3B). In this case, P/A showed a greater dependence on these predictor variables than ID, given the magnitude of the coefficients of the binomial model (Supplementary file 5) and the characterized change relationships in the marginal response curves (Fig. 3B).

Potential distribution of *Plasmodiophora brassicae* and clubroot in host-producing areas in Colombia based on ecological niche models

From the initial set of predictors, 14 environmental variables were discarded due to autocorrelation with temperature and precipitation, using a threshold of 0.8 (Supplementary file 2). Of the remaining 17 (Supplementary file 3), 7 variables were retained as a result of the iterative reduction of predictors with permutation importance below 5% (Fig. 4). These predictors were presumed to better explain the risk of the presence of *P. brassicae* and were distributed among the variables: soil (3) temperature (3), and precipitation (1). These variables were: mean diurnal range (BIO2), climatic moisture index (CMI), volumetric fraction of coarse fractures (CFVO), seasonality of temperature (BIO4), nitrogen content (N), soil cation exchange capacity (CEC) and mean temperature of the warmest quarter (BIO10). In terms of importance, the following decreasing order was observed: BIO2 and CMI, with contributions of 33.6% and 21.6%, respectively, followed by CFVO (10.7%) and BIO4 (10.4%), N (9.4%), CEC (7.6%), and BIO10 (6.6%). Similarly, these variables were considered to summarize the environmental variability explained by the predictors initially discarded, and their contributions depended to some extent on the model calibration parameters (Fig. 4A).

Starting from an initial process associated with good practices in ENM (multinomial niche estimation) in which spatial autocorrelation of presence points and collinearity of predictor variables were eliminated, a baseline model with a good fit to the data and adequate predictive capacity was obtained according to the evaluation metrics AUC (0.85 train, 0.73 test), TSS (0.72 train, 0.43 test), and AICc (1032.04). Subsequently, an optimization was achieved to assess the correct combination between the set of predictor variables (initially 16) and the algorithm parameters (FC and β value), resulting in a model with a combination of FC of type lph, a β of 0.65, and 7 predictor variables (Fig. 4). This parameterization presented the most adequate levels of AUC and TSS metrics (Fig. 4B–C), parameters higher than the baseline model, indicating that the improvement and optimization process was correct.

The final optimized model according to the AUC and TSS metrics for training and testing was 0.87–0.76 and 0.60–0.51, respectively. This indicated a model with good predictive capacity. It should be noted that these values were above the minimum level, indicating acceptable performance for classification algorithms according to these metrics (Fig. 4C). Similarly, according to the ROC curve (Fig. 4B), the relationship between true positive rate (TPR) and false positive rate (FPR) for both the train set and the test set indicated a model with an adequate classification level. On the other hand, the model presented an AICc of 694.42, indicating a low level of complexity, especially attributed to the reduction of predictors.

From the selection of the best MNE in the calibration process and their subsequent projection areas currently or potentially seeded with their hosts (Fig. 4C), a differential probability of risk of *P. brassicae* presence was presented, with higher risk zones represented in red and lower risk areas in blue. In this sense, the regions of Cundinamarca, Santander, Nario, Huila, Boyacá and Cauca corresponded to zones of high-risk presence, while the other production regions presented a lower risk (Fig. 4C). These areas represented only the environmental suitability for the presence of the pathogen and not the disease, in which case its development would be conditioned not only by suitable soil and climatic conditions, but also by the presence of a susceptible host.

Discussion

Our analysis aimed to approach the subject on two scales. On one hand, a plot-level analysis was conducted to identify the relationships between edaphoclimatic variables and epidemiological parameters associated with *P. brassicae*, such as presence-absence and inoculum quantity. The findings partially corroborated those of other authors (Botero *et al*. 2022a; Botero *et al*. 2022b), however, we proposed a more robust analysis based on GLMs with a detailed modeling

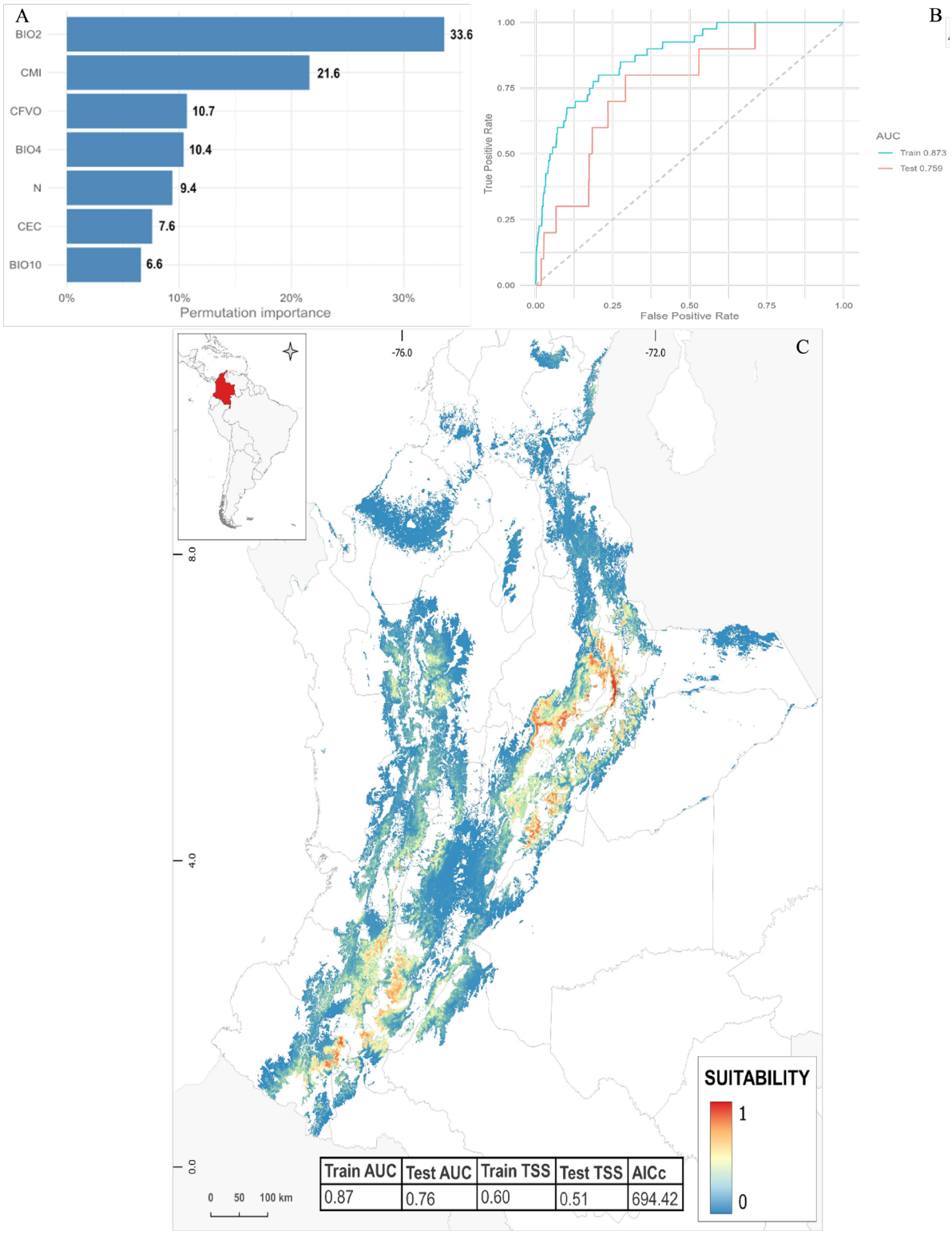

Fig. 4. Importance of the permutation and MaxEnt model of *Plasmodiophora brassicae* in Colombia according to the Maxent algorithm for environmental predictors with the greatest contribution to the risk of presence.
A – ROC curve for calibration and test data sets; B – continuous risk model for the presence of *P. brassicae* in environmentally suitable areas for cruciferous production; C – continuous risk model the presence of *P. brassicae* in areas with environmental suitability for crucifer production using ecological niche models approach

approach, aiming to meet all assumptions and ensure the reproducibility of the results. On the other hand, and as something novel in the present approach, a regional (meso-scale) epidemiological analysis was proposed, which indicates that this disease should also be viewed from this perspective. This approach aims to guide management measures, not only by avoiding planting in conducive climatic zones but

also by implementing a series of practices to prevent the spread of the inoculum.

Local causal factors (plot) derived from statistical analyses using GLM, a robust and flexible tool for epidemiological analysis, indicated how the behavior of *P. brassicae* was closely associated with edaphoclimatic factors in host-producing areas under tropical conditions, as in the case of Colombia. These results provided additional evidence that the physical and chemical properties of soils play an important role in the development of disease (Dixon 2009a).

For the disease quantification variables and epidemiological importance of *P. brassicae* such as P/A and ID, the effective cation exchange capacity (CEC) and the content of basic cations such as Ca, Mg, Na, and K represent the factors that contribute the most to both variables. When the relationships between these predictors and the response variable are characterized, it is observed that high values of the content of these cations imply a reduction in the probability of the pathogen's presence. This result is supported by the fact that soil alkalinity has been reported to be a suppressive edaphic factor for this disease (Webster and Dixon 1991).

In this regard, the content of Mg, Ca, Na, and K cations can create basic conditions in the soil, which affect the growth and development of *P. brassicae* by inhibiting spore germination and plasmodium maturation (Bhering *et al.* 2017; Webster and Dixon 1991). This explains the low dependence of ID on alkalinity and, generally, on the evaluated factors, since this response variable represents the content of resistance structures in the soil as the primary source of inoculum and does not imply a process of germination, colonization, and establishment (Botero-Ramirez *et al.* 2022a, b). In contrast, the relationship between CEC and P/A suggests that high values increase the probability of the pathogen's presence, which could be due to the underlying exchangeable acidity as a fundamental component of CEC.

According to risk models on the meso- and macro scales of the amount of inoculum and the presence of *P. brassicae* estimated using Kernel density approaches and MNEs (maximum likelihood estimators), a good approximation of this important epidemiological parameter and its representation in geographic space was achieved. These two analyses, which differ in terms of analysis scale, algorithms, and statistical foundations, were consistent with each other in showing that the risk of this disease is highly localized and strongly associated with plots and regions when high levels of inoculum are present under field conditions planted with hosts of this microorganism. This significant finding indicates that the spread of this disease occurs from plot to plot, possibly influenced by short-distance dispersal factors such as machinery, water, and personnel, which are involved in the movement of inoculum through soil and infected plant material (Cao *et al.* 2009; Rennie *et al.* 2011; Strelkov and Hwang 2014).

With macrolevel modeling, some causality relationships related to epidemiological behavior at the local level (mesoscale) were identified. In this sense, the predictors of greatest importance for the geographical distribution of the presence of *P. brassicae* were the mean diurnal range (BIO2) and the climatic moisture index (CMI). These variables explain variations in temperature and precipitation, respectively, and represent environmental factors reported as conducive to the pathogenic expression of *P. brassicae*. Temperature has been extensively studied as a parameter associated with diseases caused by this pathogen (Sharma *et al.* 2011; Dixon 2009a; Gossen *et al.* 2015). In terms of precipitation variables, rainy climates are conducive to increasing the movement of the pathogen's spores resting in the soil, and increases in humidity have a direct effect on disease favorability (Webster and Dixon 1991; Dixon 2009a).

Within a correct and epidemiologically sensible framework, good practices in the use of ENMs play a fundamental role in ensuring that this approach can produce results with biological and ecological significance and provide an accurate representation of the epidemiology of the disease studied (Ramírez-Gil *et al.* 2021). As demonstrated in this study, in ENMs, an optimization process is necessary with the aim of finding models that are as simple as possible with response curves that are straightforward for easy interpretation (Merow *et al.* 2013).

Under this criterion, an ecological niche model can be useful for developing intelligent management strategies, as potential areas of pathogen presence are established, which can subsequently lead to syndromes in specific crops. Thus, one management approach could involve exclusion, avoiding zones where conditions are not optimal for the pathogen's development (Ramírez-Gil *et al.* 2021).

Under the analytical proposal presented in this work, advances were made in the epidemiology of *P. brassicae* for better understanding of spatial aspects and related edaphoclimatic variables, at both local (plot) and regional (producer zones) scales. Additionally, a contribution was made in terms of analytical methods and their application exploring new approaches for spatial and temporal analyses and the behavior of plant diseases with respect to potential related variables. Based on the findings of this research and the limitations of this approach, it is believed that future work needs to be carried out to better understand the mechanisms of inoculum dispersion and movement at the regional level (machinery, plant material, workers, etc.), specifically to better understand the response of different genotypes in hosts, the

genetic variability of the pathogen, persistence in infected plots, the role of other hosts, and many other epidemiological parameters.

By understanding epidemiological aspects related to the spatial distribution of *P. brassicae,* the risk posed by the inoculum of this microorganism and its role in the development of the associated disease can be assessed. This knowledge enables the mitigation of various factors, such as the movement of personnel and machinery between infected plots, scheduling rotation systems in high-risk areas, and implementing management methods for pathogen populations based on antagonistic microorganisms, etc.

Conclusions

In the present work, a series of unconventional analysis methods were implemented and validated for the identification of soil variables conducive to the infection of *P. brassicae* in the host under tropical conditions. A risk analysis was carried out both on the mesoscale (plots) and on the macroscale (country), which was represented through maps that predicted the current and potential distribution of this pathogen in geographic areas suitable for host development in Colombia. The presented approach also suggested that the occurrence of this disease is highly clustered at the regional level, associated with plots with a high level of inoculum of its pathogen, and potential short-distance dissemination mechanisms.

References

Allouche O., Tsoar A., Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43 (6): 1223–1232. DOI: https://doi.org/10.1111/j.1365-2664.2006.01214.x

Barve N., Barve V., Jiménez-Valverde A., Lira-Noriega A., Maher S.P., Peterson A.T., Soberón J., Villalobos F. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling 222: 1810–1819. DOI: https://doi.org/10.1016/j.ecolmodel.2011.02.011

Bhering A.S., do Carmo M.G.F., Matos T.M., Lima E.S.A., Sabrihno N.M.B.A. 2017. Soil factors related to the severity of clubroot in Rio de Janeiro, Brazil. Plant Disease 101: 1345–53. DOI: https://doi.org/10.1094/PDIS-07-16-1024-SR

Bithell F. 1990. An application of density estimation to geographical epidemiology. Statistics in Medicine 9: 691–701. DOI: https://doi.org/10.1002/sim.4780090616

Botero-Ramírez A., García C., Gossen B.D., Strelkov S.E., Todd C.D., Bonham-Smith P.C., Pérez-López E. 2019. Clubroot disease in Latin America: distribution and management strategies. Plant Pathology 68 (5): 827–833. DOI: https://doi.org/10.1111/ppa.13013

Botero-Ramírez A., Padilla-Huertas F.L., García C. 2022a. Soil, climate, and management practices associated with the prevalence of clubroot in Colombia. Agronomía Colombiana 40 (2): 228–236. DOI: https://doi.org/10.15446/agron.colomb.v40n2.101461

Botero-Ramírez A., Padilla-Huertas F.L., Strelkov S.E., García-Domínguez C. 2022b. The occurrence of clubroot in Colombia and iIts relationship with climate and agronomic practices. Horticulturae 8 (8): 711. DOI: https://doi.org/10.3390/horticulturae8080711

Cao T., Manolii V.P., Strelkov S.E., Hwang S.-F., Howard R.J. 2009. Virulence and spread of *Plasmodiophora brassicae* [clubroot] in Alberta, Canada. Canadian Journal of Plant Pathology 31 (3): 321–329. DOI: https://doi.org/10.1080/07060660909507606

Cobos M.E., Peterson A.T., Barve N., Osorio-Olvera l. 2019. kuenm: an R package for detailed development of ecological niche models using Maxent. PeerJ 7: e6281. DOI: https://doi.org/10.7717/peerj.6281

Conrad O., Bechtel B., Bock M., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V., Böhner J. 2020. System for automated geoscientific analyses (SAGA) v. 2.1.4. Geoscientific Model Development 8: 1991–2007. DOI: https://doi.org/10.5194/gmd-8-1991-2015

DANE. 2020. Boletín técnico Encuesta nacional agropecuaria (ENA) 2019. [Online][Available from: https://www.dane.gov.co/files/investigaciones/agropecuario/enda/ena/2019/boletin_ena_2019.pdf] [Accessed 20 February 2020]

Diederichsen E., Frauen M., Ludwig-Müller J. 2014. Clubroot disease management challenges from a German perspective. Canadian Journal of Plant Pathology 36 (sup1.): 85–98. DOI: https://doi.org/10.1080/07060661.2013.861871

Dixon G.R. 2009a. *Plasmodiophora brassicae* in its Environment. Journal of Plant Growth Regulation 28 (3). DOI: https://doi.org/10.1007/s00344-009-9098-3

Dixon G. R. 2009b. The Occurrence and Economic Impact of *Plasmodiophora brassicae* and Clubroot Disease. Journal of Plant Growth Regulation 28 (3): 194–202. DOI: https://doi.org/10.1007/s00344-009-9090-y

Dobson A.J, Barnett A.G. 2018. An Introduction to Generalized Linear Models. 4th ed. Chapman and Hall/CRC. New York, USA. DOI: https://doi.org/10.1201/9781315182780

Donald C., Porter I., Porter I. 2009. Integrated control of clubroot. Journal of Plant Growth Regulation 28 (3): 289. DOI: https://doi.org/10.1007/s00344-009-9094-7

Elith J., Phillips S.J., Hastie T., Dudík M., Chee Y.E., Yates C.J. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17 (1): 43–57. DOI: https://doi.org/10.1111/j.1472-4642.2010.00725.x

Escobar L.E. 2020. Ecological niche modeling: an introduction for veterinarians and epidemiologists. Frontiers in Veterinary Science 7: 519059. DOI: https://doi.org/10.3389/fvets.2020.519059

Fick SE., Hijmans R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37: 4302–4315. DOI: https://doi.org/10.1002/joc.5086

Garrett K.A., Madden L.V., Hughes G. Pfender W.F. 2004. New applications of statistical tools in plant pathology. Phytopathology 94: 999–1003. DOI: https://doi.org/10.1094/PHYTO.2004.94.9.999

Gossen B.D., Strelkov S.E., Manolii V.P., Rennie D.C., Cao T., Hwang S.F., Peng, G., McDonald M.R. 2015. Spread of *Plasmodiophora brassicae* on canola in Canada, 2003–2014: Old pathogen, new home. Canadian Journal of Plant Pathology 37 (4): 403–413. DOI: https://doi.org/10.1080/07060661.2015.1105871

Hengl T., Mendes de Jesus J., Heuvelink G.B.M., Ruiperez M., Kilibarda M., Blagotić A., Shangguan W., Wright M.N., Geng X., Bauer-Marschallinger B., Guevara M.A., Vargas R., MacMillan R.A., Batjes N.H., Leenaars J.G.B., Ribeiro E., Wheeler I., Mantel S., Kempen B., Bond-Lamberty B. 2017. SoilGrids250m: Global gridded soil information based on machine learning. Plos One 12 (2): e0169748. DOI: https://doi.org/10.1371/journal.pone.0169748

Hill T.B., Daniels G.C., Feng J., Harding M.W. 2022. Hard to kill: inactivation of *Plasmodiophora brassicae* resting spores using chemical disinfectants. Plant Disease 106 (1): 190–196. DOI: https://doi.org/10.1094/PDIS-05-21-1055-RE

Hittorf M., Letsch-Praxmarer S., Windegger A., Bass D., Kirchmair M., Neuhauser S. 2020. Revised Taxonomy and Expanded Biodiversity of the Phytomyxea (Rhizaria, Endomyxa). Journal of Eukaryotic Microbiology 67 (6): 648–659. DOI: https://doi.org/10.1111/jeu.12817

Hwang S.F., Ahmed H.U., Zhou Q., Rashid A., Strelkov S.E., Gossen B.D., Peng G., Turnbull G.D. 2013. Effect of susceptible and resistant canola plants on *Plasmodiophora brassicae* resting spore populations in the soil. Plant Pathology 62 (2): 404–412. DOI: https://doi.org/10.1111/j.1365-3059.2012.02636.x

Javed M.A., Schwelm A., Zamani-Noor N., Salih R., Silvestre Vañó M., Wu, J., González M., Marten T., Luo C., Prakash P., Pérez- López E. 2023. The clubroot pathogen *Plasmodiophora brassicae*: A profile update. Molecular Plant Pathology 24 (2): 89–106. DOI: https://doi.org/10.1111/mpp.13283

LeDell E., Poirier S. 2020. H2O AutoML: Scalable automatic machine learning. p. 1–16. In: 7th ICML Workshop on Automated Machine Learning (2020). July 2020, USA.

McCullagh P. 2019. Generalized Linear Models. 2nd ed. Routledge, New York, USA. DOI: https://doi.org/10.1201/9780203753736

Madden L.V., Hughes G., van den Bosch F. 2007. The study of plant disease epidemics. p 421. St. Paul USA: American Phytopathological Society (APS Press). DOI: https://doi.org/10.1094/9780890545058

Merow C., Smith M.J., Silander J.A. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36: 001–012. DOI: https://doi.org/10.1111/j.1600-0587.2013.07872.x

Ministerio de Agricultura y Desarrollo Rural. 2023. Agronet: Área, producción y rendimiento nacional por cultivo. [Online] [Available from: https://www.agronet.gov.co/estadistica/paginas/home.aspx?cod=1] [Accessed 26 February 2023]

Molnar C. 2022. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2nd ed. p. 328. [Available on: christophm.github.io/interpretable-ml-book/] [Accessed on 9 October 2023]

NASA Shuttle Radar Topography Mission. 2013. 30 m Digital Elevation Model. [Online] [Available on: www.earthdata.nasa.gov] [Accessed: 9 October 2023]

Nelder J.A., Wedderburn R.W.M. 1972. Generalized linear models. Journal of the Royal Statistical Society: Series A. 135: 370–384. DOI: https://doi.org/10.2307/2344614

Nelson M.R., Orum T.V., Jaime-Garcia R., Nadeem A. 1999. Applications of geographic information systems and geostatistics in plant disease epidemiology and management. Plant Disease 83: 308–319.

Osorio-Olvera L., Lira-Noriega A., Soberón J., Peterson A.T., Falconi M., Contreras-Díaz R.G., Barve N. 2020. ntbox: An r package with graphical user interface for modelling and evaluating multidimensional ecological niches. Methods in Ecology and Evolution 11 (10): 1199–1206. DOI: https://doi.org/10.1111/2041-210X.13452

Phillips S., Anderson R., Schapire R. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190 (3–4): 231–259. DOI: https://doi.org/10.1016/j.ecolmodel.2005.03.026

Ramírez-Gil J.G., Morales-Osorio J. G., Peterson A.T. 2021. The distribution of *Phytophthora cinnamomi* in the Americas is related to its main host (*Persea americana*), but with high potential for expansion. Phytopathologia Mediterranea, 60 (3): 521–534. DOI:https://doi.org/10.36253/phyto-12327

Rennie D.C., Manolii V.P., Cao T., Hwang S.F., Howard R.J., Strelkov S.E. 2011. Direct evidence of surface infestation of seeds and tubers by *Plasmodiophora brassicae* and quantification of spore loads. Plant Pathology 60 (5): 811–819. DOI: https://doi.org/10.1111/j.1365-3059.2011.02449.x

Saupe E.E., Barve V., Myers C.E., Soberón J., Barve N., Hensz C.M., Peterson A.T., Owens H.L., Lira-Noriega A. 2012. Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. Ecological Modelling 237: 11–22. DOI: https://doi.org/10.1016/j.ecolmodel.2012.04.001

Shah D.A., De Wolf E.D., Paul P.A., Madden L.V. 2021. Accuracy in the prediction of disease epidemics when ensembling simple but highly correlated models. PLoS Computational Biology 17 (3): e1008831. DOI: 10.1371/journal.pcbi.1008831

Sharma K., Gossen B.D., McDonald M.R. 2011. Effect of Temperature on Cortical Infection by *Plasmodiophora brassicae* and Clubroot Severity. Phytopathology® 101 (12): 1424–1432. DOI: https://doi.org/10.1094/PHYTO-04-11-0124

Silverman B. 1998. Density Estimation for Statistics and Data Analyses. 1st ed. New York, USA.

Skelsey P. 2021. Forecasting Risk of Crop Disease with Anomaly Detection Algorithms. Phytopathology 111: 321–332. DOI: https://doi.org/10.1094/PHYTO-05-20-0185-R

Strelkov S.E., Hwang S.-F. 2014. Clubroot in the Canadian canola crop: 10 years into the outbreak. Canadian Journal of Plant Pathology 36 (sup. 1.): 27–36. DOI: https://doi.org/10.1080/07060661.2013.875338

Tewari J.P., Strelkov S.E., Orchard D., Hartman M., Lange R.M., Turkington T.K. 2005. Identification of clubroot of crucifers on canola (*Brassica napus*) in Alberta. Canadian Journal of Plant Pathology 27 (1): 143–144. DOI: https://doi.org/10.1080/07060660509507206

Timila R.D., Correll J.C., Duwadi, V.R. 2008. Severe and Widespread Clubroot Epidemics in Nepal. Plant Disease 92 (2): 317–317. DOI: https://doi.org/10.1094/PDIS-92-2-0317B

Title P.O., Bemmels J.B. 2018. ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. Ecography 41 (2): 291–307. DOI: https://doi.org/10.1111/ecog.02880

Torres E. 1972. Reacción de algunas crucíferas al ataque de *Plasmodiophora brassicae* Woronin en Manizales, Colombia. Acta Agronómica 22 (3–4): 185–207.

Wallenhammar A.C. 1996. Prevalence of *Plasmodiophora brassicae* in a spring oilseed rape growing area in central Sweden and factors influencing soil infestation levels. Plant Pathology 45 (4): 710–719. DOI: https://doi.org/10.1046/j.1365-3059.1996.d01-173.x

Vignali S., Barras A.G., Arlettaz R., Braunisch V. 2020. SDMtune: An R package to tune and evaluate species distribution models. Ecology and Evolution 10 (20): 11488–11506. DOI: https://doi.org/10.1002/ece3.6786

Webster M.A., Dixon G.R. 1991. Calcium, pH and inoculum concentration influencing colonization by *Plasmodiophora brassicae*. Mycological Research 95 (1): 64-73. DOI: https://doi.org/10.1016/S0953-7562(09)81362-2