Agriculture, Ecosystems and Environment 179 (2013) 25–34

Contents lists available at ScienceDirect

Agriculture, Ecosystems and Environment

journal homepage: www.elsevier.com/locate/agee

Environmental and social factors account for Mexican maize richness and distribution: A data mining approach

Carolina Ureta[a], Constantino González-Salazar[a,b,*,1], Edgar J. González[c],
Elena R. Álvarez-Buylla[d], Enrique Martínez-Meyer[a]

[a] Departamento de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico
[b] C3 – Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 México, D.F., Mexico
[c] Departamento de Ecología y Recursos Naturales, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico
[d] Laboratorio de Genética Molecular, Desarrollo y Evolución de Plantas, Instituto de Ecología, Universidad Nacional Autónoma de México, Centro de Ciencias de la Complejidad, Cd. Universitaria, Mexico City, Mexico

ARTICLE INFO

Article history:
Received 7 January 2013
Received in revised form 19 June 2013
Accepted 22 June 2013

Keywords:
Zea mays
Agrobiodiversity
In situ conservation
Center of crop origin and diversification
Ethnic groups
Data mining
Races distribution

ABSTRACT

Food security is a key topic for human welfare worldwide. In this context, the agrobiodiversity in centers of origin and diversification (COD) for the world's staple crops will be critical for feeding the world under changing environmental and social conditions. Maize is one of the most widely cultivated cereal and is the staple food for African and Latin-American countries, including its COD: Mexico, harboring more than 60% of the world's diversity for this crop. In this study we implemented a data mining approach that allowed us to evaluate spatial relationships of environmental (altitude, climate, slope and soil) and social factors (education and ethnic groups) with the spatial distribution of Mexican races, as well as the areas that can potentially harbor the highest number of races (PRA). In contrast to commonly used species distribution approaches, the data mining method implemented here allowed the integration of contrasting types of variables and their spatial relationships with the focal entity. Our results indicate that altitude, which is related with climate, was the factor with highest predictive power for most races. However, different factors showed different degrees of association with the spatial distribution of particular races. In any case, the performance of the model increased when using all evaluated factors. In our example study case, the highly vulnerable race Palomero Tolqueño was mainly influenced by climate, implying that climate change might threaten its preservation by reducing the areas with favorable conditions for its cultivation. Importantly, however, for this and other eleven races analyzed in detail, is that the ethnic group was the factor with the greatest predictive power. This finding further reinforces the key importance of *in situ* conservation by supporting local indigenous communities who are in charge of preserving, adapting to changing challenges and cultivating local races.

1. Introduction

The preservation of agrobiodiversity at Centers of Crop Origin and Diversification (COD) is critical for global food security (Thrupp, 2000; Esquinas-Alcazar, 2005). Such diversity is dynamically and communally generated at COD and hence, can only be preserved *in situ* (Bardsley and Thomas, 2005; Ortega Paczka, 2007). Therefore, an integrated understating of the role of diverse environmental and social factors on the distribution of individual races and the areas that can potentially harbor the highest number of races (PRA) becomes critical. Yet, such analyses are scarce or have not been conducted for the staple crops of the world (FAO, 2012).

Maize (*Zea mays* L.L.) is one of the most widely cultivated cereals in the world (Nuss and Tanumihardjo, 2010). It is the basic food of many African and Latin-American countries (CIMMYT, 2012), including Mexico, its COD, where around 59 maize races with thousands of varieties comprise 60% of the maize genetic diversity in the world (Ruiz Corral et al., 2008). Although the race unit is not commonly used in cultivated plants, it has been quite useful to systematize maize diversity. Anderson and Cutler (1942) defined a race as "a group of related individuals with enough characteristics in common to permit their recognition as a group".

Such maize diversification has responded to environmental and biotic conditions, but also to social variables since it was

---

* Corresponding author at: Ciudad Universitaria, 04510, Instituto de Biología. D.F., Mexico. Tel.: +52 56229169; fax: +52 55680847.
E-mail addresses: carolina.ureta@hotmail.com (C. Ureta), cgsalazar7@gmail.com (C. González-Salazar), emm@ib.unam.mx (E.J. González), edgarjgonzalez@ciencias.unam.mx (E.R. Álvarez-Buylla), eabuylla@gmail.com (E. Martínez-Meyer).
[1] Present address: University of California, Berkeley, College of Natural Resources, Department of Plant and Microbial Biology, 431 Koshland Hall, Berkeley, CA 94720, United States. Tel.: +1 510 642 6731/510 642 5601.

0167-8809/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.agee.2013.06.017

domesticated around 5000 years ago (Kato et al., 2009). Mexican maize production is highly dependent on environmental factors, but it is also influenced by complex and dynamic socio-cultural processes (Sánchez González and Goodman, 1992; Bellon and Brush, 1993; Brush and Perales, 2007; Kato et al., 2009; Bellon and Hellin, 2011). Indeed, 75–80% of maize cultivation depends on small-hold farmers (mestizos and indigenous peasants) using traditional methods (Brush and Perales, 2007). Consequently, Mexican maize diversity conservation should rely on criteria that incorporate natural and social factors. Despite the acknowledgment, a study providing an integrated analysis of both types of factors is lacking and is critical for future policies concerning the preservation and use of Mexican maize diversity upon a plethora of challenges faced by maize production and diversity conservation. Among such challenges, the alteration of the distribution of adequate regions for maize cultivation due to climate change is outstanding (Ureta et al., 2012). Also of importance is the loss of rare and critical races to maintain production under specific environmental threats, meeting specific cultural or social needs, preserving maize genetic diversity for future breeding programs, and preserving to avoid or diminish farmers' vulnerability under changing environmental, sociopolitical and economic conditions.

Here, we developed such study using a recently published data mining approach (Stephens et al., 2009; González-Salazar et al., 2013). This method allows the incorporation of biotic and abiotic variables and has been used to model the distribution of wild species, leading to more accurate distribution maps in which it is possible to evaluate the relative importance of the different variables considered, and which should be combined to better understand and explain the distribution of any given species. Under the same principle, we integrated the role of environmental and social factors to evaluate and model the spatial distribution of individual races and of the areas that can potentially harbor the highest number of races (PRA). In comparison with traditional species' distribution approaches that have been previously used (Jackson and Robertson, 2011; Loarie, 2008; Prates-Clark et al., 2008), data mining allows variables of contrasting nature and format to be combined and considered.

Specifically, we: (a) identify environmental and social factors associated with races potential distribution (commonly represented by regions with suitable environmental conditions; in this specific study we also incorporated social factors) and PRA (b) evaluate which of these factors have the highest predictive power of the races potential distribution areas, (c) create probability distribution maps of the potential distribution of individual races and PRA. With the aim of presenting the methodological process and determining the factors influencing the races current distribution, we provide a detailed analysis for an exemplar race. For this aim we have selected the vulnerable Palomero Tolqueño (Ureta et al., 2012) and include the results for this race in the main body of this paper. However, the data and analyses for the rest of the 46 evaluated Mexican races are presented in the Appendixes.

2. Methods

2.1. Data sources

We obtained a georeferenced Mexican maize database from the Mexican Commission for Biodiversity CONABIO (2011) (Acevedo et al., 2011). This database contains 50 years of data collection containing 21,848 records, but for the purpose of our analyses, we eliminated spatially duplicated recordsat the cell level (see below) for each race because they do not provide additional useful information, thus remaining 7949 unique records. Also, from the Mexican races identified in the database, only 47 were considered in this study due to existing taxonomic synonyms (Bofo = Elotes Occidentales and Chiquito = Nal-tel de Altura) and because the status of some races is still under debate.

We considered environmental and social factors and evaluated their influence on the distribution of each Mexican race and on the distribution of the areas that can potentially harbor the highest number of races (24–35 races) (PRA). We also identified the socio-environmental drivers of races and richness.

The environmental variables included in this study are: climate, soil type, altitude, and slope. For climate we used 19 bioclimatic variables that provide extreme, seasonal and annual temperature and precipitation patterns (Téllez et al., 2011), whereas altitude and slope came from the Hydro 1 k database (USGS, 2010). Both climatic and topographic variables were originally in raster format at 1 km resolution. We also incorporated 21 soil types countrywide in vector format at a spatial scale of 1: 250,000 (CONABIO, 1995), therefore, in order to analyze our information, categorical data were used in its native format, but continuous variables (climate, altitude and slope) were each discretized into ten categories (Appendix 1).

In terms of social factors, we evaluated the average years of education and ethnic groups. Ethnic groups were established based on the ethnic language spoken (68 different languages were identified) and years of education (INEGI, 2005) were taken as a proxy of overall socioeconomic level, as previously done by the Mexican government (INEGI, 2012). This factor was input as a categorical variable with five categories, ranging from unfinished basic education to bachelors or higher degree (Appendix 1). Ethnic groups were established base on the indigenous language spoken and we worked with 62 out of 68 languages that have been recognized because six of them had just one or two unique georeferenced data points (INALI, 2011), and were insufficient to carry out the analysis.

2.2. Evaluation of factors influencing the distribution of races

We used a recently published non-parametric data mining method (Stephens et al., 2009; González-Salazar et al., 2013) to find significant geographical or spatial associations between particular races and the factors being considered. This method is based on geographical co-occurrences of factors and a target entity: presence of a particular maize race or a unit area with more than 24 races. First, we divided the map of Mexico into a 10 km× 10 km grid (resulting in 20,719 cells); then, we assigned the presence/absence of each race to map cells and the corresponding value for all the associated factors. Such spatial association allowed us to calculate the probability of finding our race under specific conditions using the following formula:

$$P(A_i|I_k) = \frac{N_{Ai\&Ik}}{N_{Ik}} \tag{1}$$

where, $N_{Ai\&Ik}$ is the number of cells in which there is a co-occurrence of the distribution of the Mexican race A_i and factor $\mathbf{I}_k$; and $N_{\mathbf{I}k}$ is the number of cells in which the factor I_k is distributed.

However, $P(A_i|I_k)$ is a simple probability that does not take into account sample size which is important to determine significance. To incorporate the sample size and avoid bias in our results, we considered the following statistical test:

$$\varepsilon(A_i|I_k) = \left(\frac{N_{Ik}(P(A_i|\mathbf{I}_k) - P(A_i))}{N_{\mathbf{I}k}P(A_i)(1 - P(A_i))}\right)^{1/2} \tag{2}$$

Where $P(A_i)$ is the probability of finding our entity under study; $P(A_i|\mathbf{I}_k)$ the probability of finding the latter when factor I_k is present; and $N_{\mathbf{I}k}$ is the number of cells in which the factor I_k is distributed. This method assumes a normal distribution of epsilon values (ε) and uses as a critical value $\varepsilon(A_i|I_k) \geq 2$, which evaluates the statistical dependency between A_i and I_k relative to the null hypothesis that the distribution of A_i is independent of $\boldsymbol{I}_k$ and is randomly distributed over the grid. The critical value of 2 represents 2 standard deviations (95% confidence interval) (Stephens et al., 2009; González-Salazar et al., 2013).

2.3. Calculating score values to create probability maps

We created probability maps for each race by calculating score values $S(A_i|\boldsymbol{I}')$through a Bayes approximation:

$$S(A_i|\boldsymbol{I}\prime) = \sum_{k=1}^{N} S(A_i|I_k) = \sum_{k=1}^{N} \ln\frac{P(I_k/A_i)}{P(I_k/B_i)}, \qquad (3)$$

Where A_i is the number of cells in which maize is present; B_i is the number of cells in which maize is absent; I_k is the number of cells in which a factor is present; and $S(A_i|\boldsymbol{I}')$ is a measure of the probability of finding the distribution of A_i when niche profile is $\boldsymbol{I}'$. $P(I_k|A_i)$ is the probability of finding A_i when factor I_k is present; and $P(I_k|B_i)$ is the probability of not finding A_i when factor I_k is present (Stephens et al., 2009; González-Salazar et al., 2013).

We ranked the score values of each cell for all factors together; and divided them in deciles, where the 10th decile represented the highest probability to find a specific race. To demonstrate the different contribution of factors to the distribution of each race, we ranked race's score values for each factor individually.

Finally, we named as the suitable conditions the socio-environmental factors in the 10th decile that presented a spatially significant relationship with the distribution of races' areas.

2.4. Evaluation of factors' influencing race richness

To evaluate factors influencing PRA we firstly had to identify those areas by overlaying the individual probability maps of all races. Individual races maps had to be transformed into binary maps (presence/absence) in order to be added by taking the minimum score value where the race was present as a threshold value. We categorized richness potential distribution areas in 10 classes, with 24–35 races being the richest category (PRA).

Once we recognized PRA, we evaluated $\varepsilon(A_i|I_k)$ and $S(A_i|\boldsymbol{I}')$ to find the suitable socio-environmental conditions that better explained the distribution of such areas. In other words, we identified the factors spatially associated with PRA.

2.5. Validation

To evaluate the predictive capacity of the factors under study, we performed a cross-validation test.This procedure uses part of the original data to fit the model with a specific factor and the rest of the data is used to evaluate its ability to correctly predict them. In our case, we randomly split the occurrence records of maize races into training (70%) and validating (30%) datasets. The training data were used to calculate score values associated to each factor via a Bayes approximation (cross-validation). Each race occurrence was assigned to a cell, and cells were sorted according to their associated score value (that depends on the factor evaluated) and divided into deciles. Good models are expected to have most of their validation points in the higher deciles. Next, we counted how many validation cells were found in the top decile. This procedure was repeated 100 times for each race to obtain the mean of validation cells on each decile and its associated standard deviation. Finally, to determine the importance of factors for model predictability, we carried out a t-test for independent samples between the numbers of occurrences predicted by different factors. This procedure only took place for individual races, because the richness map was constructed only by the addition of these individual maps.

3. Results

3.1. Influence of environmental factors on the distribution of areas that can potentially harbor the highest number of races and of individual races

Areas with the potential of holding 24–35 races were considered the richest (PRA). In these areas we found significant spatial associations with environmental and social factors ($\varepsilon > 2$): altitudes ranging 1000–4952 m.a.s.l., slopes 3.43–30.7°, and 7 different soil types (Regosols, Rankers, Planosol, Gleysol, Chernozem, Cambisol and Acrisol) were identified as the environmental factors with higher influence on PRA distribution (Table 1). The climatic factors significantly associated to the spatial distribution of PRA could be divided into four main groups: (a) annual climatic factors, (b) climatic factors related to spring-summer season, (c) extreme climatic factors and (d) factors representing climatic variability.

Among the annual climatic factors associated with PRA, we found annual mean temperatures between 12.5 and 17.6 °C and annual precipitations between 982 and 1980 mm to be the most important. In terms of seasonal climatic factors, we found that these areas were related to mild-hot temperatures between 9.3 and 24.4 °C and precipitations between 234 and 1222 mm in the warmest and wettest quarter of the year, and precipitation of 19 to 65 mm and temperature of 15.2–21.6 °C during the coldest and driest quarter of the year, respectively. Additionally, precipitation between 201–376 mm (from a range of 9–819 mm) in the wettest month, temperatures not higher than 32.7 °C with a minimum of 12.8 °C in the warmest month, and a minimum temperature between 4.7–12 °C in the coldest month were important extreme climatic factors (Table 1). Finally, factors regarding climatic variability also played an important role, since all of them were significantly associated with the areas that can potentially harbor the highest number of races (Table 1).

Factors that best predicted the presence of individual races varied among races (Table 2). However, Fig. 1b shows that altitude is the environmental factor with highest predictive power for most of the races, followed by slope, soil and climate (see also Table 2). For the race analyzed in detail, Palomero Toluqueño, altitude was the best predictor followed by climate, slope and soil; whereas in other cases, e.g., Elotero de Sinaloa, soil type was the best predictor followed by slope, altitude and climate.

3.2. Influence of social factors on the distribution of areas that can potentially harbor the highest number of racesand of particular races

The level of education was associated with the distribution of PRA, as long as a high-school or lower educational level was being considered. In addition, the presence of 25 different ethnic groups was associated with such areas (see Table 1).

Interestingly, when evaluating the presence and distribution of each race individually, our results show that 12 races (Apachito, Blando, Chiquito, Dulce, Dulcillo, Nal-tel, Olotón, Palomero Toluqueño, Serrano Mixe, Tabloncillo Perla, Zamorano Amarillo, and Zapolote Grande) were better predicted by the presence of a particular ethnic group than by any of the environmental factors (Table 2). For example, the Purepecha was the only ethnic group significantly associated with the presence of Zamorano Amarillo and was also that best explained the distribution pattern of this race.

In terms of individual maize races, we found significant spatial relationships with different levels of education, going from unfinished elementary school to finished high-school. Bachelor or higher degrees of education was not found in the suitable profile of any race (Appendix 2). Additionally, 49 out of the 62 ethnic groups were also significantly associated with particular races (Appendix 2). The ethnic groups more frequently present in suitable profiles of races were Zapoteco, Tojolobal, Otomí, Mixteco, Chatino, Cuicateco, Huave, Mam, Mazateco, Mixe, and Náhuatl. Interestingly, Zapoteco was in the suitable profile of 19 races, being the most frequent ethnic group in the suitable profiles, suggesting that Zapoteco cultures have been involved in the diversification of many races that are still preserved in Mexico.

3.3. Probability maps

Areas supporting high maize richness, as identified by the probability maps, included the southern states of Chiapas, Oaxaca, Guerrero, the western state of Jalisco, and central Mexico. Nevertheless, suitable areas for at least one maize race cover practically the entire Mexican territory (Fig. 2).

Moreover, the probability distribution map of Palomero Tolqueño serves to show the contribution of individual factors to the distribution model (Fig. 3). Occurrences in the highest probability areas of Palomero Toluqueño for each factor map suggest that ethnic group, altitude and climate are the factors with highest predictive power. Additionally, predictive power increases when all factors are used simultaneously; increasing the number of occurrences predicted in the highest decile. Probability maps for each race using all variables are available in Appendix 3.

4. Discussion

There has been a longstanding interest in understanding the effect of various natural and social factors in explaining the

Table 1
Factors significantly associated with Mexican potential maize-rich areas (24–35 races).

Factors	[a]Category	[b]Range	[c]Epsilon	[d]Score
Environmental				
Altitude	1041–4942 m.a.s.l.	0–5500 m.a.s.l.	8.019	2.146
Annual mean temperature	12.5–17.6 °C	7.8–29.1 °C	3.858	0.841
Annual precipitation	982–1980 mm	55–4517 mm	6.388	1.236
Isothermality	65–76	38–85	8.861	1.678
Max temperature of warmest month	13.8–32.7 °C	138–425 °C	6.449	1.519
Mean diurnal range	120–163	60–205	3.768	0.806
Mean temperature of coldest quarter	15.2–21.6 °C	3.2–28.2 °C	5.51	1.113
Mean temperature of driest quarter	16.3–20.2 °C	4.6–29.3 °C	5.704	1.074
Mean temperature of warmest quarter	9.3–24.4 °C	9.3–32.5 °C	5.588	1.263
Min temperature of coldest month	4.7–12 °C	138.425 °C	5.677	1.067
Precipitation of coldest quarter	40–65 mm	1–823 mm	3.519	0.562
Precipitation of driest month	6–12 mm	0–145 mm	4.91	0.765
Precipitation of driest quarter	19–35 mm	0–485 mm	6.423	0.954
Precipitation of warmest quarter	234–556 mm	3–1475 mm	4.528	0.931
Precipitation of wettest month	201–376 mm	9–819 mm	7.676	1.473
Precipitation of wettest quarter	525–1222 mm	21–2187 mm	6.673	1.307
Precipitation seasonality	88–102	29–129 mm	5.985	1.155
Slope	3.43–30.7°	0–34.12°	8.236	1.712
Soil type	Regosol, Ranker, Planosol, Gleysol, Chernozem, Cambisol, Acrisol		3.498	1.381
Temperature annual range	16.7–23.8 °C	12.1–39.4°	7.133	1.254
Temperature seasonality	379–2021	379–9348	11.418	1.885
Social factors				
Ethnic group	Akateko, Cuicateco, Chatino, Chocolteco, Cora, Mixe, Huave Jakalteko, Mam, Mazahua, Mazateco, Mixteco, Náhuatl, Otomí, Purepecha, Qanjobal, Teko, Tlahuica, Tlapaneco, Tojolabal, Triqui de la Baja, Tsetsal, Tsotsil, Zapoteco, Zoque		4.515	2.076
Education	unfinished elementary-finished high school	unfinised elementary-bachelor or higher degree	5.752	0.831

[a] Category = factors' category that presented a significant relationship with the richest areas (24–35 races).
[b] Range = the range of the factor evaluated.
[c] Epsilon = the epsilon value resulted from the spatial interaction between the richest areas (24–35 races) and the factors evaluated.
[d] Score = factors contribution to maize presence.

patterns and extents of distribution of particular races and landraces of maize (Sánchez González and Goodman, 1992; Bellon and Brush, 1993; Brush and Perales, 2007; Kato et al., 2009; Bellon and Hellin, 2011). These studies have suggested that environmental factors are the main driving forces of Mexican maize diversification. Nonetheless, previous studies have also shown the importance of seed selection, landraces yield, market and culture on maize diversity conservation.

However, as far as we know, our study is the first attempt to evaluate the relative importance of social and environmental factors on the distribution of Mexican maize. Despite the fact that we were able to consider many more environmental factors in

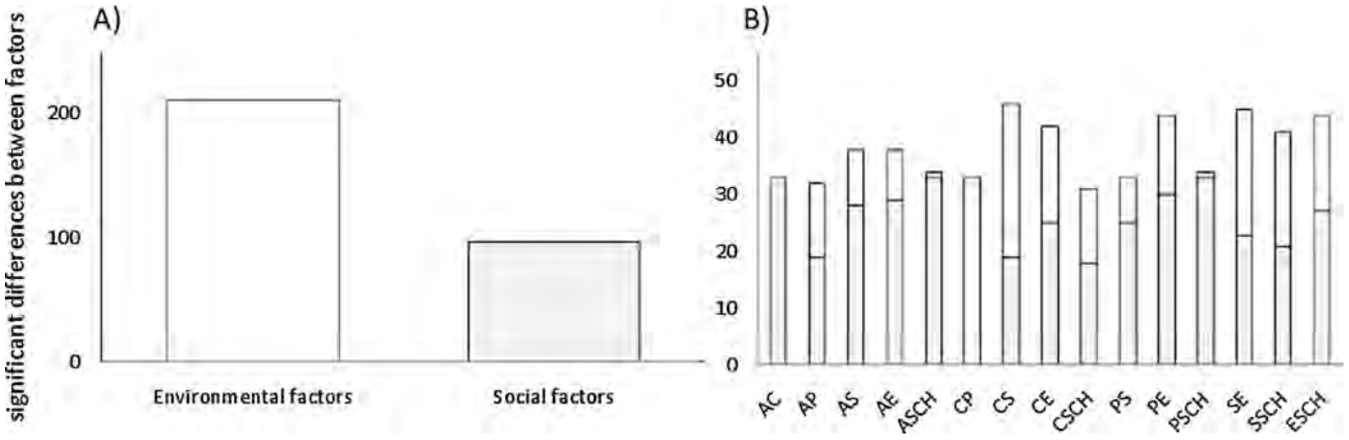

Fig. 1. Significant predictive differences between factors for all Mexican maize races distribution areas. (A) Number of significant differences between all factors evaluated for all races. Environmental factors = number of significant differences in which the environment (altitude, climate, slope and soil) was a better predictor than social factors (ethnic group and education). Social factors = number of significant differences in which the social factors were better predictors. (B) Number of significant differences in all races. Gray = times in which the first factor was significantly better predictor than the second. White = times in which the second factor was a significantly better predictor than the second factor. AC = altitude vs. climate, AP = altitude vs. slope, AS = altitude vs. soil, AE = altitude vs. ethnic group, ASCH = altitude vs. education, CP = climate vs. slope, CS = climate vs. soil, CE = climate vs. ethnic group, CSCH = climate vs. education, PS= slope vs. soil, PE = slope vs. ethnic group, PSCH = slope vs. education, SE = soil vs. ethnic group, SSCH = soil vs. education and ESCH= ethnic group vs. education.

Table 2
T-tests between predictability of variables for all races evaluated.

	A vs. C	A vs. P	A vs. S	A vs. E	A vs. SCH		C vs. P	C vs. S	C vs. E	C vs. SCH	P vs. S		P vs. E	P vs. SCH	S vs. E	S vs. SCH	E vs. SCH
Ancho	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Ancho	NS	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	Ancho	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Apachito	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Apachito	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Apachito	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Arrocillo Amarillo	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Arrocillo Amarillo	$P\le 0.05$ (−)	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (−)	Arrocillo Amarillo	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Azul	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	Azul	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	Azul	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)
Blando	NS	NS	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	Blando	NS	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	Blando	$P\le 0.05$ (−)	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Bofo	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Bofo	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	Bofo	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)
Bolita	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Bolita	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Bolita	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)
Cacahuacintle	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Cacahuacintle	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Cacahuacintle	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Celaya	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Celaya	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	Celaya	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)
Chalqueño	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Chalqueño	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Chalqueño	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)
Chapalote	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	NS	Chapalote	NS	$P\le 0.05$ (+)	NS	NS	$P\le 0.05$ (+)	Chapalote	NS	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	NS
Chiquito	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Chiquito	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Chiquito	$PP\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Comiteco	NS	NS	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	Comiteco	$P\le 0.05$ (−)	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (−)	Comiteco	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Conejo	NS	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Conejo	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Conejo	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Cónco	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Cónco	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	Cónco	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)
Cónico Norteño	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	NS	Cónico Norteño	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Cónico Norteño	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)
Coscomatepec	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Coscomatepec	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Coscomatepec	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Cristalino	NS	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	Cristalino	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	Cristalino	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)
Dulce	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Dulce	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)		Dulce	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Dulcillo	NS	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	Dulcillo	NS	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	Dulcillo	$P\le 0.05$ (−)	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Dzit-bacal	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Dzit-bacal	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	NS		Dzit-bacal	$P\le 0.05$ (−)	NS	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Elotero de Sinaloa	$P\le 0.05$ (+)	NS	NS	NS	$P\le 0.05$ (+)	Elotero de Sinaloa	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	NS	Elotero de Sinaloa	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Elotes Cónicos	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	Elotes Cónicos	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	Elotes Cónicos	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Gordo	NS	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	Gordo	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	Gordo	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	NS
Jala	NS	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Jala	NS	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Jala	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Mushito	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Mushito	$P\le 0.05$ (−)	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Mushito	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Nal-tel	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	Nal-tel	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	NS	NS	Nal-tel	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Olotillo	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Olotillo	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	NS	NS	Olotillo	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)(−)
Olotón	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	Olotón	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Olotón	NS	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Onaveño	NS	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	Onaveño	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Onaveño	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$PP\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)
Palomero Toluqueño	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	Palomero Toluqueño	NS	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Palomero Toluqueño	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Pepitilla	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	NS	Pepitilla	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	Pepitilla	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)
Ratón	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	Ratón	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	NS	NS	Ratón	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)
Reventador	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Reventador	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	Reventador	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Serrano	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Serrano	NS	$\mathrm{P}\le 0.05$ (+)		$\mathrm{P}\le 0.05$ (+)	$\mathrm{P}\le 0.05$ (+)	Serrano	$\mathrm{P}\le 0.05$ (−)	$\mathrm{P}\le 0.05$ (+)	$\mathrm{P}\le 0.05$ (−)	$\mathrm{P}\le 0.05$ (−)	$P\le 0.05$ (+)
Tablilla de Ocho	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Tablilla de Ocho	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Tablilla de Ocho	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Tabloncillo	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)	$P\le 0.05$ (+)	NS	Tabloncillo	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)	Tabloncillo	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)
Tabloncillo Perla	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	Tabloncillo Perla	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	NS	Tabloncillo Perla	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Tehua	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Tehua	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Tehua	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Tepecintle	NS	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Tepecintle	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Tepecintle	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Tuxpeño	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Tuxpeño	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	NS	NS	Tuxpeño	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)
Tuxpeño Norteño	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Tuxpeño Norteño	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	Tuxpeño Norteño	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)
Vandeño	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	NS	Vandeño	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	NS	Vandeño	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	NS	$P\le 0.05$ (−)
Zamorano Amarillo	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Zamorano Amarillo	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Zamorano Amarillo	$P\le 0.05$ (−)	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)
Zapalote Chico	$P\le 0.05$ (+)	$P\le 0.05$ (+)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	Zapalote Chico	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	NS	$P\le 0.05$ (−)	Zapalote Chico	$P\le 0.05$ (−)	NS	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)
Zapalote Grande	NS	NS	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	Zapalote Grande	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (−)	Zapalote Grande	$P\le 0.05$ (−)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (+)	$P\le 0.05$ (−)

A= altitude, C= climate, P = slope, S= soil, E= ethnic group, SCH= education, (+) =significant difference between the means of the number of predicted validation points; being the mean of the first variable higher than the second variable, (−) = significant difference between the means of the number of predicted validation points; being the mean of the second variable higher than the mean of the first variable, NS= no significant.

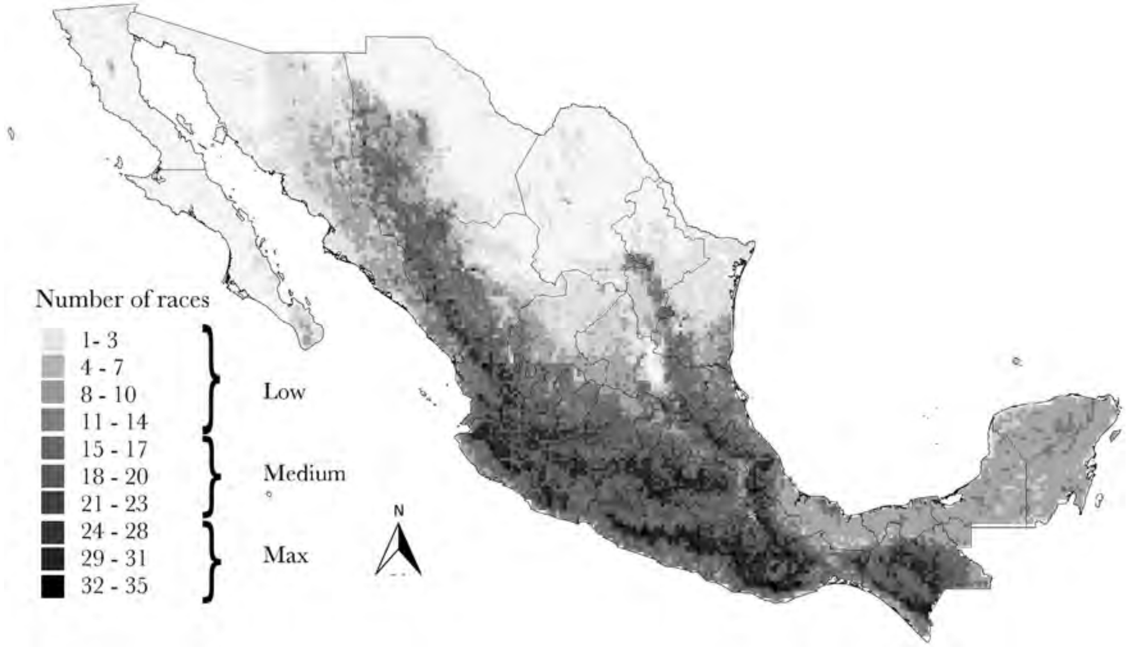

Fig. 2. Maize-rich areas distribution map. Probability map created by adding all Mexican maize race distribution areas. The darker areas in the map represent higher richness.

comparison to social ones, the data mining approach that we have used in this paper has enabled us to identify which are the most informative variables for the distribution of individual races and to gain a better understanding of the factors geographically associated with the areas that can potentially harbor the highest number of races (PRA). However, as in all modeling methods, quality of data does affect the analysis. We are aware that the quality of the data sets for the social factors is inferior to that for the environmental factors, and this likely affected our analyses. Still, the ethnic group factor results to be critical in order to explain the actual distribution of individual maize races, and the potential areas of highest richness in Mexico.

4.1. Methodological issues

Although it is a known fact that scale and resolution of spatial analysis have an influence on the analysis (Peterson et al., 2011), their real impact on results strongly depends on the question to be answer. In our analyses we used a 10 km × 10 km grid to balance between fine-scale social and environmental relationships with maize races, and the coarse-grained countrywide scope of the study. Furthermore, the method implemented in this paper has already been tested with different cell sizes by Stephens et al. (2009) (5, 10, 50, and 100 km) and found that although absolute epsilon values do change as a function of cell size (the number of co-occurrences changes), relative epsilon values remained stable. In other words, the relationship pattern was maintained independently of the cell size. Similarly, Kok and Veldkamp, 2001 evaluated the effects of different spatial resolutions (15 to 75 km) on land use patterns in Central America and found that resolution did not significantly affect the spatial relationship between land use patterns and their determinant factors.

On the other hand, our validation consisted in a cross-validation procedure in which we randomly split our data in a 70:30 proportion, repeating the process 100 times. A cross-validation would ideally use 50% of data to calibrate (fit) the model and 50% to validate it. However, in our case we had races with 18 occurrences, thus carrying out the model with only 9 points implies using a reduced number of degrees of freedom to fit the model. Yet, too few validation data might not let us see how well the model is performing. We considered that the way we split our data represent a good balance between the data used for fitting the model and for validating it. We kept this split ratio throughout the races for the purpose of comparability and because such ration was good enough for analyses with greater amount of data.

Finally, we want to clarify that we created the PRA map, by adding the modeled distributions for individual races, rather than using the actual data for race distribution, thus the resulting map overestimates the number of races per area. For example, we found areas that can potentially hold up to 35 races, when in the field the actual highest number of races found per unit is 22. We decided to add the expected presence of individual races, instead of modeling race richness, because we consider that maize richness is not an entity that responds to environmental or social factors (i.e., it does not have a socio-ecological niche, although it can be spatially related to some social and environmental factors). The latter becomes clear when we compare two equally rich areas that have a contrasting set of races. Consequently, we decided to use the maps for the individual races that have a socio-ecological niche and added such data only to gain a better insight of which factors mostly associated spatially to the areas of highest richness, independently of the races that are found within them. Therefore, for the race-richest areas, ε values represent only geographical associations with factors, and not its niche profile as it was considered for the analysis of each individual race.

4.2. The role of environmental factors on the distribution of areas that can potentially harbor the highest number of races and of particular races

Even though our PRA distribution map is similar to the one presented by Ureta et al. (2012), where only climatic factors were used under an ensemble niche modeling procedure, the present study indicated that climate was not the most important predictor compared to other environmental factors. Nonetheless, altitude, which resulted to be the best predictor in this study, is strongly related to climatic factors, and consequently our maps largely coincide with those modeled with climatic factors only. Altitude is actually an environmental variable that summarizes several climatic factors

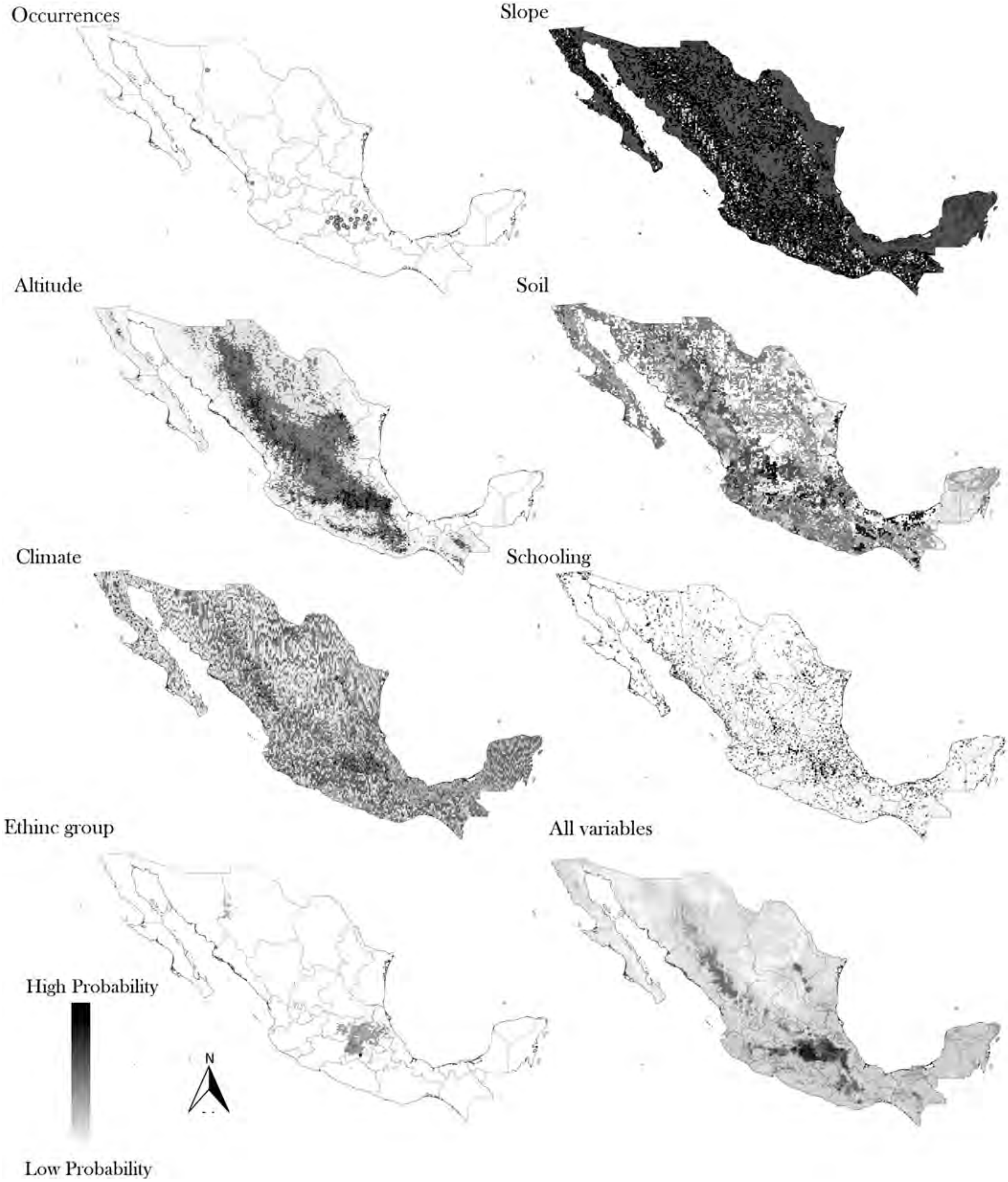

Fig. 3. PalomeroToluqueño's case.Probability maps created by each factor evaluated (altitude, climate, slope, soil, ethnic group; and education) and all together.

acting simultaneously and consequently it might be veiling the important role of climate on the distribution of PRA. Still, the importance of altitude on the distribution of maize is undeniable and has been acknowledge before (Sánchez González and Goodman, 1992; Brush and Perales, 2007).

In terms of climate, we found that PRA is associated with mild temperatures and intermediate precipitation. Such climatic conditions at a given geographical location and their distribution are expected to change in the near future as temperature and rain patterns are altered due to global climate changes (Conde et al., 2008).

When evaluating slopes, such areas were found at 3.43–30.7°. Most Mexican maize-races are distributed along the Mexican mountain ranges. Additionally, Mexican states with highest richness also present very abrupt landscapes. Therefore, adaptation of several races to mid altitudes and slopes can be expected and such areas are among the richest ones for native maize. Finally, the soil types to which PRA is more frequently associated with, include soils of tropical and subtropical areas (Gleysols); dry or very cold areas (Regosols); those that are generally very suitable for agriculture (Cambisols and Chernozems), moderately suitable (Planasols and rankers), or not suitable at all (Acrisols) (FAO, 2012). These results suggest that maize races have diversified in a wide variety of soil types, and have been bred to thrive even in marginal soil conditions.

Compared with more conventionally used distribution modeling methodologies, the data-mining approach used in this study has identified which particular factors and respective categories are significantly associated with the geographic distribution of each one of most Mexican races (Appendix 2). For example, for Palomero Tolqueño it was shown that the environmental factors with highest predictive power are not the same as the best predictors for most races. In this particular case, climate type became the second best predictor instead of the worst. Furthermore, the presence of this race is strongly associated with mild annual temperatures, low annual precipitation, mild temperature during the hottest month, temperatures above 0 °C during the coldest month, low to medium precipitation in the summer-spring season and very low precipitations in the autumn-winter season, as well as with strong isothermality and temperature seasonality (Appendex 3). Its strong association with climatic variables leads to postulate that a temperature increase due to climate change may become an important threat for this race.

Most individual races were better predicted by environmental variables (Fig. 1 and Table 2), however; using social factors to produce the probability maps increased the predictability of distribution maps as shown with Palomero Toluqueño (Fig. 3), for which all factors together gave rise to a more robust predictive distribution map than individual factors. Similar results were shown by González-Salazar et al. (2013) where the addition of biotic variables into abiotic distribution maps increased predictability and reliability of distribution models.

4.3. The role of social factors in the distribution of areas that can potentially harbor the highest number of races and of particular races

For PRA, ε values indicated that four educational level categories and 25 ethnic groups showed significant associations. The only educational level category that was not significantly related with PRA was bachelors or higher degree levels, indicating that most highly educated people are not likely to grow or produce native maize, but farmers with some level of education seem to be more likely involved in crop production, than those who have not received any education. This result is expected because most highly educated persons live in large cities and are not involved in primary production (INEGI, 2012); otherwise, they rather practice large-scale production, (INEGI, 2012).

In contrast, there is a close relationship between the richness of maize races and ethnic groups in Mexico. PRA comprises around 6.9% of the Mexican territory and is associated to areas where more than 40% of the ethnic groups of Mexico live. The results presented here agree with Boege (2010), who also showed that the presence of indigenous communities of diverse ethnic groups are closely associated and have been key for creating and preserving a dynamic diversification of maize races. However, as seen in Brush and Perales (2007), “mestizo” rural communities have also played an important role in the *in situ* generation and conservation of native maize varieties, and they continue to be actively involved in such processes. It is noteworthy, that both types of peasants, indigenous and “mestizo,” share the same type of relatively low formal educational level, and generally rely on traditional technological practices, rather than industrialized ones (Bellon et al., 2005). Therefore, the preservation of maize native diversity relies on smallholder farmers, as well as on the cultural diversity associated to different ethnic groups that are strongly associated to specific maize races. Until recent extension programs, modern technology has not reached or aimed at Mexican smallholder farmers needs (Bellon et al., 2005). Exceptions are some studies by Mexican agronomists who have achieved significant increases in yield or other race traits, with various traditional breeding approaches, such as masal selection (Smith et al., 2001).

For the last decade, the Mexican government has supported a few participatory improvement efforts (Aragón-Cuevas et al., 2003; Segura et al., 2011) to collect and evaluate races from different regions of the country. The goal is to be able to collect an important part of the maize race varieties used by local farmers, and promote breeding programs of different sorts. Such programs should effectively support local producers and races in order to further promote and improve their cultivation and production in all the Mexican territory.

Although seed-banks are an alternative tool to preserve agrodiversity in Mexico they have clear drawbacks (lack of representation, sample losses, bad storage conditions, classification problems and inadequate sampling procedures) with respect to *in situ* conservation (Ortega Paczka, 2007); and consequently should not be taken as the only and best way to preserve maize diversity, but rather be considered only as a complementary tool that feedbacks to *in situ* conservation programs that are based on the leaderships of indigenous and campesino communities.

In terms of the distribution of particular maize races, social variables included in our analyses were not as overall good predictors in comparison to environmental data, but they are also important components of the set of factors that seem to be related to the presence of particular races. In fact, twelve races were better predicted by the presence of particular ethnic group, rather than by any environmental variable (Fig. 1, Appendix 2). It is important to consider that the social data set is not as complete as the environmental one, and this could bias the importance of the former. However, epsilon values used in our analyses take into account the bias on the number of factors considered for each type of data and therefore, the suitable profile found for each race helps us identify the key factors mostly associated to the distributions of races, even when they were not good predictors. Future studies should further consider social factors that might be critical for the presence and distribution of other staple crops in developing countries, as well as for determining the location of the richest race areas.

For our case study, Palomero Toluqueño, the ethnic group was the variable with the highest predictive power. The ethnic groups that were related to this race were: Thlahuica, Pima, Otomí, Náhuatl and Mazahua. In this case, associations between most of these ethnic groups and Palomero Tolqueño had already been reported (Romero Contreras et al., 2006; Boege, 2010), except for the case of the Pima. A better understanding of the relationship among ethnic groups and maize race diversification and distribution could help agronomists and biologists identify where endangered races are being grown and how different ethnic groups are actively contributing to the generation and conservation of maize diversity. For instance, Purepecha is the only ethnic group associated with Zamorano Amarillo when the ethnic group was the most important factor in terms of spatial relations and predictive power. Additionally, ethnic groups such as Zapoteco are spatially related to 19 races, meaning that they are playing an important role on the *in situ* preservation of Mexican maize richness. These examples should make clear that preserving maize races implies following the leaderships and supporting the ethnic indigenous communities that are in charge of preserving and further breeding maize races. In other cases, "mestizo" peasants are also key for maize race diversification and preservation. Such *in situ* diversity might be critical to face climate change challenges, as well as to analyze which areas and races are critical to face such novel conditions (see also Ureta et al., 2012).

5. Concluding remarks

Our methodology facilitated the analysis of the complex system of growing maize in Mexico, as it integrates a range of different social and environmental variables, which would be very difficult to combine in a single model, given the most common distribution algorithms that have been proposed. The data-mining methodology used here allowed us to produce maps with predicted distributions of individual races, or of the areas harboring the potential highest richness of races (PRA) by using combined data sets. This contrasts with previous analyses that have only relied on environmental data.

The main results of our study are: (1) Altitude which is strongly related to climate was the factor with highest predictive power for most races. (2) Twelve races potential distribution areas were better predicted by an ethnic group than by any other factor. (3) Predictive power increases when all factors are used simultaneously. (4) Areas supporting high maize richness included Chiapas, Oaxaca, Guerrero, the western state of Jalisco, and central Mexico. Nevertheless, suitable areas for at least one maize race cover practically the entire Mexican territory. (5) PRA comprises around 6.9% of the Mexican territory and are associated to 40% of the ethnic groups of Mexico. (6) Interestingly, Zapoteco cultures have probably been involved in the diversification of many races that are still preserved in Mexico.

It is likely that as data sets for social factors improve and become more complete, this will be more important than suggested by our analyses. Nevertheless, we strongly recommend including current social factors in studies evaluating spatial distributions and associations of other crops. We envision that the presence of ethnic groups will be also critical to explain the distribution of crops associated to traditional agriculture, other than maize, such as native pumpkins (*Cucurbitapepo* L.) and beans (*Phaseoulus vulgaris* L.). The consideration of the leadership of the communities of indigenous communities is thus central to agrobiodiversity conservation at Centers of Crop Origin and/or diversification. On the other hand, additional biological factors such as pests are likely to improve the predicted distribution and impact of possible global climate change or pest crisis on race richness and distribution at Centers of Crop Origin and/or diversification.

Acknowledgements

We deeply appreciate the advice, ideas and overall input of Professor Hugo Perales Rivera during different stages of this study. We would like to thank el Posgrado en Ciencias Biológicas and CONACyT for the PhD grant given to the first three authors and the projects 167705, 152649, 105678, 180380, 180098, DGPA, UNAM (IN204011-3; IN203113-3; IN226510-3; IB201212-2) and UC-MEXUS CN.12-623 and 12–571. We would also like to thank the Dirección General de Análisis y Prioridades of CONABIO for providing the maize database, the Academic Writing course of the UNAM-CANADÁ and the Academic Writing course at the Instituto de Ecología, UNAM.ERAB was supported by the Miller Institute for Basic Research in Science at the University of California, Berkeley, California, USA

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.agee.2013.06.017.

References

Acevedo, F., Huerta, E., Burgeff, C., Koleff, P., Sarukhán, J., 2011. Is transgenic maize what Mexico really needs? Nat. Biotechnol. 29, 23–24.

Anderson, E.F., Cutler, H.C., 1942. Races of *Zea mays*: I Their recognition and classification. Ann. Mo. Botanical Garden 29, 69–88.

Aragón-Cuevas, F., Taba, S., Castro-García, F.H., Hernández-Casillas, J.M., Cabrera-Toledo, M., Osorio-Alcalá, L., Dillánes-Ramírez, N., 2003. *In situ* conservation and use of local maize races in oaxaca, mexico: a participatory and decentralized approach. In: Proceedings of a Workshop held at CIMMYT, April 7–10, 2003.

Bardsley, D., Thomas, I., 2005. Agrobiodiversity conservation for regional development in Nepal. GeoJournal 62, 27–39.

Bellon, M.R., Brush, S., 1993. Keepers of Maize in Chiapas, Mexico. Econ. Bot. 48, 196–209.

Bellon, M.R., Hodson, D., Bergvinson, D., Beck, D., Martínez-Romero, E., Montoya, Yinha, 2005. Targeting agricultural research to benefit poor farmers: relating poverty mapping to maize environments in Mexico. Food Policy 30, 476–492.

Bellon, M.R., Hellin, J., 2011. Planting Hybrids, Keeping Landraces: Agricultural Modernization and Tradition Among Small-Scale Maize Farmers in Chiapas, Mexico. World Development 39, 1434–1443.

Boege, E., 2010. El Patrimonio biocultural de los pueblos indígenas de México. In: Hacia la conservación in situ de la biodiversidad y agrodiversidad en los territorios indígenas. Instituto Nacional de Antropología e Historia, México, D. F.

Brush, S.B., Perales, H.R., 2007. A maize landscape: Ethnicity and agro-biodiversity in Chiapas Mexico. Agriculture. Ecosyst. Environ. 121, 211–221.

CIMMYT, 2012. Centro Internacional de Mejoramiento de Maíz y Trigo. http://www.cimmyt.org/

CONABIO, 1995. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. In: Mapa edafológico. Escalas 1:250 000 y 1:1000 000. México. http://www.conabio.gob.mx/informacion/metadata/gis/eda251mgw.xml?_httpcache=yes&_xsl=/db/metadata/xsl/fgdc_html.xsl&_indent=no

CONABIO, 2011. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. In: Base de datos de maíz [Maize data base]. http://www.biodiversidad.gob.mx/genes/genes.html#NULL, http://www.conabio.gob.mx/

Conde, C., Martínez, B., Sánchez, O., Estrada, F., Fernández, A., Zavala, J., Gay, C., 2008. Escenarios de Cambio Climático (2030 y 2050) para México y Centro América. Temperatura y Precipitación. (Documento en línea). http://www.atmosfera.unam.mx/gcclimatico/escenarios/Escenarios_de_cambio_climatico_2008.htm

Esquinas-Alcazar, J., 2005. Protecting crop genetic diversity for food security: political, ethical and technical challenges. Nat. Rev. Genet. 6, 946–953.

FAO, 2012. Food and Agriculture Organization. http://www.fao.org/

González-Salazar, C., Stephens Ch, R., Marquet, P.A., 2013. Comparing the relative contributions of biotic and abiotic factors as mediators of species distributions. Ecol. Model. 248, 57–70.

INALI, 2011. Instituto Nacional de Lenguas Indígenas. http://www.inali.gob.mx/

INEGI, 2005. Instituto Nacional de Estadística y Geografía. Conjunto de datos vectoriales de uso de suelo y vegetación Serie III (continuo nacional), escala 1:250 000.

INEGI, 2012. Instituto Nacional de Estadística y Geografía. Regiones Socioeconómicas de México. http://sc.inegi.org.mx/niveles/index.jsp

Jackson, C.R., Robertson, M.P., 2011. Predicting the potential distribution of an endangered cryptic subterranean mammal from few occurrence records. J.Nature Conserv. 19, 87–94.

Kato, T.A., Mapes, C., Mera, L.M., Serratos, J.A., Bye, R.A., 2009. Origen y diversificación del maíz: una revisión analística. Universidad Nacional Autónoma de México, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México, DF, pp. 116.

Kok, K., Veldkamp, A., 2001. Evaluating impact of spatial scales on land use pattern analysis in Central America. Agri. Ecosyst. Environ. 85, 205–221.

Loarie, S.R., 2008. Climate change and the future of California's endemic flora. PLos ONE 3 (6), e2502, doi:2510.1371/journal.pone.0002502.

Nuss, E.T., Tanumihardjo, S.A., 2010. Maize: a paramount staple crop in the context of global nutrition. Compr. Rev. Food Sci. Food Safety 9, 417–436.

Ortega Paczka, R., 2007. La diversidad del maíz en México. Sin maíz no hay país. Consejo Nacional Para la Cultura y las Artes, México, D.F., pp. 123–154.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2011. Ecological Niches and Geographic Distributions. Princeton University Press, U.K., London, pp. 125–126, 23-46.

Prates-Clark, C.D.C., Saatchi, S.S., Agosti, D., 2008. Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data. Ecol. Model. 211, 309–323.

Romero Contreras, T., González Días, L., Reyes Reyes, G., 2006. Geografía e historia del maíz Palomero toluqueño (*Zea mayseverta*). Ciencia Ergo Sum 13, 47–56.

Ruiz Corral, J.A., Durán Puga, N., Sánchez González, J.J., Ron Parra, J., González Euiarte, R., Holland, J.N., Medina García, G., 2008. Climatic adaptation and ecological descriptors of 42 Mexican races. Crop Sci. 48, 1502–1512.

Sánchez González, J.J., Goodman, M., 1992. Relationships among Mexican Races of Maize. Econ. Bot. 46, 72–85.

Segura, J.C., Dzib-Aguilar, L.A., Ortega-Paczka, R., Latournerie-Moreno, L., 2011. Diallelic crosses among maize landraces from Yucatan and improved populations. Trop. Subtrop. Agroecosyst. 14, 119–127.

Smith, M.E., Castillo, F., Gómez, F., 2001. Participatory plan breeding with maize in Mexico and Honduras. Euphytica 122, 551–565.

Stephens, C.R., Giménez-Heau, J., Gonzálezz, C., Ibarra-Cardeña, C., Sánchez-Cordero, V., Gonzálezz-Salazar, C., 2009. Using biotic interaction networks for prediction in biodiversity and emerging diseases. Plos ONE 4, e5725.

Téllez, O., Hutchinson, M.A., Nix, H.A., 2011. Desarrollo de coberturas digitales climáticas para México. Universidad Autónoma del estado de Hidalgo, Pachuca, México.

Thrupp, L.A., 2000. Linking agricultural biodiversity and food security: the valuable role of agrobiodiversity for sustainable agriculture. Int. Aff. 76, 283–297.

Ureta, C., Martínez-Meyer, E., Perales, H.R., Álvarez-Buylla, E., 2012. Projecting the effects of climate change on the distribution of maize races and their wild relatives in Mexico. Global Change Biology 18, 1073–1082.

USGS, 2010. United States Geological Survey, HYDRO1K http://www.usgs.gov/