ASIAN JOURNAL OF AGRICULTURE
Volume 4, Number 2, December 2020
Pages: 71-82

E-ISSN: 2580-4537
DOI: 10.13057/asianjagric/g040203

Species distribution modeling and phenotypic diversity reveals collection gap in the *Musa balbisiana* germplasm conservation in Philippines

ROEL C. RABARA[1,♥], RACHEL C. SOTTO[2,♥♥], ERIC ARIEL L. SALAS[3]
[1]University of Virginia. Charlottesville, VA 22903, USA. ♥email: roel.rabara@jacks.sdstate.edu
[2]Institute of Plant Breeding, University of the Philippines Los Baños. College, Laguna 4031, Philippines. Tel.: +63-49-5365287, ♥email: roel.rabara@jacks.sdstate.edu, ♥♥ rcsotto@up.edu.ph
[3]Agricultural Research Development Program, Central State University. Wilberforce, Ohio 45384, USA

Manuscript received: 5 July 2020. Revision accepted: 12 August 2020.

Abstract. *Rabara RC, Sotto RC, Salas EAL. 2020. Species distribution modeling and phenotypic diversity reveal collection gap in the* Musa balbisiana *germplasm conservation in Philippines. Asian J Agric 4: 71-82. Musa balbisiana* is one of the progenitors of cultivated bananas. Its conservation is crucial because of its important phenotypic traits for breeding and genetic improvement of future *Musa* crops. In the Philippines, the largest ex-situ germplasm collection of *M. balbisiana* was characterized to assess their phenotypic diversity. Diversity analyses of the 97 plant samples revealed high diversity in the collection. Of the 80 quantitative traits that showed diversity, 51% showed high diversity indices (H′=0.76 to 0.99) while 20% and 29% of the traits have medium and low diversity, respectively. Nineteen percent of the scored morphological traits were invariants, which may reflect the need to increase diversity in the germplasm holdings. Species distribution analysis revealed that annual mean temperature and annual precipitation influenced the species distribution models suggesting the importance of these two environmental factors to species establishment. Overall, phenotypic diversity analyses and species distribution models revealed gaps in the collection. Further collection trips are recommended to increase diversity on traits with low diversity indices and to increase representative accessions from southern Philippines.

Keywords: Crop wild relatives, correlation, ex-situ conservation, SDM ensemble, morphological traits

INTRODUCTION

Banana ranks fourth in world’s staples succeeding rice, wheat, and maize (Churchill 2011). Its cultivation is a major employment source and main staple food of over 400 million people (Hölscheret et al. 2014). In 2013, the Philippines was the second largest global banana producer, generating nearly 94% of Asia’s banana export. But its production drastically declined in 2014-2015 from 3.68 to 1.85 million tons because of long drought (FAO 2014, 2017).

Ranking seventh in species diversity and endemism (Altoveros and Borromeo 2007), the Philippines is central to Musa diversity with indigenous species: *M. balbisiana*, *M. acuminata* and *M. textilis* (Lalusin and Villavicencio 2015). *M. balbisiana* or 'Butuhan' to Filipino farmers is widely utilized as food (male bud), animal feeds (pseudo-stem), and food preparation media (leaves).

Musa balbisiana is known for resistance traits against bacterial wilt (Kumakechet al. 2013) diseases. Its resistance to bunchy top and mosaic virus diseases proved useful in breeding abaca (*M. textilis*) (Bernardo and Umali 1956; Lalusin and Villavicencio 2015), Philippines’ primary export fiber crop. Most significantly, hybridizations between *M. balbisiana* Colla (B genome) and *M. acuminate* Colla (A genome) produced commercial bananas (Heslop-Harrison and Schwarzacher 2007; Davey et al. 2013). Recently, *M. balbisiana* genome’s association with improved vigor and tolerance to biotic and abiotic stresses improved its hold in Musa breeding programs (Davey et al. 2013).

Wild *Musa*, like other crop wild relatives (CWRs) possess useful crop improvement traits and are gene sources of disease resistance, tolerance to extreme temperatures, salinity and drought (Lane 2006; Hajjar and Hodgkin 2007; Heslop-Harrison and Schwarzacher 2007). CWRs benefits are valued at US$68 and US$120 billion for commercial and priority crops, respectively, about $196 billion when soybean and maize are included (PwC 2013; Fielder et al. 2015; Tyack and Dempewolf 2015).

Uses of CWRs in breeding increased owing to biotechnology advances like omics approaches and to geographic and ecological variation metrics to analyze gaps for improving CWRs conservation efficiency (Zhang et al. 2017), including geographic information system (GIS), species distribution modeling and other predictive tools (Hunter and Heywood 2012). GIS helped strategize conservation plans for Peruvian wild potato species (Hijmans and Spooner 2001), *Phaseolus* beans gene pool collections (Ramírez-Villegas et al. 2010) and collecting missions for Paraguay’s rare wild pepper (Jarvis et al. 2005). Ecogeographic studies defined core collections and predicted gaps for optimal GenBank management (Guarino 1995; van Zonneveld et al. 2011; Parra-Quijano et al. 2012; Rabara et al. 2014).

Overall, this study focused on *M. balbisiana* characterization for future utility in banana improvement. Explicitly, this study analyzed the phenotypic diversity and geographic distribution of *M. balbisiana* conserved by a research GenBank in Laguna, Philippines that could guide breeders for specific desirable traits boosting food and fiber production and tolerances to environmental stresses of future *Musa* varieties. Results could influence future collections and conservation of *M. balbisiana* diversity and make *Musa* germplasm management efficient and cost-effective, where conserved collections remain true representations of a species' phenotypic and geographical diversities.

MATERIALS AND METHODS

Germplasm collection and phenotyping

A total of 97 *M. balbisiana* accessions were considered for this study, that were collected through field study explorations in different parts of the Philippines (Sotto and Rabara 2000). The *M. balbisiana* accessions were established at the field GenBank of the National Plant Genetic Resources Laboratory (NPGRL), Institute of Plant Breeding, University of the Philippines Los Baños as part of a national ex-situ conservation program. Phenotyping of the conserved germplasm at different stages of development were done using 99 selected morphological traits comprising both qualitative and quantitative traits (IPGRI-INIBAP/CIRAD 1996). All phenotypic traits were measured from three biological replicates, representing three mats for each accession. Multivariate data analyses were done using NTSYS software (Rohlf 2002) following the protocol described by Rabara et al. (2014). Correlation analysis was done to identify phenotypic traits that are associated with each other following methodology of Taylor (1990).

Georeferencing and species distribution modeling

Germplasm passport data were used to generate georeferenced data points for each collection using the Philippine gazetteer database downloaded from the DIVA-GIS (Hijmans et al. 2001). Locations not found in the gazetteer were data mined from online databases (http://www.geonames.org/). Figure 1summarized the processes involved in generating potential distribution of *M. balbisiana* in the Philippines following the Overview, Data, Model, Assessment and Prediction (ODMAP) protocol outlined by Zurell et al. (2020).

We utilized 19 raster-based bioclimatic variables from the WorldClim datasets (Fick and Hijmans, 2017). The set of variables was used to describe the present environmental conditions and explore the relationship between bioclimatic conditions and species distribution patterns. WorldClim provides climate models downscaled to 30 seconds, roughly 900 m at the equator (Table 1). To model the climatic habitat of *M. balbisiana*, we selected the following SDMs: Random Forest (RF) (Breiman, 2001), Boosted Regression Tree (BRT) (Elith et al. 2008), Maxent (Phillips et al. 2006), and Multivariate Adaptive Regression Splines (MARS) (Leathwick et al. 2006). These SDMs were selected based on their performance with presence-only data. We utilized a tool provided by Talbert (2012) to create a modeling workflow and develop bioclimatic-envelope models for present day conditions. The tool used source codes shared within Github (https://github.com/talbertc-usgs/sahm). Since species lacked absence data, the tool randomly generated 10,000 background points (i.e., pseudo-absences) (Phillips and Dudík, 2008). When multiple species occurrences were present within a given pixel of the climatic data, the tool consolidated them to a single occurrence per pixel. We summed up binary maps generated from probability surfaces from each statistical modeling algorithm (Liu et al. 2005). We used specificity = sensitivity as the threshold in discretizing the probability maps. This has previously been identified as the optimal threshold (Manel et al. 2001). The final map consisted of pixel values that showed the number of models in agreement that a particular pixel is suitable for the species. A pixel with a value of zero meant that none of the models identified bioclimatic suitability for the species at that location, while a value of four meant there was agreement across all four SDMs.

We removed one of each pair of highly correlated ($r > 0.7$) (Salas et al. 2017) environmental variables to avoid collinearity among predictors (Gama et al., 2015). We made the choices between variables based on the results of a species-specific literature search. We selected variables that were identified in one or more studies regarding the *M. balbisiana* as influencing its distribution.

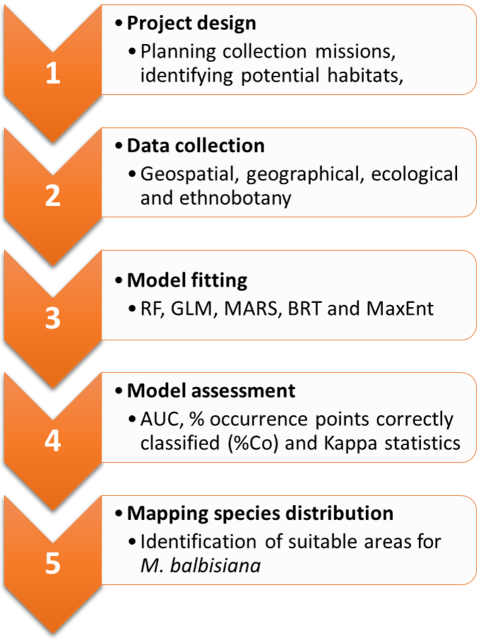


Figure 1. Generalized workflow in generation of *Musa balbisiana* distribution map following the ODMAP guidelines (Zurell et al. 2020). Model abbreviations are as follows: GLM: Generalized Linear Model, MARS: Multivariate Adaptive Regression Splines, BRT: Boosted Regression Tree, and RF: Random Forest.

Table 1. List of 19 bioclimatic variables used in bioclimatic-envelope model development. Names and descriptions are based on WorldClim. One of each pair of highly correlated (r > 0.7) variables was removed from the models to avoid collinearity among variables (Dormann, 2013). An asterisk (*) denotes final variables used for modeling the climatic habitat.

Variable	Description
Bioclim 1	Annual Mean Temperature*
Bioclim 2	Mean Diurnal Range
Bioclim 3	Isothermality*
Bioclim 4	Temperature Seasonality
Bioclim 5	Maximum Temperature of the Warmest Month
Bioclim 6	Minimum Temperature of the Coldest Month
Bioclim 7	Temperature Annual Range*
Bioclim 8	Mean Temperature of Wettest Quarter
Bioclim 9	Mean Temperature of Driest Quarter
Bioclim 10	Mean Temperature of Warmest Quarter
Bioclim 11	Mean Temperature of Coldest Quarter
Bioclim 12	Annual Precipitation
Bioclim 13	Precipitation of Wettest Month
Bioclim 14	Precipitation of Driest Month
Bioclim 15	Precipitation Seasonality*
Bioclim 16	Precipitation of Wettest Quarter*
Bioclim 17	Precipitation of Driest Quarter
Bioclim 18	Precipitation of Warmest Quarter*
Bioclim 19	Precipitation of Coldest Quarter

We assessed confidence in individual model results in terms of concordance among the different distribution models. We had higher confidence that environmental conditions were suitable for a species when three or more (at least 75% of) algorithms agreed. Information was compiled on various measures of model performance, including the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) for the test data, correct classification rate (%Co) (Warren and Seifert, 2011), and the kappa statistic (Allouche et al. 2006) for each algorithm by species combination. Swets (1988) classified values of AUC as follows: those > 0.9 indicated high accuracy, from 0.7 to 0.9 indicated good accuracy, and those < 0.7 indicated low accuracy.

RESULTS AND DISCUSSION

Phenotyping and diversity analysis of germplasm

Climate is an important factor that defines the distribution of plant species (Silva-Flores et al. 2014) and is a commonly used variable in species distribution modeling of plant species (Austin and Van Niel 2011). Philippine climate is tropical and maritime and is formally categorized into four types (Figure 2) based on the prevalence of the southwest and northwest monsoons and monthly distribution of rainfall (Altoveros and Borromeo 2007).

This study analyzed the phenotypic diversity of 97 accessions of *M. balbisiana* that are ex-situ conserved at the NPGRL field GenBank of the UPLB-Institute of Plant Breeding. Originally, the accessions were collected through field explorations from 21 provinces in the Philippines, where majority of the areas have 506 m elevation (Figure 2).

General appearance of the germplasm collections showed similar green pseudo-stem except for accession 5942 (Pisang KlutukWulung) which has distinct black pseudo-stem. Detailed phenotyping of this germplasm showed diversity in various morphological characters. Calculated Shannon-Weaver diversity indices (H′) showed that 51% of the characters measured have high diversity (H′≥ 0.76 to 0.99) while moderate (H′ ≥ 0.46 to 0.75) and low (H′ ≤ 0.45) diversity comprised 20% and 29%, respectively (Figure3). The fading color of bract base (BRBFC) had the highest diversity index (0.999) which reflected that all states under this descriptor were observed in the evaluated germplasm collections. Figure 3 summarizes the diversity indices calculated for all the 80 morphological characters scored. More than half (51%) of the scored characters have high diversity indices, while 30% of the scored characters showed low diversity. Nearly two-thirds (61%) of the characters showing low diversity were inflorescence characters. These include characters like bract imbrication, style shape and rachis position that showed the lowest diversity (H′=0.087). Only one state was observed for each of these characters which resulted in their having low diversity. This may indicate that there is no diversity in these characters or further collection and characterization are needed to capture diversity in these characters. Also, invariant traits in 19 characters were observed while phenotyping the evaluated germplasm (Table 2). Close to half (42%) of these invariant characters are description of plant habits.

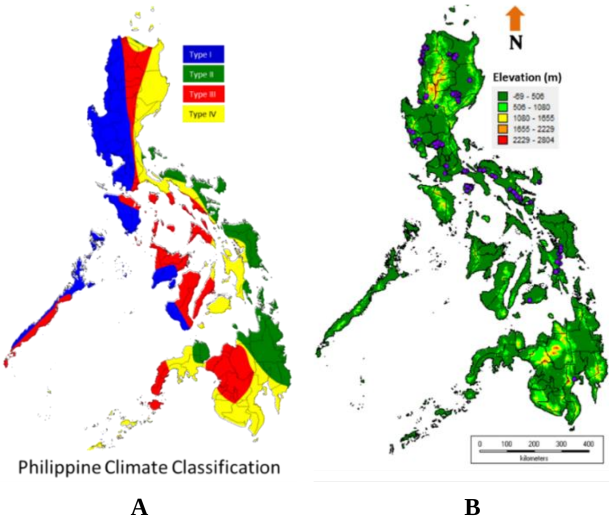


Figure 2. Map of the Philippines: (A) Climate classification based on Coronas system(Altoveros and Borromeo 2007): Type I-two pronounced seasons: dry from November to April, wet the rest of the year; Type II- No dry season with a very pronounced rainfall from November to April and wet the rest of the year; Type III-Seasons are not very pronounced, relatively dry from November to April and wet the rest of the year; Type IV-Rainfall is more or less evenly distributed throughout the year. (B) Geographical locations of *Musa balbisiana* germplasm (purpledots) collected from different areas in the Philippines.

Table 2. Invariant characters observed in the *Musa balbisiana* collections.

Trait	Descriptor state	
	Observed	**Observable states**
Plant habit		
Sap color	Watery	1. Watery 2. Milky 3. Red-purple 4. Other
Petiole margins	Winged and clasping the pseudo-stem	1. Winged and undulating 2. Winged and not clasping the pseudo-stem 3. Winged and clasping the pseudo-stem 4. Not winged and clasping the pseudo-stem 5. Not winged and not clasping the pseudo-stem
Wing type	Dry	1. Dry 2. Not dry
Petiole margin color	Green	1. Green 2. Pink-purple to red 3. Purple to blue 4. Other
Edge of petiole margin	Without color line along the petiole	1. Colorless 2. With a color line along
Shape of leaf blade base	Both sides rounded	1. Both sides rounded 2. One side rounded, one-pointed 3. Both sides pointed
Color of cigar leaf dorsal surface	Green	1. Green 2. Red-purple 3. Other
Peduncle hairiness	Hairless	1. Hairless 2. Slightly hairy 3. Very hairy, short hairs 4. Very hairy, long hairs
Male bud		
Rachis type	Present	1. Truncated 2. Present
Male bud type	Normal	1. Normal (present) 2. Degenerating before maturity (like false-horn 'Plantain') 3. Like true-horn ‘Plantain’ (absent)
Color on the bract apex	Tinted with yellow	1. With yellow tint 2. No yellow tint
Bract behavior before falling	Not revolute	1. Revolute (rolling) 2. Not revolute (not rolling)
Filament color	Yellow	1. White 2. Cream 3. Yellow
Anther color	Yellow	1. White 2. Cream 3. Yellow 4. Grey 5. Brown/rusty brown 6. Pink/pink-purple 7. Black (anthers aborted) 8. Other
Pigmentation on style	Without pigmentation	1. Without pigmentation 2. Purple
Bunch/fruits		
Pedicel surface	Hairless	1. Hairless 2. Hairy
Fusion of pedicels	No sign of fusion	1. Very partially or no visible sign of fusion 2. Partially fused 3. Totally fused
Pulp in fruit	With pulp	1. Without pulp 2. With pulp
Pulp color at maturity	White	1. White 2. Cream 3. Ivory 4. Yellow 5. Orange 6. Beige-pink 7. Other

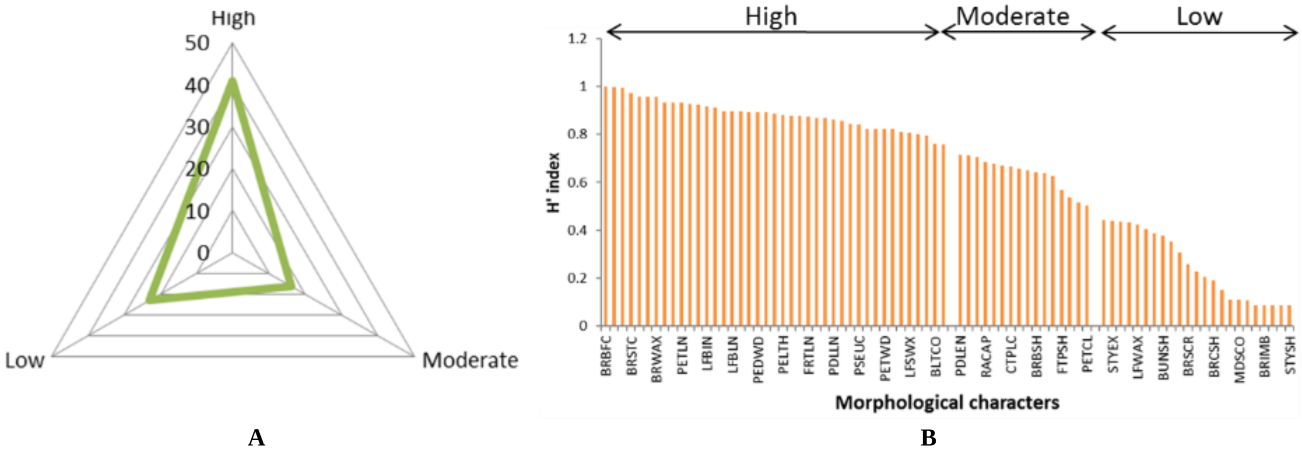


Figure 3. Calculated diversity indices (H′) of each measured and scored morphological characters of *Musa* germplasm collection at NPRGL. **(A)** Frequency distribution of diversity indices classified into high, moderate, and low (Rabara et al. 2014). **(B)** Diversity indices of each character measured.

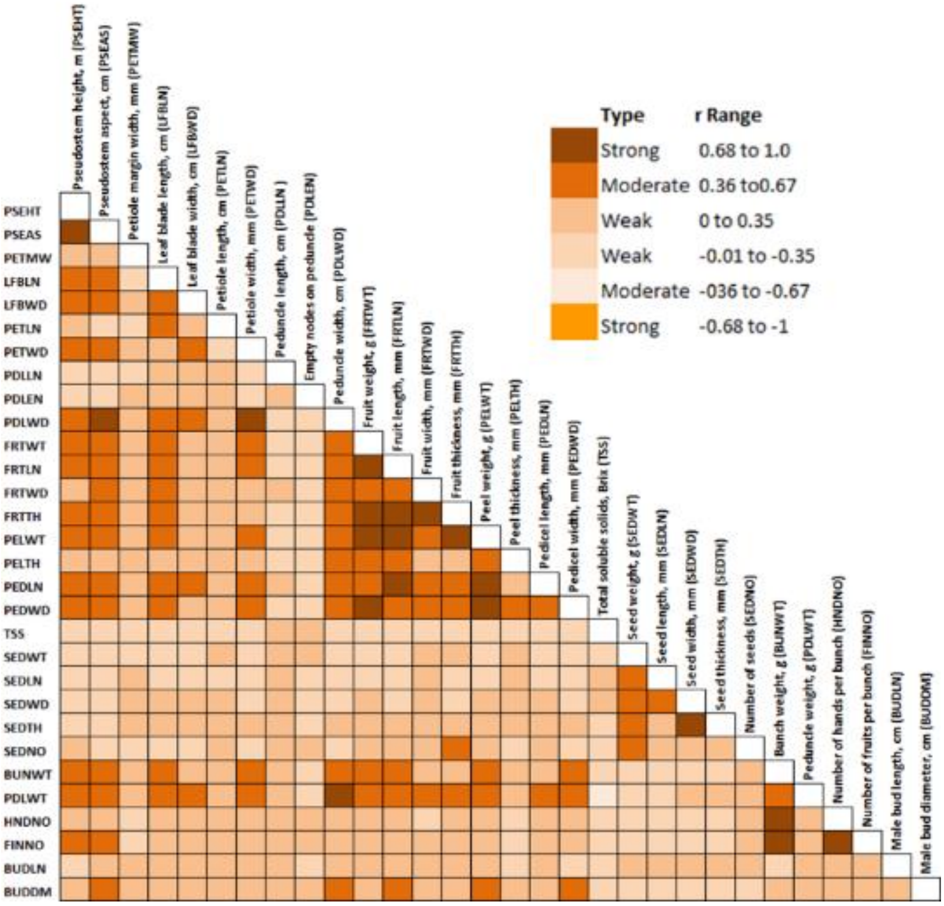


Figure 4. Heat map representing calculated Pearson's product-moment correlation coefficients among morphological traits measured in 97 *Musa* germplasm characterized. Taylor (1990) classified correlation coefficients as weak (r ≤ 0.35), moderate (r ≥ 0.36) and strong (r > 0.68).

To assess the relationship among quantitative characters measured, correlation analysis was conducted on 30 morphological characters (Figure 4). Correlation analysis can be a valuable tool in germplasm conservation and management, as highly correlated characters can be prioritized during collections. Correlation analysis showed that approximately 4.4% of the calculated correlation values have strong positive correlations. These were mostly fruit characters like fruit number, weight, length, and peel weight. Correlation between the number of hands and number of fingers present in a bunch showed the highest correlation (r=0.93). A high correlation between plant diameter and plant height (r=0.87) was also observed. Positive moderate and weak correlations constituted 21.8% and 44.1%, respectively. Negative correlation among the characters was also observed which constituted 29.4% of the total values observed.

Cluster analysis was done to assess the relationship among the evaluated germplasm based on their morphological characteristics. Generated dendrogram showed three distinct clusters with Cluster I having the greatest number of members (Figure 5). This group is composed of tall accessions with an average height of 4.2 m and pseudo-stem circumference of 67 cm. The group also produced the heaviest fruit (finger) with an average of 129 g which could be attributed to thick peel (2.32 mm) and long pedicel (26 mm). Moreover, Cluster I had the greatest number of fingers per bunch (99 fingers) but had similar number of hands per bunch (seven hands) with the accessions in Cluster II.

In comparison to Cluster I, members in Cluster II had relatively shorter height averaging 4.1 m with pseudo-stem circumference of 65 cm. Although Cluster I had lighter fruit weight (115 g) on the average compared to Cluster II, it had the greatest number of seeds (258 seeds per finger) that could have led to having the heaviest fruit bunch (15.4 kg). On the other hand, members of Cluster III had the shortest height averaging 3.6 m and smallest pseudo-stem diameter (57 cm). The group only averaging six hands per bunch, had the least number of fingers per bunch (67 fruits) and the lightest fruit weight (103 g).

Fruit weight was one of the fruit traits with high diversity index (H′=0.893). To classify the accessions based on their fruit and quantify their diversity, color-coded geogrids of the germplasm collections based on their fruit weight class were generated and their variations were measured (Figure 6). Accessions having heavy fruit weight (167-209 g) were mostly collected from the northern parts of the country. Lightest fruit weight (≤ 42 g) was measured from accessions collected in the central part of the country. However, both the northern and central regions of the country showed high diversity in fruit weight as shown by the calculated coefficient of variation (CV) values ranging between 61 and 76.

Aside from phenotypic diversity analyses, clustering in terms of geolocations was examined. Figure 4 shows the locations of members of each cluster and the type of climate of the original locations where the banana germplasm accessions were collected. Cluster I have the largest number of accessions were mostly collected from the northern parts of the country which have a Type I climate (pronounced wet and dry season); Type III (wet and dry seasons not clearly defined) or Type IV climate (rainfall evenly distributed year-round). Members of Cluster II were mostly from regions with Type I and IV climatic conditions. On the other hand, Cluster III have similar profile with Cluster I where accessions collected grew in Types I, III and IV climatic conditions. No plant samples were collected in areas having Type II climate.

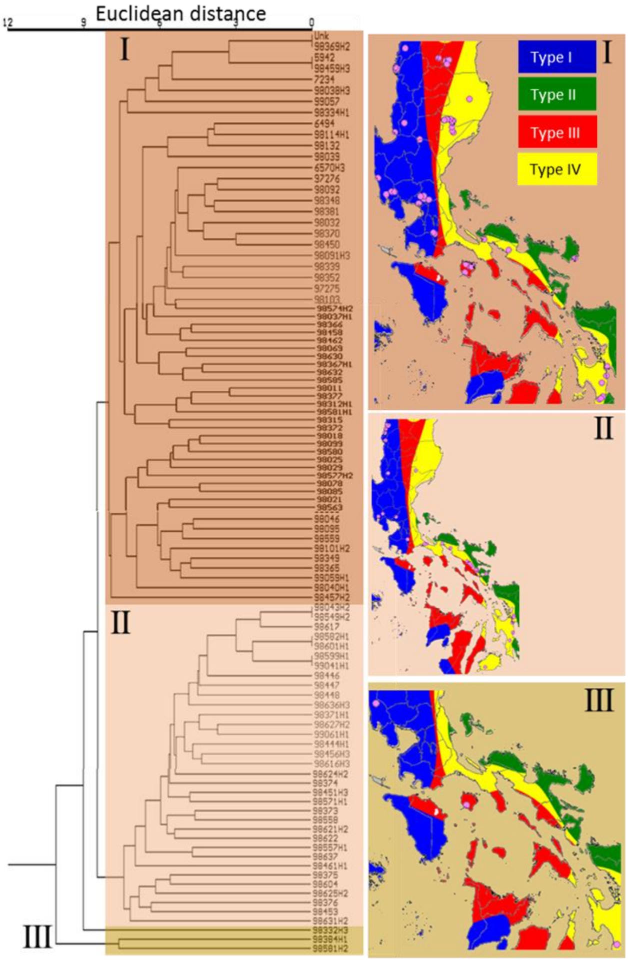
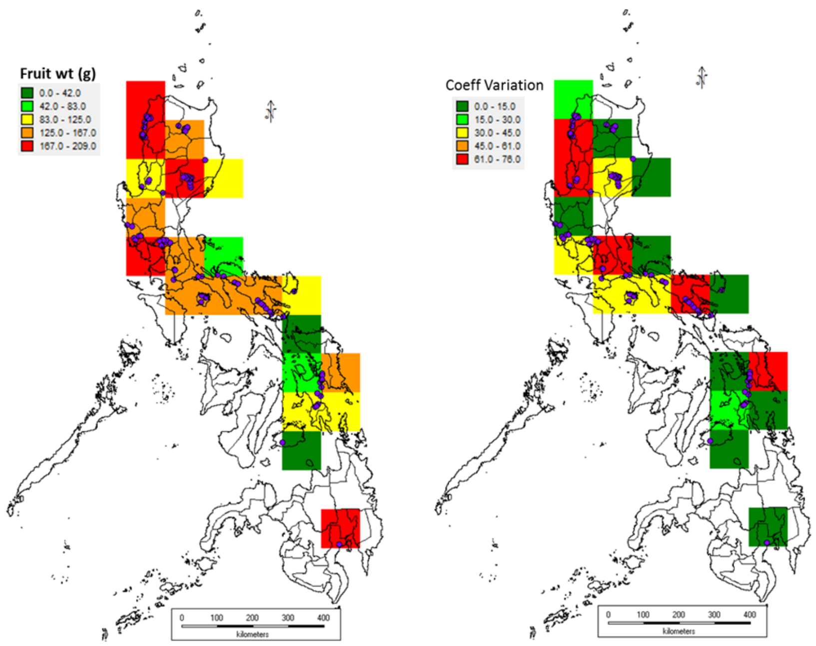


Figure 5. Cluster analysis of the phenotypic data of *Musa balbisiana* collection showing the geolocation or collection sites of accessions for each cluster. Different colors denote regional climate classifications: Type I (Blue), Type II (Green), Type III (Red) and Type IV (Yellow).

Species distribution modeling of germplasm

Using five common species distribution models, we generated climatic suitability maps derived from current climatic conditions for *M. balbisiana* in the country based on presence data (Figure 7). The models showed excellent levels of predictive performance as indicated by the area under the receiver operating characteristic (ROC) curve value (AUC) of test data and the percentages of occurrence points correctly classified (%Co) for the five different models (Table 3). Among the five models, BRT, RF and Maxent showed high AUC and %Co with BRT showing the highest computed AUC and %Co with 0.96 and 94%, respectively. The K values varied across models but follows the trend observed in AUC and %Co. Like the results for AUC, the BRT, RF, and Maxent models performed well in terms of K.BRT scored the highest (K = 0.89) while GLM scored the lowest among five models with K = 0.72. Overall, GLM had the lowest accuracy based on both the %Co and K relative to the other four SDMs.



A **B**

Figure 6. Using DIVA-GIS ver.7.5.0 (Hijman et al. 2001), generated maps show diversity in fruit weight among the evaluated 97 *Musa balbisiana* collection: (A) fruit weight measured and (B) coefficient of variation in fruit weight. Colored grids on the maps denote different values in fruit weight (A) and coefficient of variation (B).

Among the climatic variables, annual mean temperature (Bioclim 1)and precipitation of warmest quarter(Bioclim 18) were the top two most important predictors for all five SDMs (Table 4). BRT, RF, and Maxent models ranked precipitation seasonality (Bioclim 15) third. Isothermality (Bioclim 3) was the least important among six variables for four of the five SDMs.

The models predicted high probability of presence in areas close to where the *M. balbisiana* germplasm used in this study have been collected as shown in Figure 7. Interestingly, the GLM model (Figure 6.b) showed many areas of the country to have high probability of presence for *M. balbisiana*. These include the islands of Panay, Negros and Cebu in the Central Visayas region of the country which has no representative germplasm collected in these regions. Panay and Negros islands have Types I and III climate while Cebu has Types III and IV. The island of Bohol has high probability of presence of the species in all five models. The island of Mindanao which has one representative germplasm showed to have high probability of *M. balbisiana* presence in areas with Types II and IV climates shown in four models (Figures 7.a-d). Both these climate types have pronounced precipitation throughout the year. In RF (Figure 7.e), the whole island of Mindanao was predicted to be suitable for *M. balbisiana* at varying degree of probability. Combining models for probability of presence improved the potential suitable areas for *M. balbisiana* in the country (Figure 8). The improved suitable maps showed that at least 60% of the algorithms agree that the bioclimatic conditions of the area are suitable for the species. This analysis identified suitable areas where no presence data was observed such as the island of Palawan (western part of the country) and the islands in the southern tip of the country such as Basilan, Jolo and Tawitawi.

Discussion

Germplasm characterization is a core activity in germplasm conservation to enhance efficient management and utilization of available germplasm in breeding programs (Rabara et al. 2014; Rabara et al. 2015). However, the management of GenBank is usually based on intuition rather than reason because of inadequate characterization data of germplasm collections (McCouchet al. 2012). This is one of the main reasons this study on extensive phenotyping of *M. balbisiana* germplasm collection conserved at NPGRL GenBank was conducted. The existing collections' phenotypic and genetic diversity need assessment to develop conservation and management strategies for current and future use in *Musa* varietal improvement. De Vicente et al. (2005) emphasized the importance of germplasm characterization in decision-making for germplasm conservation and utilization.

Based on calculated diversity indices, results clearly showed that the current collection is highly diverse on some of traits that were measured and scored. Also, results of genetic diversity assessment indicated high diversity in the collection. Some 19 traits were found to be invariants and should be considered as target traits in future collection trips to enhance diversity of these traits in the ex-situ conserved *Musa* germplasm.

Table 3. The Areas under the Curve (AUC) and the kappa statistics (K)associated with the test data and the percentages of occurrence points were correctly classified (%Co) for the five different models. Model abbreviations are as follows: GLM: Generalized Linear Model, MARS: Multivariate Adaptive Regression Splines, BRT: Boosted Regression Tree, and RF: Random Forest.

Species	**GLM**			**MARS**			**BRT**			**RF**			**Maxent**		
Musa	AUC	%Co	K	AUC	%Co	K	AUC	%Co	K	AUC	%Co	K	AUC	%Co	K
	0.79	76.2	0.72	0.85	82.7	0.77	0.96	93.8	0.89	0.94	92.3	0.87	0.95	91.2	0.87

Table 4. Ranking of the important climatic predictors for each statistical algorithm for *Musa balbisiana*. The two most important variables are annual mean temperature and precipitation of warmest quarter. Model abbreviations are as follows: BRT: Boosted Regression Tree, GLM: Generalized Linear Model, MARS: Multivariate Adaptive Regression Splines, Maxent, and RF: Random Forest.

Rank	**GLM**	**MARS**	**BRT**	**RF**	**Maxent**
1	Annual Mean Temperature	Annual Mean Temperature	Annual Mean Temperature	Annual Mean Temperature	Annual Mean Temperature
2	Precipitation of Warmest Quarter	Precipitation of Warmest Quarter	Precipitation of Warmest Quarter	Precipitation of Warmest Quarter	Precipitation of Warmest Quarter
3	Precipitation of Wettest Quarter	Precipitation of Wettest Quarter	Precipitation Seasonality	Precipitation Seasonality	Precipitation Seasonality
4	Precipitation Seasonality	Temperature Annual Range	Temperature Annual Range	Temperature Annual Range	Precipitation of Wettest Quarter
5	Temperature Annual Range	Precipitation Seasonality	Precipitation of Wettest Quarter	Isothermality	Temperature Annual Range
6	Isothermality	Isothermality	Isothermality	Precipitation of Wettest Quarter	Isothermality

The species distribution analyses showed that annual mean temperature and precipitation of warmest quarter greatly influenced results of diversity indices, ranking first and second, respectively, in all five statistical algorithms. Such attests that environmental factors play an important role in species diversification. In dryland habitats, seasonal pattern of rainfall is one of the driving forces in bio-diversification (Bonkoungou 2001). Similarly, rainfall pattern significantly influences floristic differentiation among Atlantic rainforests in Southeastern Brazil (Oliveira-Filho and Fontes 2000). A survey conducted by Zhao et al. (2018) of 2700 plants species in China has shown that mean annual precipitation and mean annual temperature predictive contribution to woody fleshy-fruited species distributions and herbaceous fleshy-fruited species distributions, respectively. Species distribution of *M. balbisiana*, an herbaceous fleshy-fruited species, was highly influenced by annual precipitation while the minimum temperature of the coldest month contributed 24% to the model. Ramirez et al. (2011) reported that decrease in rainfall can affect rates of photosynthesis and leaf emergence in bananas since they have rapid physiological response water deficit in the soil.

The strength of the models depended on both the selected predictors as well as the methodology in building them. The spatial resolution of the environmental variables used in the modeling in our study fit the spatial resolution of our species records which address one of the common mistakes pointed out by Sillero and Barbosa (2020). The bioclimatic variables that were highly correlated (r > 0.7) were removed from the models to avoid collinearity among variables, to avoid over-fitting of these variables to the model (Pradhan, 2016). The models we ran were based on the climatic data alone and excluded non-climatic variables. We believe that the distribution of suitable conditions for *M. balbisiana* would be driven by the climatic variables that we focused on. Apart from the lack of datasets projected according to the climatic dataset, scale is also an issue as non-climatic variables such as soil properties, for instance, are available at finer resolutions than the climatic dataset. Since our results for important climatic variables were based on concordance of five models, they are likely to have high transferability.

Knowledge of species distribution is an important component in biodiversity conservation programs. SDMs provide tools for GenBank curators and conservationists to identify areas for *in-situ* conservation of CWRs (Vincent et al. 2019) as well as to prioritize collecting for ex-situ conservation (Ramirez-Villegas et al. 2020). Vincent et al. (2019) proposed 150 global sites for *in-situ* conservation of CWR that will conserve 66% of major crop CWRgenepools. Ramirez-Villegas et al. (2020) used SDM to prioritize collecting of Andean and Mesoamerican bean gene pools to fill up the gap in ex-situ collection of common bean landraces. Our goal for this study was to identify potential gaps in our ex-situ germplasm collection by combining phenotypic diversity and species distribution data. In our study, we observed that *M. Balbisiana* accessions with bigger fruits were mostly found in the northern area of the country, in provinces with pronounced drier seasons like the province of Quirino. However, this observation may not hold true if compared to *M. balbisiana* accession growing in the southern parts of the country since there is only one germplasm collected in the area among the 97 accessions evaluated. Having only a single collection in the Mindanao area is partly due to inaccessibility of the region owing to conflicts and safety issues during exploration trips. This by itself requires the need for collection activities in Mindanao to increase germplasm collection and to ensure representative accessions are conserved. Increasing the number of germplasms collected from this region might improve the distribution model in the area as previous studies reported that distribution of *M. balbisiana* is influenced by precipitation. For example, *M. balbisiana* is commonly found farther north of India where dry seasons are prominent (Subbaraya et al. 2006).

One setback encountered during field exploration and collection of *M. balbisiana* germplasm was the accessibility of the natural habitat of this wild species. Due to difficulty of accessing far-flung areas in most of the provinces, majority of the germplasm collected was close to road networks (Figure 9). The geographic locations of the 97-germplasm evaluated in this study showed collection bias, one of the most commonly cited problems in field exploration(Hermann 1988; von Bothmer and Seberg 1995; Hijmans et al. 2000).

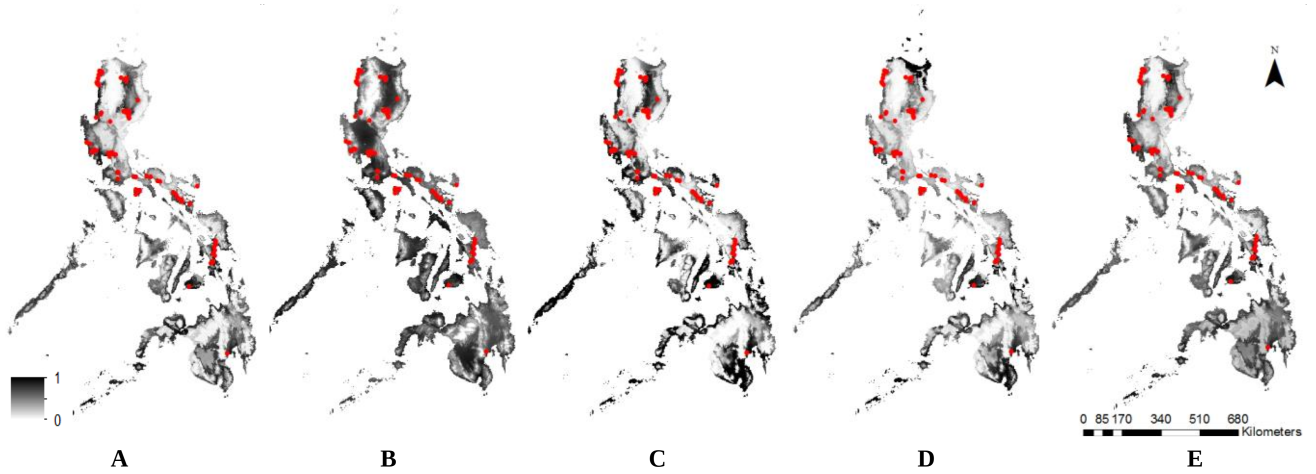


Figure 7. Climatic probability maps derived from current climatic suitable conditions for Musa (*Musa balbisiana*) using five common species distribution models: A. BRT: Boosted Regression Tree, B. GLM: Generalized Linear Model, C. MARS: Multivariate Adaptive Regression Splines, D. Maxent, and E. RF: Random Forest. Species occurrence points are shown in red dots. The map values range between 0 (white, low probability) to 1 (black, high probability).

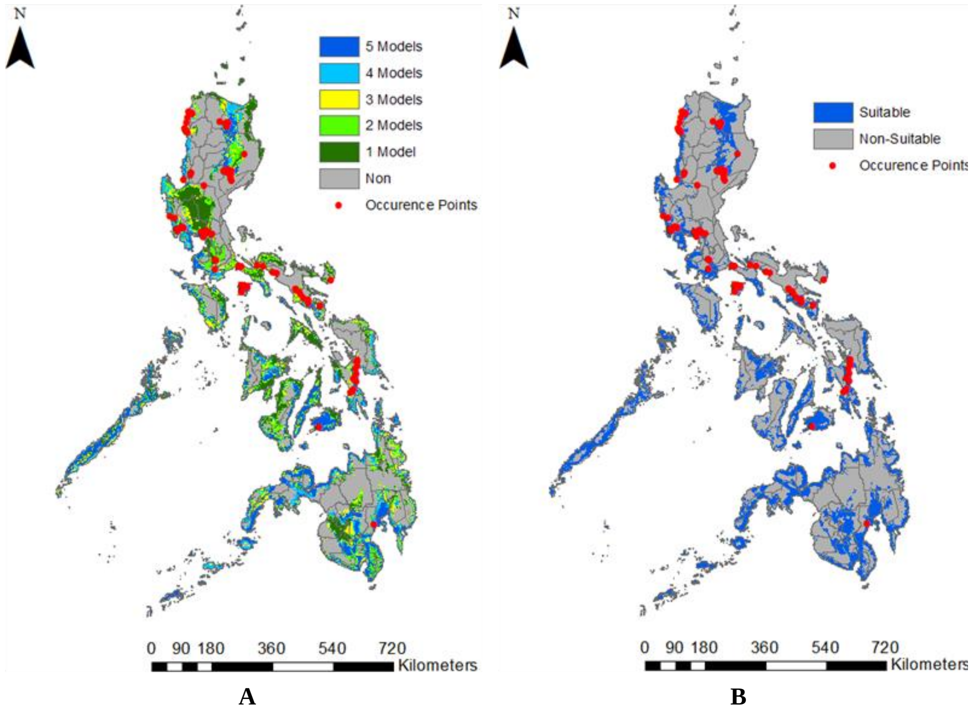


Figure 8. Combined present-day climatic models for Musa (*Musa balbisiana*): A. A score of 5 means that all species distribution models agreed and identified that pixel as containing suitable bioclimatic conditions for the species, B. With three or more (at least 60% of) algorithms agree that the pixel contains suitable bioclimatic conditions for the species. Red dots represent the species occurrence data.

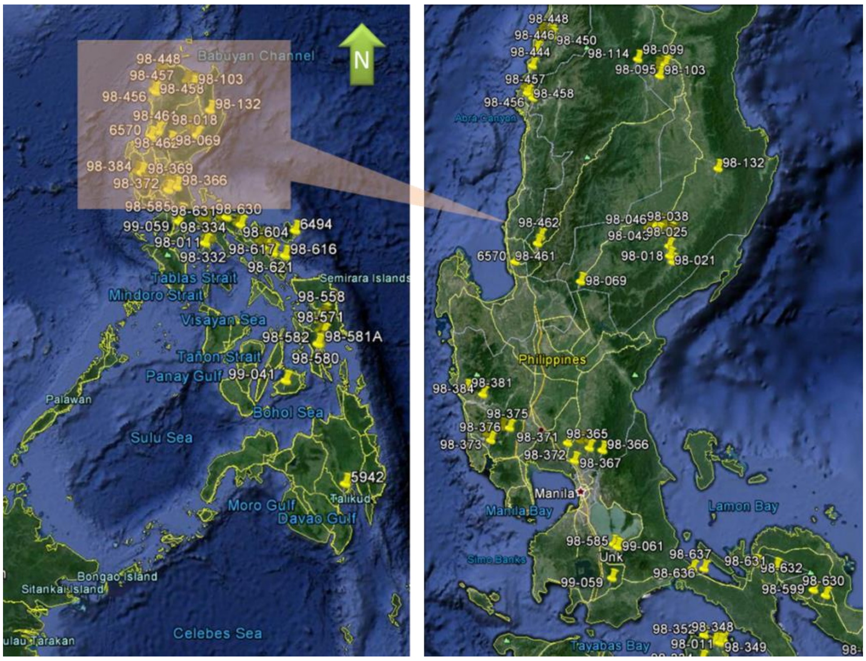


Figure 9. Map of *Musa balbisiana* germplasm collections showing accessions collected close to road networks. Maps were generated using Google Earth software

Collection bias due to infrastructure is common in collection missions because collectors tend to follow roads connected to main towns for reasons of efficiency, logistics and convenience (Hijmans et al. 2000) as well as safety and security. This bias was observed in many previously reported germplasm collections and studies (Hermann 1988; von Bothmer and Seberg 1995; Hijmans et al. 2000). For example, the Andean tuber crop collection sites were located largely in Pan-American Highway and other major roads in Ecuador (Hermann 1988) and collection sites in the distribution map of *Elymus cordilleranus* were mostly around major cities like La Paz (Bolivia), Lima (Peru) and the Pan-American Highway in Ecuador (von Bothmer and Seberg 1995). Similarly, 68% of the total germplasm holdings of Bolivian wild potatoes collection were collected within 2 km from the nearest roads (Hijmans et al. 2000). In this study, partnering with local officials made collections trips efficient and productive owing to their knowledge of local dialects and potential locations where *Musa* germplasm materials thrive.

In conclusion, following the extensive analyses of *M. balbisiana* germplasm, totaling 97 accessions, study results indicate the need for more implementation of conservation efforts to enhance the diversity of conserved germplasm for posterity as well as improve the germplasm's utility for current and future breeding activities on *Musa* species through related studies described below : (i) Exploration and collection trips with focus on invariant traits found in the *M. balbisiana* germplasm collections. (ii) Conduct collection trips in the southern region of the country, particularly Mindanao and neighboring islands to further conserve and assess the diversity of *Musa* species in the Philippines. (iii) Assessment of phenotypic responses of *Musa* germplasm to various biotic and abiotic stresses with the objective of identifying potential donor parental lines for breeding programs. Accordingly, *M. balbisiana* a wild species of *Musa*, is known to have traits that make it resistant to devastating banana diseases. Its high diversity in terms of phenotypic traits, as revealed in this study, could be source of desired characteristics in future plant types. Evaluation of this germplasm to biotic and abiotic stresses would further provide wealth of information for plant breeders for the utilization of this germplasm to abaca improvement. Results of further studies into *M. balbisiana* species' genome structure, regulatory networks and allelic diversity would elucidate the development of improved banana cultivars having desirable quality traits and resistance to biotic and abiotic stresses.

ACKNOWLEDGEMENTS

We would like to acknowledge DOST-PCARRD for supplementing the UPLB-IPB funds for this study. We are thankful to Eladio Werba for assisting in the phenotyping activities and maintaining the field GenBank as well as Mr. IsidoroBanasihan for helping establish and maintain the *Musa* field GenBank. We also acknowledge Ms. Jennifer Jara-Rabara for reviewing and editing the manuscript. Funding for this study was provided in part by the Department of Science and Technology-Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development (DOST-PCARRD) and University of the Philippines Los Baños- Institute of Plant Breeding (UPLB-IPB). RCR and RCSdesigned and performed the experiment. RCR analyzed the data and drafted the manuscript. ES generate SDM models and RCS contributed to the writing of the manuscript.

REFERENCES

Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43: 1223-1232.

Altoveros NC, Borromeo TH. 2007. The state of the plant genetic resources for food and agriculture of the Philippines (1997-2006). A Country Report. Bureau of Plant Industry, Department of Agriculture, Manila, Philippines.

Austin MP, Van Niel KP. 2011. Improving species distribution models for climate change studies: Variable selection and scale. J Biogeogr 38: 1-8. DOI: 10.1111/j.1365-2699.2010.02416.x.

Bernardo FA, Umali DL. 1956. Possible sources of resistance to abaca mosaic and bunchy top. Phil Agricul 40: 277-84.

Bonkoungou EG. 2001. Biodiversity in drylands: Challenges and opportunities for conservation and sustainable use. In: Niamir-Fuller M (ed) Challenge Paper: The Global Drylands Initiative UNDP Drylands Development Centre, Nairobi, Kenya.

Breiman L. 2001. Random forests. Mach Learn 45: 5-32.

Churchill AC. 2011. *Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control. Mol Plant Pathol 12 (4): 307-328. DOI: 10.1111/j.1364-3703.2010.00672.x.

Davey MW, Gudimella R, Harikrishna JA, Sin LW, Khalid N, Keulemans J. 2013. A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter-and intra-specific *Musa* hybrids. BMC Genomics 14: 683. DOI: 10.1186/1471-2164-14-683.

De Vicente MC, Guzmán FA, Engels J, Rao VR. 2005. Genetic characterization and its use in decision-making for the conservation of crop germplasm. The role of biotechnology for the characterisation and conservation of crop, forestry, animal and fishery genetic resources. Turin, Italy, 5-7 March 2005.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JR, Gruber B, Lafourcade B, Leitao PJ, Münkemüller T. 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography 36 (1): 27-46. DOI: 10.1111/j.1600-0587.2012.07348.x.

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. J Anim Ecol 77 (4): 802-813. DOI: 10.1111/j.1365-2656.2008.01390.x.

Fick SE, Hijmans RJ. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37: 4302-15. DOI: 10.1002/joc.5086.

Fielder H, Brotherton P, Hosking J, Hopkins JJ, Ford-Lloyd B, Maxted N. 2015. Enhancing the conservation of crop wild relatives in England. PLoS One 10 (6): e0130804. DOI: 10.1371/journal.pone.0130804.

Food and Agriculture Organization of the United Nations (FAO). 2017. Banana market review 2015-2016. FAO, Rome.

Food and Agriculture Organization of the United Nations (FAO). 2014. Banana market review and banana statistics 2012-2013. Rome. Accessed at http://www.fao.org/docrep/019/i3627e/i3627e.pdf, on 02.08.2020.

Guarino L. 1995. Mapping the ecogeographic distribution of biodiversity. Collecting plant genetic diversity. Technical guidelines. CAB International, Wallingford, United Kingdom.

Hajjar R, Hodgkin T. 2007. The use of wild relatives in crop improvement: A survey of developments over the last 20 years. Euphytica 156: 1-13. DOI: 10.1007/s10681-007-9363-0.

Hermann M. 1988. Progress report of the IBPGR research project on the genetic resources of Andean tuber crops. AGP-IBPGR report 892.

Heslop-Harrison JS, Schwarzacher T. 2007. Domestication, genomics and the future for banana. Ann Bot 100 (5): 1073-1084. DOI: 10.1093/aob/mcm191.

Hijmans RJ, Garrett K, Huaman Z, Zhang D, Schreuder M, Bonierbale M. 2000. Assessing the geographic representativeness of GenBank collections: The case of Bolivian wild potatoes. Conserv Biol 14: 1755-1765.

Hijmans RJ, Spooner DM. 2001. Geographic distribution of wild potato species. Am J Bot 88 (11): 2101-2112. DOI: 10.2307/3558435.

Hijmans RJ, Guarino L, Cruz M, Rojas E. 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. Plant Genet Resour Newslett 127: 15-19.

Hölscher D, Dhakshinamoorthy S, Alexandrov T, Becker M, Bretschneider T, Buerkert A, Crecelius AC, De Waele D, Elsen A, Heckel DG, Heklau H, Hertweck C, Kai M, Knop K, Krafft C, Maddula RK, Matthäus C, Popp J, Schneider B, Schubert US, Sikora RA, Svatoš A, Swennen RL. 2014. Phenalenone-type phytoalexins mediate resistance of banana plants (*Musa* spp.) to the burrowing nematode *Radopholus similis*. Proc Natl Acad Sci USA 111: 105-110.

Hunter D, Heywood V. 2012. Crop wild relatives: A manual of in situ conservation. Routledge, London.

IPGRI-INIBAP/CIRAD. 1996. Descriptors for banana (*Musa* spp.). International Plant Genetic Resources Institute, Rome, Italy/International Network for the Improvement of Banana and Plantain, Montpellier, France/Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, France.

Jarvis A, Williams K, Williams D, Guarino L, Caballero P, Mottram G. 2005. Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. Genet Resour Crop Evol 52: 671-682. DOI: 10.1007/s10722-003-6020-x.

Kumakech A, Kiggundu A, Okori P. 2013. Reaction of *Musa balbisiana* to banana bacterial wilt infection. Afr Crop Sci J 21: 337-346.

Lalusin AG, Villavicencio MLH. 2015. Abaca (*Musa textilis* Nee) breeding in the Philippines. In: Cruz VMV,Dierig DA (eds)Industrial crops: Breeding for bioenergy and bioproducts. Springer Science+Business Media, New York.

Lane A. 2006. An introduction to crop wild relatives. In: Raymond DR, Moore C (eds) Crop wild relatives. Bioversity International, Rome, Italy.

Leathwick JR, Elith J, Hastie T. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecol Model 199: 188-196. DOI: 10.1016/j.ecolmodel.2006.05.022.

Liu C, Berry PM, Dawson TP, Pearson RG. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28 (3): 385-393. DOI: 10.1111/j.0906-7590.2005.03957.x.

Manel S, Williams HC, Ormerod SJ. 2001. Evaluating presence–absence models in ecology: The need to account for prevalence. J Appl Ecol 38 (5): 921-931. DOI: 10.1046/j.1365-2664.2001.00647.x

McCouch SR, McNally KL, Wang W, Hamilton RS. 2012. Genomics of gene banks: A case study in rice. Am J Bot 99 (2): 407-423. DOI: 10.3732/ajb.1100385.

Oliveira-Filho AT, Fontes MAL. 2000. Patterns of floristic differentiation among Atlantic forests in Southeastern Brazil and the influence of Climate. Biotropica 32 (4): 793-810. DOI: 10.1111/j.1744-7429.2000.tb00619.x.

Parra-Quijano M, Iriondo JM, Torres E. 2012. Ecogeographical land characterization maps as a tool for assessing plant adaptation and their implications in agrobiodiversity studies. Genet Resour Crop Evol 59: 205-217. DOI: 10.1007/s10722-011-9676-7.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecol Model 190: 231-259. DOI: 10.1016/j.ecolmodel.2005.03.026.

Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography 31 (2): 161-175. DOI: 10.1111/j.0906-7590.2008.5203.x

PwC (PricewaterhouseCoopers LLP). 2013. Crop wild relatives: A valuable resource for crop development. PricewaterhouseCoopers LLP, United Kingdom. Accessed at http://pwc.blogs.com/files/pwc-seed-bank-analysis-for-msb-0713.pdf, on 02.08.2020.

Pradhan P. 2016. Strengthening MaxEnt modelling through screening of redundant explanatory bioclimatic variables with variance inflation factor analysis. Researcher 8 (5): 29-34. DOI: 10.7537/marsrsj080516.05.

Rabara RC, Ferrer MC, Calayugan MIC, Duldulao MD, Jara-Rabara J. 2015. Conservation of rice genetic resources for food security. Adv Food Technol Nutr Sci SE (1): S51-S56. DOI: 10.17140/AFTNSOJ-SE-1-108.

Rabara RC, Ferrer MC, Diaz CL, Newingham MCV, Romero GO. 2014. Phenotypic diversity of farmers' traditional rice varieties in the Philippines. Agronomy 4: 217-241. DOI: 10.3390/agronomy4020217.

Ramirez J, Jarvis A, Van den Bergh I, Staver C, Turner D. 2011. Changing climates: Effects on growing conditions for banana and plantain (*Musa* spp.) and possible responses. In: Yadav SS, Redden RJ, Hatfield JL, Lotze-Campen H, Hall AE (eds) Crop adaptation to climate change. John Wiley & Sons, Ltd.

Ramírez-Villegas J, Khoury C, Jarvis A, Debouck DG,Guarino L. 2010. A gap analysis methodology for collecting crop genepools: A case study with *Phaseolus* beans. PLoS One 5 (10): e13497. DOI: 10.1371/journal.pone.0013497.

Ramirez-Villegas J, Khoury CK, Achicanoy HA, Mendez AC, Diaz MV, Sosa CC, Debouck DG, Kehel Z, Guarino L. 2020. A gap analysis modelling framework to prioritize collecting for ex-situ conservation of crop landraces. Divers Distrib 6: 730-742. DOI: 10.1111/ddi.13046

Rohlf FJ. 2002. NTSYSpc: Numerical taxonomy system, ver. 2.1. Exeter Publishing Ltd., Setauket, NY.

Salas EAL, Seamster VA, Harings NM, Boykin KG, Alvarez G, Dixon KW. 2017. Projected future bioclimate-envelope suitability for reptile and amphibian species of concern in South Central USA. Herpetol Conserv Biol 12: 522-547.

Sillero N, Barbosa AM. 2020. Common mistakes in ecological niche models. Intl J Geogr Inf Sci 27: 1-4. DOI: 10.1080/13658816.2020.1798968.

Silva-Flores R, Pérez-Verdín G, Wehenkel C. 2014. Patterns of tree species diversity in relation to climatic factors on the Sierra Madre Occidental, Mexico. PLoS One 9 (8): e105034. DOI: 10.1371/journal.pone.0105034.

Sotto RC, Rabara RC. 2000. Morphological diversity of *Musa balbisiana* Colla in the Philippines. Info Musa 9: 28-30.

Subbaraya U, Lutaladio NB, Baudoin WO. 2006. Farmers'knowledge of wild *Musa* in India. FAO, Rome.

Swets JA. 1988. Measuring the accuracy of diagnostic systems. Sci 240: 1285-1293.

Talbert C. Software for assisted habitat modeling package for VisTrails (SAHM: VisTrails). Ver. 1.

Taylor R. 1990. Interpretation of the correlation coefficient: A basic review. J Diagn Med Sonogr 6: 35-39.

Tyack N, Dempewolf H. 2015. The economics of crop wild relatives under climate change. In: Redden R, Yadav SS, Maxted N, Dulloo ME, Guarino L, Smith P (eds) Crop Wild Relatives and Climate Change, John Wiley & Sons, Inc., New York.

van Zonneveld M, Thomas E, Galluzzi G, Scheldeman X. 2011. Mapping the ecogeographic distribution of biodiversity and GIS tools for plant germplasm collectors. In: Collecting plant genetic diversity: Technical guidelines – 2011 update, CAB International, Wallingford, United Kingdom.

Vincent H, Amri A, Castañeda-Álvarez NP, Dempewolf H, Dulloo E, Guarino L, Hole D, Mba C, Toledo A, Maxted N. 2019. Modeling of crop wild relative species identifies areas globally for in situ conservation. Commun Biol 2 (1): 136. DOI: 10.1038/s42003-019-0372-z

von Bothmer R, Seberg O. 1995. Strategies for the collecting of wild species. In: Collecting plant genetic diversity: Technical guidelines. CAB International, Wallingford, United Kingdom.

Warren DL, Seifert SN. Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. Ecol Appl 21 (2): 335-342.

Zhang H, Mittal N, Leamy LJ, Barazani O, Song BH. 2017. Back into the wild-Apply untapped genetic diversity of wild relatives for crop improvement. Evol Appl 10 (1): 5-24. DOI: 10.1111/eva.12434

Zhao Y, Cao H, Xu W, Chen G, Lian J, Du Y, Ma K. 2018. Contributions of precipitation and temperature to the large-scale geographic distribution of fleshy-fruited plant species: Growth form matters. Sci Rep 8: 17017. DOI: 10.1038/s41598-018-35436-x

Zurell D, Franklin J, König C, Bouchet PJ, Dormann CF, Elith J, Fandos G, Feng X, Guillera-Arroita G, Guisan A, Lahoz-Monfort JJ. 2020. A standard protocol for reporting species distribution models. Ecography 43: 1-17. DOI: 10.1111/ecog.04960