Elodie Stoetman
s1966529

Modelling the Effects of Future Climate Changes on the Distribution of the Banana Plant *Musa Paradisiaca*.

Introduction

Musa is a plant genus of extraordinary significance providing many important foods in the world today. One such important food is the tropical Banana fruit, one of the first harvested fruits in primitive agriculture and remaining as one of the most important sources of calories for human diets worldwide. Edible *Musa* plants are characterized into different genomic groups and their wide diversity allows for a potentially wide range of different applications. The current most commonly cultivated *Musa* species include the dessert bananas (AA or AAA) and the cooking bananas or plantains (AAB, ABB or BBB).[1] The modern cultivated *Musa* species, *Musa Paradisiaca*, is a hybrid of two wild Banana plant species *Musa Acuminata* and *Musa Balbisiana* and the relative proportion of the genomes from these two gives different phenotypes, such as increasing proportions of B chromosome in the hybrids genome from *Musa Balbisiana* encoding increased drought resistance.[2]

The *Musa* species is natively distributed throughout the Indo-Malaysian, Asian and Australian tropical habitats. However, they are currently also being cultivated within the humid tropics and subtropics in the Americas and Africa and associate with many lowland forest inhabitants of all types.[3]

Tropical habitats represent a relatively stable environment throughout the year and this stability is needed to maintain the cultivation of Banana as a staple food, not only for modern day diets but also as ethnobiological importance to many cultures, particularly those throughout Indo-Malaysia.[4] With the current climate models predicting an overall average yearly increase in temperature, the potential changes in our future climate can have negative impacts on the native environment and climate required for the staple Banana fruit plants. By using species distribution modelling (SDM) to find correlations between *Musa paradisiaca* presence data and abiotic (bioclimatic) variables, the relative changes in *Musa Paradisiaca* species distribution can be modelled within these predicted future climatic changes. This information can then provide important insight into whether certain phenotypic traits may need to be selected for to help maintain *Musa Paradisiaca* cultivations in the pantropical areas. The hypothesis being, that an increase in global average temperatures could increase drought and reduce available suitable habitat which in turn can lead to a loss of overall tropical climate niche required for the current modern cultivated *Musa* species, *Musa Paradisiaca*.

Methodology

Occurrence Data Preprocessing and Cleanup:

Occurrence data for the *Musa* species, *Musa Paradisiaca* was retrieved from the GBIF database (Available from: https://www.gbif.org [3 December 2019). Occurrence data with location information and added as a human observation were downloaded. Visualization of the occurrence data was done using QGIS (QGIS Geographic Information System. Open Source Geospatial Foundation Project. https://qgis.osgeo.org) to check whether the occurrence data accurately reflected the ecological niche and no major spatial biases were present.

Overlapping coordinates were merged to one observation and from the remaining coordinates a minimum of two observations had to be made in one country for the occurrence data to be taken into account. These criteria removed singular observations in countries unlikely to represent the species natural ecological niche (such as Belgium, Germany and Croatia). The remaining 616 observations for *Musa Paradisiaca* were used in the model. No further biases were found within the data and the remaining observations were all distributed along the expected tropical regions (data not shown).

Environmental Data

Global climate data version 1.4 at 5 minutes spatial resolution was downloaded from WorldClim.[5,6] Bioclimate variables taken into the model included BIO1 (annual mean temperature), BIO4 (temperature seasonality), BIO12 (annual precipitation), BIO15 (precipitation seasonality) and BIO17 (precipitation of driest quarter). *Musa* species require a high minimum annual precipitation (around 2000 mm), temperature stability and are sensitive to drought. These variables were chosen to reflect the potential changes in these climate factors as a product of increased emissions, being of major importance to the stability of the natural habitat for the *Musa* species.

Collinearity tests using Pearon's pairwise correlation coefficient showed a strong autocorrelation between BIO4 and BIO7 and this was confirmed with the variance inflation factor (VIF, BIO4 = 15.2 and BIO7 = 16.5). Therefore, BIO7 was removed from the model as temperature seasonality (BIO4) is expected to be a better representative measure for temperature fluctuation than the annual range which is expected to be small for tropical climates.

Modelling Parameters

Musa Paradisiaca species distribution modelling (SDM) was performed using Maximum Entropy Modeling of Species Geographic Distributions in Maxent[7] version 3.4.0 using the above climate variables for modelling present and future scenarios for both world and area of interest (AOI). Future modelling was done for all four representative concentration pathways (rcp's) of greenhouse gasses in 50 years. The Jackknife setting was enabled and modelling was performed using a random seed with 10 replicates, the remaining options in Maxent were set to default values.

Results

Model Performance

The omission on test samples shows a good match to the predicted omission rate (Figure 2A), indicating that the test and training data are independent and are not spatially autocorrelated. The ROC curve (AUC) shows the model was a good fit for the training data and has a strong predictive power (Figure 2B), performing better than random..

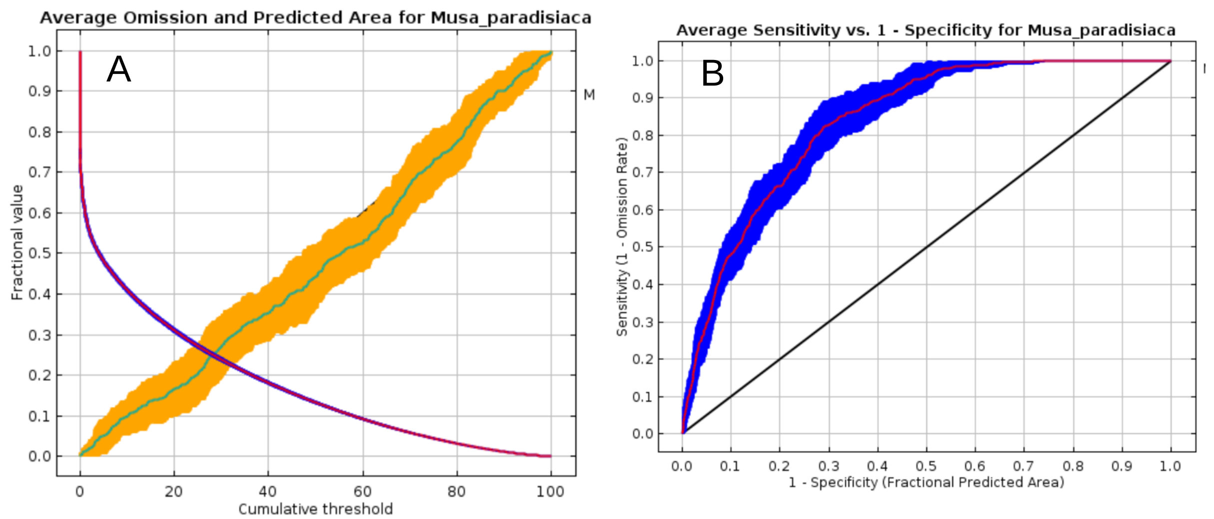

Figure 1: Model Performance. **A)** The effects of testing and training omission and their comparison to the predicted omission with the cumulative threshold (yellow: mean omission +- 1 STDev, Light Blue: Mean omission on test data +- 1 STDev, Red: mean area, Black: predicted omission) . **B)** receiver operating characteristic (ROC) curve of current model and random model AUC values (Red: mean AUC = 0.839, Blue: mean +- 1 STDev, Black: random prediction).

Overall model performance showed high predictive power with an AUC of around 0.84 using all climatic variables BIO1, BIO4, BIO12, BIO15 and BIO17 (Figure 2A). Jackknife tests results show that the climate variables BIO4 (temperature seasonality), BIO12 (Annual precipitation) and BIO17 (Precipitation of driest quarter) are the most effective variables for predicting the distribution of the occurrence data used as the test set (Table 1). Where BIO12 was able to correctly predict ~75% of test occurrence data as a single predictor (Figure 2A). These results can also be seen in the corresponding response curves (Figure 2B) where BIO4 shows a negative correlation with *Musa Paradisiaca* occurrence data. BIO1 (annual mean temperature) and BIO15 (precipitation seasonality) show much less predictive power however the relative performance of the model does not improve if these variables are removed as their corresponding AUC is lower than the model with all variables.

Variable	Percent contribution	Permutation importance
_Bio12	46.7	15.1
_Bio04	21.7	16.1
_Bio17	20.5	48.3
_Bio01	7.3	8.3
_Bio15	3.8	12.2

Table 1: Percent contribution and permutation importance for each climate variable

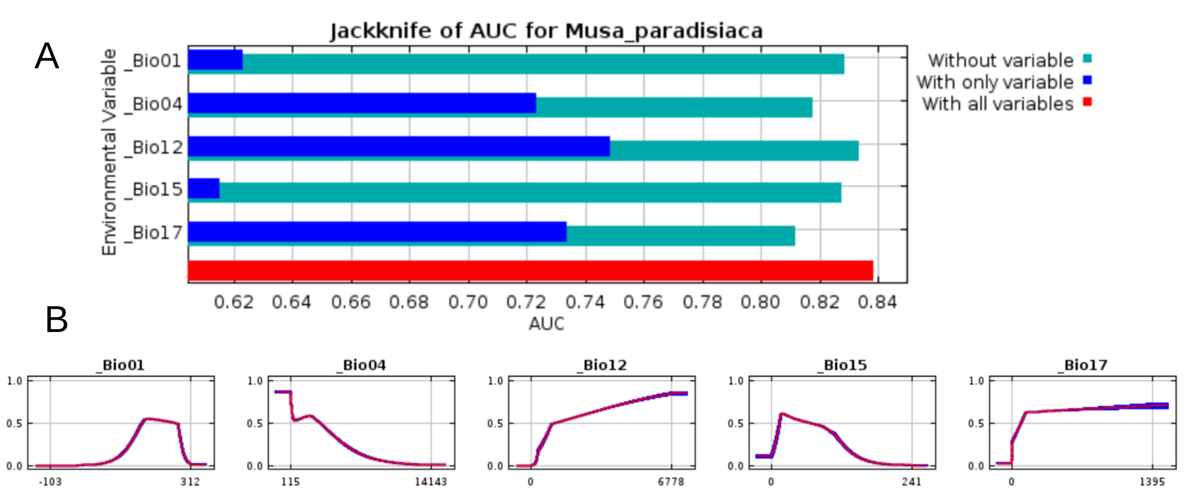

Figure 2: A) Jackknife results using AUC as measure of predictive performance **B)** Response curves per individual climate variable as the singular variable in the model.

Musa Paradisiaca Present World Projections

Occurrence predictions on the world map corresponded to the predictions within the AOI environmental layers used in the model, where highest occurrence of *Musa Paradisiaca* was predicted within the tropical regions around the equator (Figure 3A and 3B). However, a considerable inaccuracy in the occurrence prediction in Western Europe can be seen through Ireland and Norway western coastal regions representing suitable conditions. The model seems to show a slight bias towards coastal regions in general (Figure 3B).

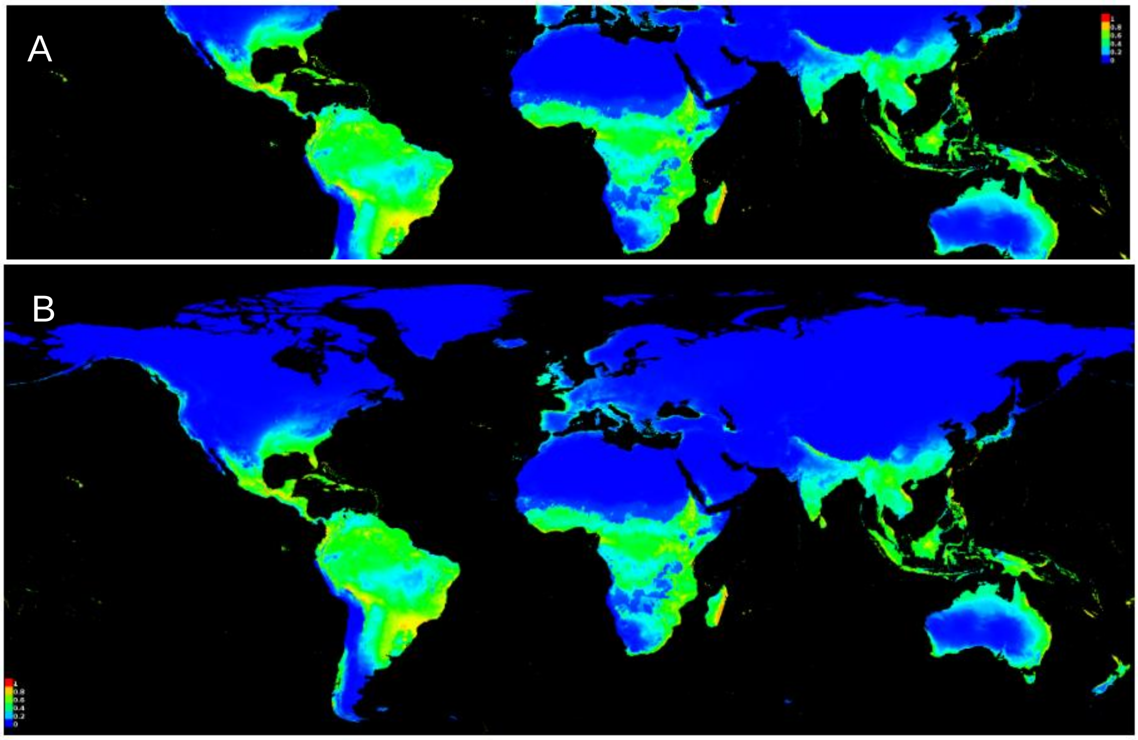

Figure 3: Point-wise mean and standard deviation of the 10 output grids from the model. **A).** Area of interest projection. **B).** World projection. Warmer colors show areas with better predicted conditions. White dots show the observation locations used for training, while violet dots show test locations.

Musa Paradisiaca Future Projections

Figure 4A shows a slight promotion in suitable environmental habitat for *Musa Paradisiaca* as a result of climate change projections under the most extreme scenario (he85bi50). However, plotting the difference in area gained and lost seems to reflect a balance between the two (Figure 4B).

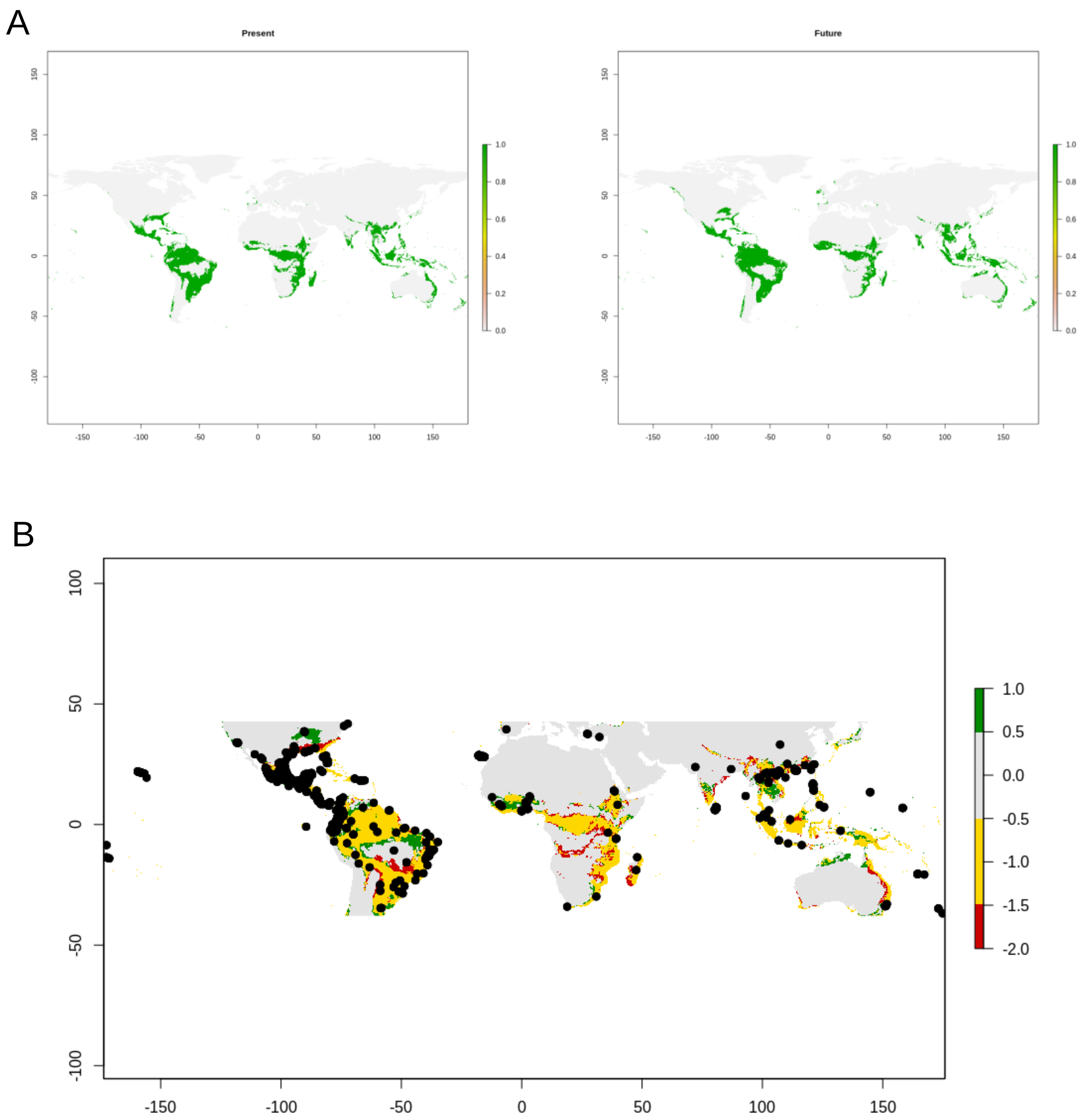

Figure 4: A). Threshold binary maps of present and future distributions of *Musa Paradisiaca*. Average of maximum training sensitivity plus specificity threshold outputted by Maxent for each replicate was 0.3891 for 10 replicates. **B).** Difference between present and future habitat area at occurrence scale, colors indicate red = lost, green = gained, yellow = remains suitable. Points show occurrence data for *Musa Paradisiaca* used in the model.

Discussion

Model performance showed a relatively high predictive power with an AUC of around 0.84 using the climatic variables BIO1, BIO4, BIO12, BIO15 and BIO17. Annual precipitation (BIO12) had the most predictive power (AUC 0.75) and contributed towards around 46.7% of the model. This high correlation with annual precipitation could be the cause of some unexpected regions showing high occurrence prediction (e.g. West coast of Ireland, France and Norway) as these regions also experience high annual precipitation. The second most predictive variable was temperature seasonality which is expected to decrease occurrence prediction for *Musa Paradisiaca* in areas with large seasonal temperature fluctuations and this can indeed be seen in the model predictions where areas with extreme temperatures such as Northern Canada, Greenland and North-East Russia represent the geographical areas least correlated with *Musa Paradisiaca* occurrence data. The bias towards coastal region could be due to spatial biases as many locations of the occurrence data were situated along the coastal regions in tropical habitats.

The set of available environmental variables may not be sufficient to describe all the parameters of the species' fundamental niche that are relevant to its distribution. Other variables which may be of importance, but were not taken into account here, are edaphic factors such as soil fertility (pH or nitrogen content). These factors are known to affect annual precipitation requirements for *Musa Paradisiaca* occurrence as well as plants in general (citation musa species pdf, Dubuis). The AUC values were computed using presence data only and thus the maximum achievable AUC is less than 1. The AUC values should therefore be seen as an index of habitat suitability as their optimality cannot be accurately determined. Using additional algorithms, such as generalized linear models (GLM), to model the occurrence data can lead to improved AUC determination and could produce better predictions than solely the generative approach of the maximum entropy modeling used here.[8]

Future projections in 50 years for all four emission scenarios showed a gradual increase in occurrence of *Musa Paradisiaca* with increasing emissions. This increase in geographical space could be due to the gradual rise in water-holding capacity causing an expected increase in the annual precipitation. However this gradual increase is minor and high fluctuations among the four scenarios as well as the final gain and loss of suitable habitat seem relatively constant. It therefore seems that drought in tropical areas is unlikely to increase with climate change and is not a threat towards the cultivation of *Musa Paradisiaca*.

Conclusion

The species distribution model represented here was able to predict with a reasonable high level of accuracy the world occurrence distribution of *Musa Paradisiaca* where annual precipitation together with temperature seasonality had the highest predictive power. Even under the high emission scenario, the model predicted that the relative change in climate could be slightly favourable for *Musa Paradisiaca* distribution. However, despite the fact that this may be the case, the effects are minor and *Musa Paradisiaca* is unlikely to colonize these new areas on its own as cultivars of *Musa Paradisiaca* are usually sterile, without seeds or viable pollen. What can be inferred from these results is that given the worst case scenario for the future climate, the effects will not be harmful towards *Musa Paradisiaca* and its important role in human diet and as a staple food is not threatened by climatic factors.

References:

1. Bello-Pérez, L. A., Agama-Acevedo, E., Osorio-Díaz, P., Utrilla-Coello, R. G. & García-Suárez, F. J. Banana and Mango Flours. in *Flour and Breads and their Fortification in Health and Disease Prevention* 235–245 (Elsevier Inc., 2011). doi:10.1016/B978-0-12-380886-8.10022-4
2. Nelson, S., Ploetz, R. & Kepler, A. Musa species (banana and plantain). *Species profiles pacific Isl. agro For.* **2006**, (2006).
3. Heslop-Harrison, J. S. & Schwarzacher, T. Domestication, Genomics and the Future for Banana. *Ann. Bot.* **100**, 1073–1084 (2007).
4. Kennedy, J. & Kennedy, J. Bananas and People in the Homeland of Genus Musa: Not just pretty fruit. *Ethnobot. Res. Appl.* **7**, 179–197 (2009).
5. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
6. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
7. Phillips, S. J. & Schapire, R. E. *A Maximum Entropy Approach to Species Distribution Modeling*. (2004).
8. Phillips Steven J, Anderson Robert P & Schapire Robert E. Maximum entropy modeling of species geographic distributions - ScienceDirect. *Ecol. Modell.* **190**, 231–259 (2006).