Project Update: December 2018

1 – Summary of activities in the period

ACTIVITY	LOCAL	PERIOD	CITY, STATE	COUNTRY
Collection of occurrence records	Museum of Natural History, Biodiversity Institute – University of Kansas – KU	10/17 – 10/31/2018	Lawrence, KANSAS	United States of America
Validation and final organisation of collected data	Laboratory of Evolution and Macroecology – University of Kansas – KU	10/17 – 11/02/2018	Lawrence, KANSAS	United States of America
Data Cleaning	Laboratory of Evolution and Macroecology – University of Kansas – KU	11/03 – 11/12/2018	Lawrence, KANSAS	United States of America
Ecological Niche Modeling	Laboratory of Evolution and Macroecology – University of Kansas – KU	11/13 – 01/15/2019	Lawrence, KANSAS	United States of America

2 – Description of activities

Seeking a better understanding of the steps by the reviewers, this presentation will follow the subdivisions: (1) Data collection, (2) Data validation or Data filtering, (3) Data cleaning and (4) Ecological Niche Modeling.

(1) Each occurrence record obtained consist of a Latitude-Longitude pair denoting the location of a species. Specimens deposited (vouchers) in the main collections and museums were accessed through pre-programmed visits and the data assigned to each of them collected.

The following Museum was visited in this period: Herpetology Sector, Museum of Natural History, Biodiversity Institute, University of Kansas – KU, Lawrence, KS, United States of America.

Concomitantly to this routine, little accurate data were eliminated from the final set characterising the step (2) "Data validation".

(2) Mismatch between the voucher and its formal name automatically prevented the specimen to compose the dataset. Inconsistencies between the locality and its lat-long pair also eliminated the record of the final set of data. For this verification all records obtained were plotted on maps using specific programmes (e.g. Google Earth©).

(3) Considering the two previous steps, all the acquired data were submitted to a rigorous routine of "cleaning" following the steps suggested and implemented in a work recently published in the scientific literature. The advantage of applying this routine and obtaining "clean" data is to increase the robustness and reliability of the results obtained from the analyses in which these data were used. According to that work, this routine is structured in two consecutive steps as shown below: "a", Data Filtering or Thinning and; "B" Data Partitioning or Splitting.

Data Filtering, in short, consists of removing occurrence records so that two of them are not close to each other given a linear distance "x" (this value of "x" being selected based on aspects of the species biology and the study area). As a result the initial dataset is transformed into a subset where records that violated the pre-selected distance have been removed. Finally, the phase formally known as "splitting" consists of randomly dividing the occurrence records into two sets; a set defined as a set of training data (records that will be used in the calibration of ecological niche models) and another set of test data (records that will be used to validate ecological niche models) (see details below).

(4) The models will be developed with the formalism of ecological niche modeling (ENM). In this procedure it is established a relationship between the occurrence data for a given species and environmental variables (predictors) using a mathematical algorithm.

Thirty-one variables derived from temperature and precipitation obtained from the *"WorldClim Global Climate Data"* (*http://worldclim.org/*) and *"Climatologies at High Resolution for Earth's Land Surface Areas-CHELSA"* (*http://chelsa-climate.org/*) were gathered. Worldclim is a database that provides global climate layers generated by interpolations of actual climate data obtained by weather stations in 30 arc-second resolution grids. Similarly, Chelsa also consists of a high-resolution climate layers repository, however the derivation process from actual data and construction of each of them differs between the two databases. While the Worldclim layers are interpolations, in the construction of the Chelsa layers is employed an algorithm based on the process of *"statistical downscaling"*. Such differences culminate in different products that lead, consequently, to the different performances of the models and for this reason both datasets were used.

Additionally, two more predictors related to the availability of water in the environment were included in the set. The first layer representing the *"Potential Evapo-Transpiration-PET"* process and the second, *"Aridity Index-AI"*. Anurans are highly vulnerable organisms to sudden environmental changes, especially those related to the availability of water in the environment, so we believe that the use of these layers, representing this type of environmental limitation, could return interesting results regarding the niche model. Both variables were used in the modeling routines and obtained from the following repository: "Consultative Group on International Agricultural Research Consortium for Spatial Information - CGIAR-CSI" (*http://cgiar.community/*).

The ecological niches were modeled using the algorithm *"MaxEnt"*. Maxent was ranked among the most effective methods for ecological niche modeling from presence-only data and for this main reason was chosen. In the end, the resulting ecological niche models will "define" the environmental conditions that best describe the spatial boundaries for each species (i.e. their ecological requirements) and then simulate their potential geographic distributions.

Ecological Niche Modeling (ENM) is a framework composed of analytical tools that, over time, has potential applications in several fields such as: conservation planning, potential impacts of climate change, biological invasions, geography of disease transmission among many others. Given the relevance of the issues in which this technique is applied, the need for increasingly robust models, which better represent the phenomena being modeled has become (and still does) more and more required. The ideal fit to the data, and consequently to the phenomenon being studied, can be achieved with multiple parametrisations or, more formally, the *"Model Calibration"*. Although its implementation adds considerable complexity to the modeling process, calibration allows better models to be obtained and, for this reason, has been implemented here.

We use "*KUENM*", a recently launched "R" package that automates important steps in the ecological modeling of niches and potentiates the chances of obtaining more robust models. Briefly, this package implements consecutively three crucial phases in the ecological modeling routine: calibration, evaluation and selection, creation of final models and extrapolation.

3 – Visit to North American museum

"HERPETOLOGICAL COLLECTION - MUSEUM OF NATURAL HISTORY, BIODIVERSITY INSTITUTE" UNIVERSTIY OF KANSAS - KU, LAWRENCE, KANSAS, USA.
Curator: Rafe M. Brown and Rich Glor
Collection Manager: Luke Welton

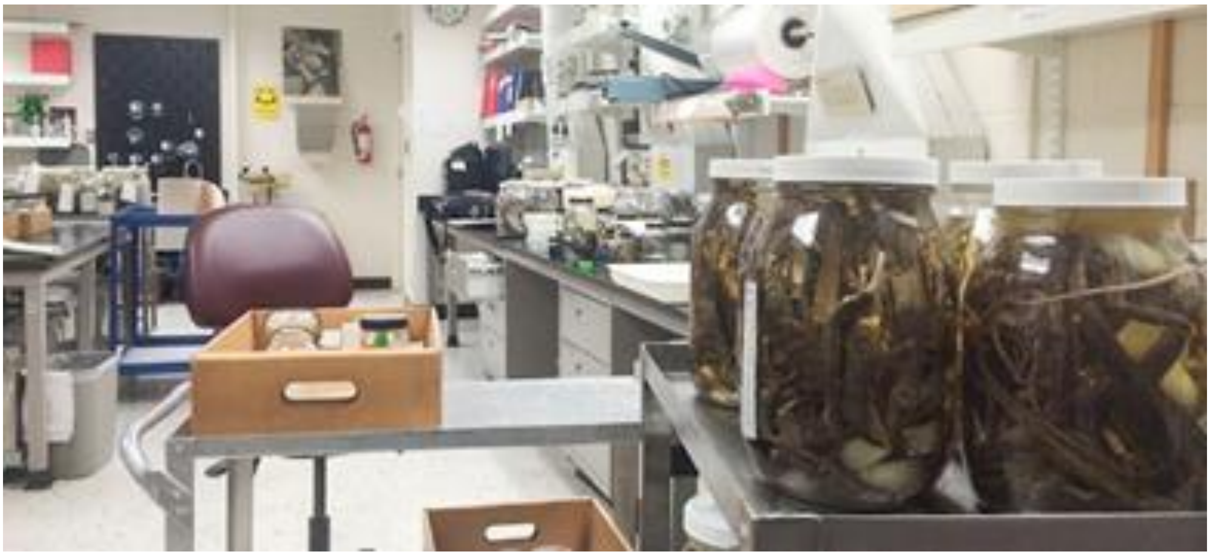


Figure 1: Inside the Herpetological Collection of Natural History Museum – Biodiversity Institute – University of Kansas – KU, Lawrence, KS, United States of America.

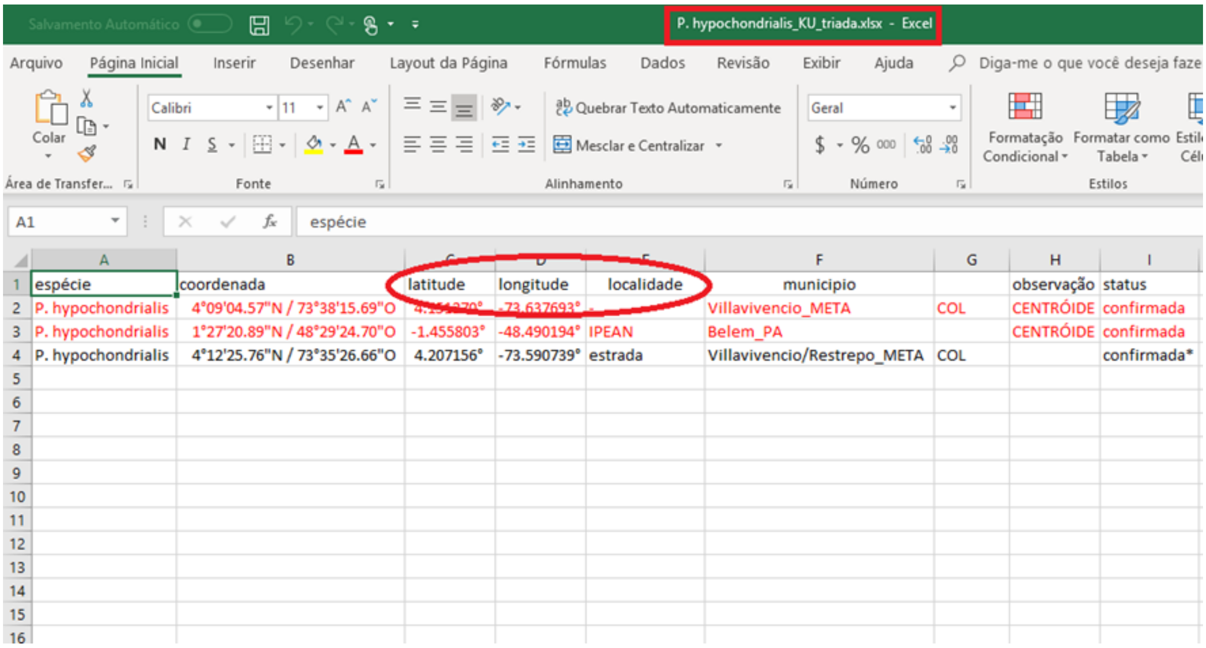
Figure 2: Facade of the Museum of Natural History, Institute of Biodiversity – University of Kansas, Lawrence, KS, United States of America.

Visit Period: 16 to 31 October 2018

Specimens Checked: A total of 313 specimens belonging to the following *Pithecopus* species were analysed: *P. ayeaye*, *P. azureus*, *P. hypochondrialis, P. palliatus* and *P. rohdei*. Specifically, one specimen of *P. ayeaye* and two of *P. azureus* were checked and none of them presented misidentifications. 213 individuals classified as *P. hypochondrialis* deposited in the Natural History Museum at KU were checked and of that total six specimens presented some problems regarding their classification. 91 specimens of *P. palliatus* were checked and none of them presented taxonomic failures. Finally, seven *P. rohdei* are collected in the museum. These specimens were analysed, and their formal taxonomic identification confirmed.

Figure 3: (clockwise). Standard label containing the basic information of the specimen. Some pots with specimens checked. Individual of P. palliatus *analyzed in dorsal and lateral view.*

Records of Occurrence: After the taxonomic screening and check of the collected specimens, I collected the information of direct interest to my project, points or records of occurrence, linked to everyone through the analysis of collection´s register books. Each record obtained was processed noted and compiled into individual worksheets (see example below, figure 4).



	A	B	C	D	E	F	G	H	I
1	espécie	coordenada	latitude	longitude	localidade	municipio		observação	status
2	P. hypochondrialis	4°09'04.57"N / 73°38'15.69"O	4.151270°	-73.637693°	-	Villavivencio_META	COL	CENTRÓIDE	confirmada
3	P. hypochondrialis	1°27'20.89"N / 48°29'24.70"O	-1.455803°	-48.490194°	IPEAN	Belem_PA		CENTRÓIDE	confirmada
4	P. hypochondrialis	4°12'25.76"N / 73°35'26.66"O	4.207156°	-73.590739°	estrada	Villavivencio/Restrepo_META	COL		confirmada*
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									

Figure 4: Example of worksheet constructed after the analysis of the specimens in the collection.

Verification and Validation: After obtaining and organising the worksheet with the data, each of the records was verified and validated. The verification process consisted of the analysis of the agreement between the coordinates provided and the registered location that was designated by the collector. After verification, we classified each specific record as appropriate or not to compose the final set of data. In addition, for valid records whose coordinates were not available a manual search was done using Google Earth© and / or by checking in Gazetteers.

Results: From the visit to the Natural History Museum – Biodiversity Institute - KU, we obtained a total of 14 valid occurrence records for the five species (*P. ayeyae*, *P. azureus*, *P. hypochondrialis*, *P. palliatus* and, *P. rohdei*). Specifically, we have one record for *P. ayeyae*, two for *P. azureus*, three occurrence records valid for *P. hypochondrialis*, five for *P. palliatus* and three valid records for *P. rohdei*. (See map below, figure 5).

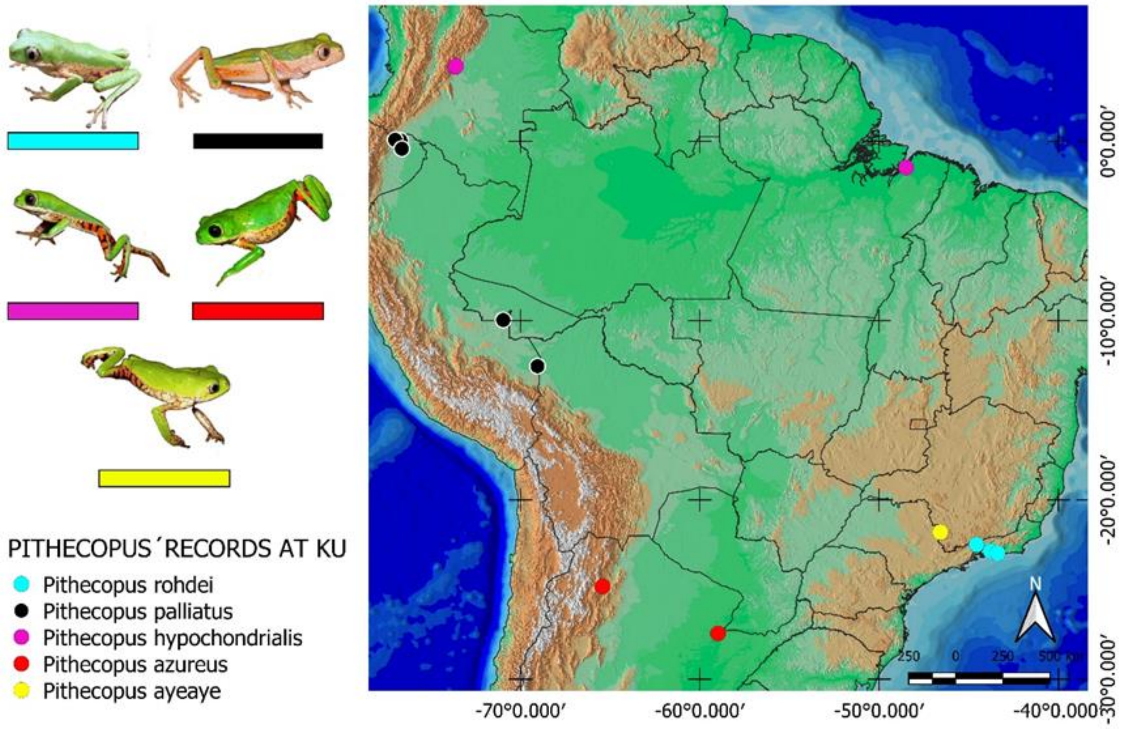


Figure 5: Illustrative map showing the locations obtained for each species through the visit to Natural History Museum at KU.

<u>4 – Visits, briefly, in numbers</u>

Aiming at a more objective presentation of all the work that has been done throughout the early stages of the project supported by "THE RUFFORD FOUNDATION ©", I hereby present compilations regarding the total number of individuals accessed and analysed for each of the species in each of the eight collections visited (see figure 6).

With the same objective of synthesis, I illustrated the number of individuals, among the total, who needed a correction. Corrections corresponded mainly to taxonomic errors associated with the analysed vouchers. Such information generally expresses the importance of the visits that have been set up as a vital step in the routine of data collection (see figure below, figure 7).

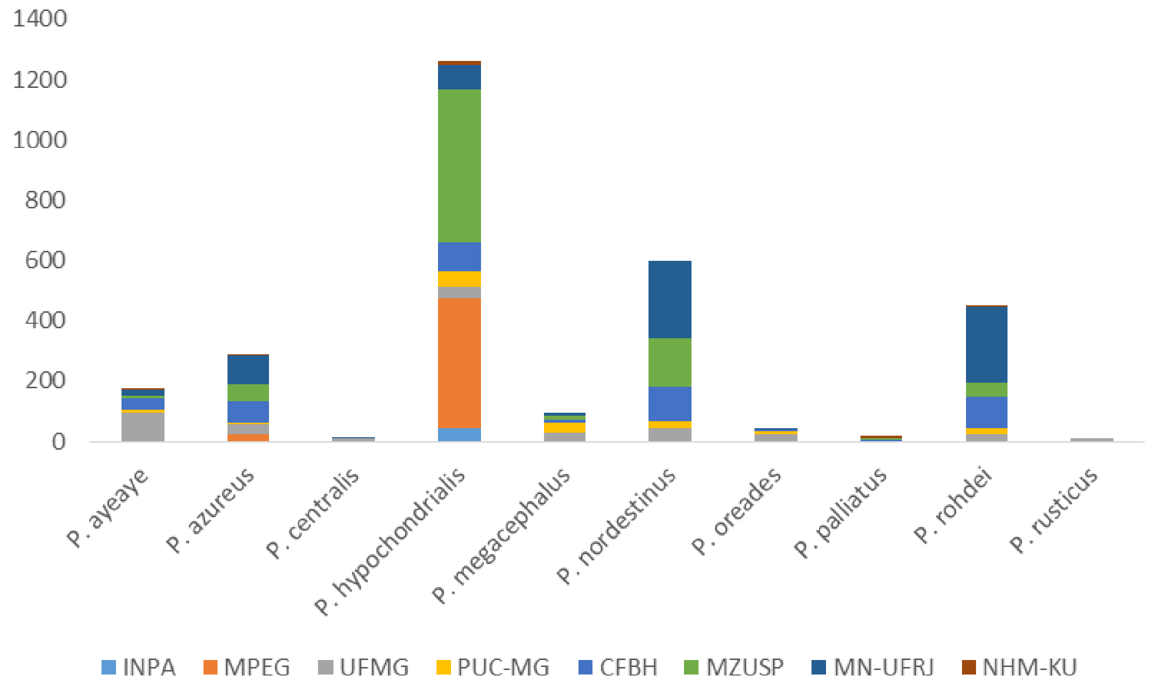


Figure 6: Graphical representation of total specimens checked, number of specimens per species and number of specimens per species per collection accessed and checked after visits to collections and museums.

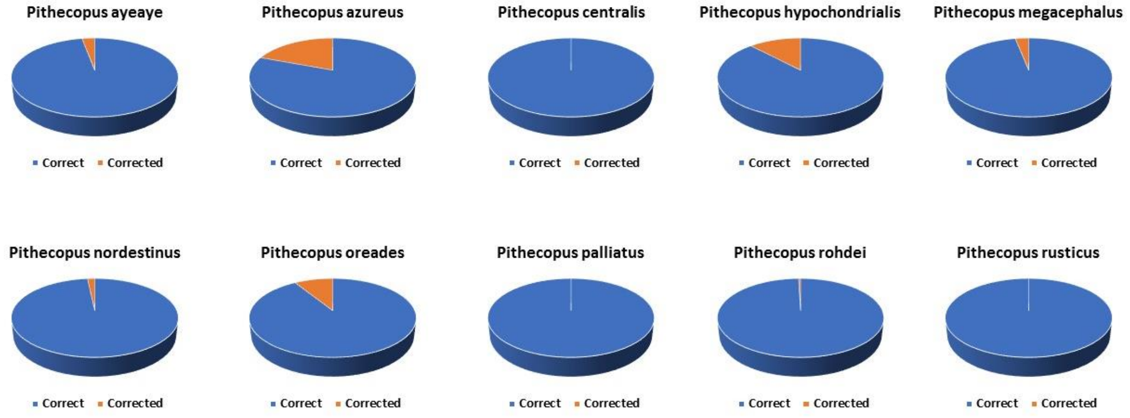


Figure 7: Graphical representation of the number of corrections (taxonomic) performed among the total number of individuals accessed. Information given by species.

Therefore, of the 2949 individuals belonging to the 10 species of the genus (100%) analysed 232 (approximately 7.9%) required a reassessment of their taxonomic attributions. As expected, the species that presented the highest number of errors were those that had more specimens collected by the collections and museums. Among all, *Pithecopus azureus* and *Pithecopus hypochondrialis* were species that presented a considerable number of individuals classified erroneously (in relation to the total of analysed specimens). Although they were intensively sampled species, which could justify such numbers of errors (as discussed above), according to the literature relevant to the taxonomy of the group, such species have a highly complex taxonomic attribution of their individuals mainly due to the high number of diagnostic characters shared with other species throughout its geographic distributions.

In general, the last synthesis that is presented here validates the need for cautious data sorting, mainly for applications such as ecological niche modeling and predictions of potential distributions for species, since it shows that even among the most important and constantly revised collections errors still endure. Moreover, the compilation presented here, although simple and superficial, justifies the use of resources for and evidences the importance of making visits to zoological collections and museums, a practice that has been increasingly neglected by researchers and routinely classified as unnecessary.

5 – The Niches on models

As a result of the consecutive accomplishment of all steps previously discussed, I present some of the ecological niche models constructed for some of the species of the genus approached by my project (figures 8, 9 and 10). As already shown, all models were built on an intensive routine that has been improved over the years. Namely, each of the models were constructed using occurrence records that were previously analysed (in an attempt to use the most reliable data possible), environmental layers (predictors) with the maximum possible resolution and that were always selected taking into account their importance for the biology and ecology of the species and application of an algorithm (which correlates geography and environment to simulate the niche) that best suited the question pursued by me. Finally, what is presented here are results of this modeling routine, the best models among a massive number of candidate models available according to the previously defined configurations. Therefore, I believe that our results carry reliable information about the ecological facet of these species, which can and will be used to understand theoretical (e.g. insights on *Pithecopus* genus diversification) and practices (e.g. how guide information in the planning or base in the improvement of strategies for the conservation of the biodiversity) issues.

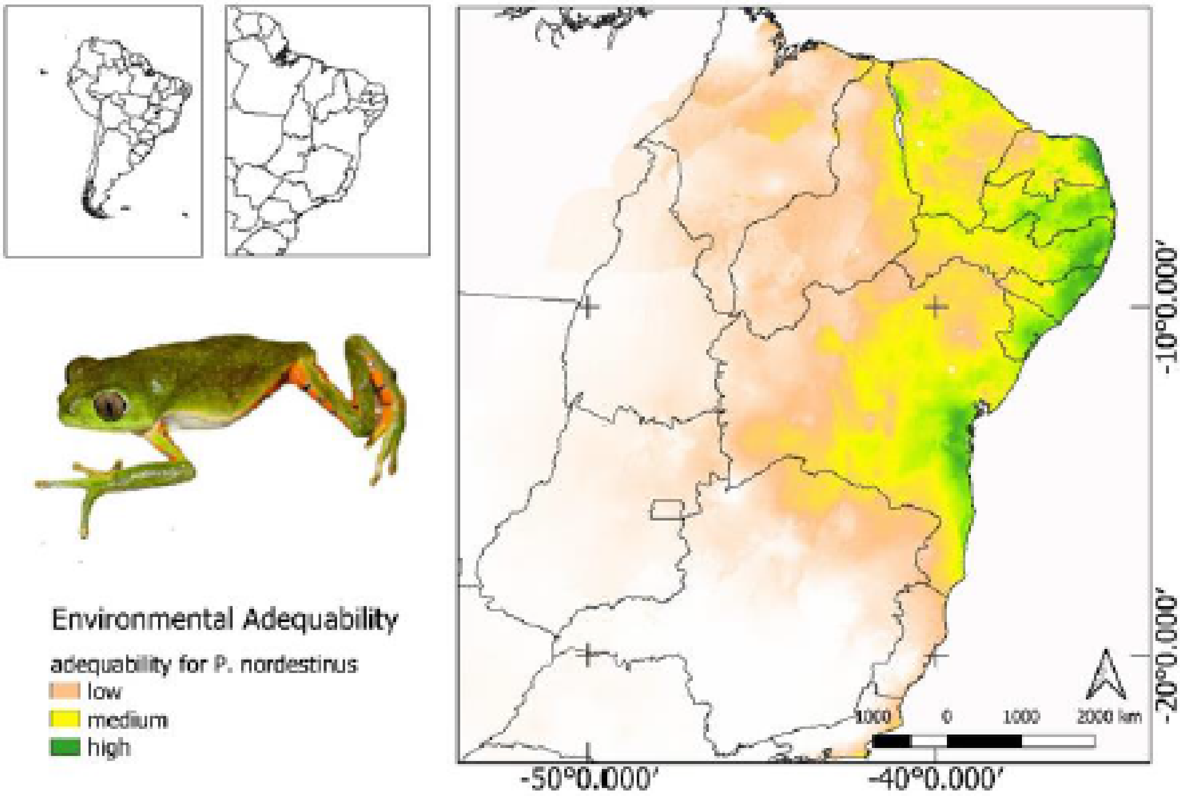


Figure 8: Pithecopus nordestinus´ ecological niche model represented in geographic space as suitability to the environment. (Model created in the analysis and programming environment "R" with the package "KUENM" through the MaxEnt algorithm).

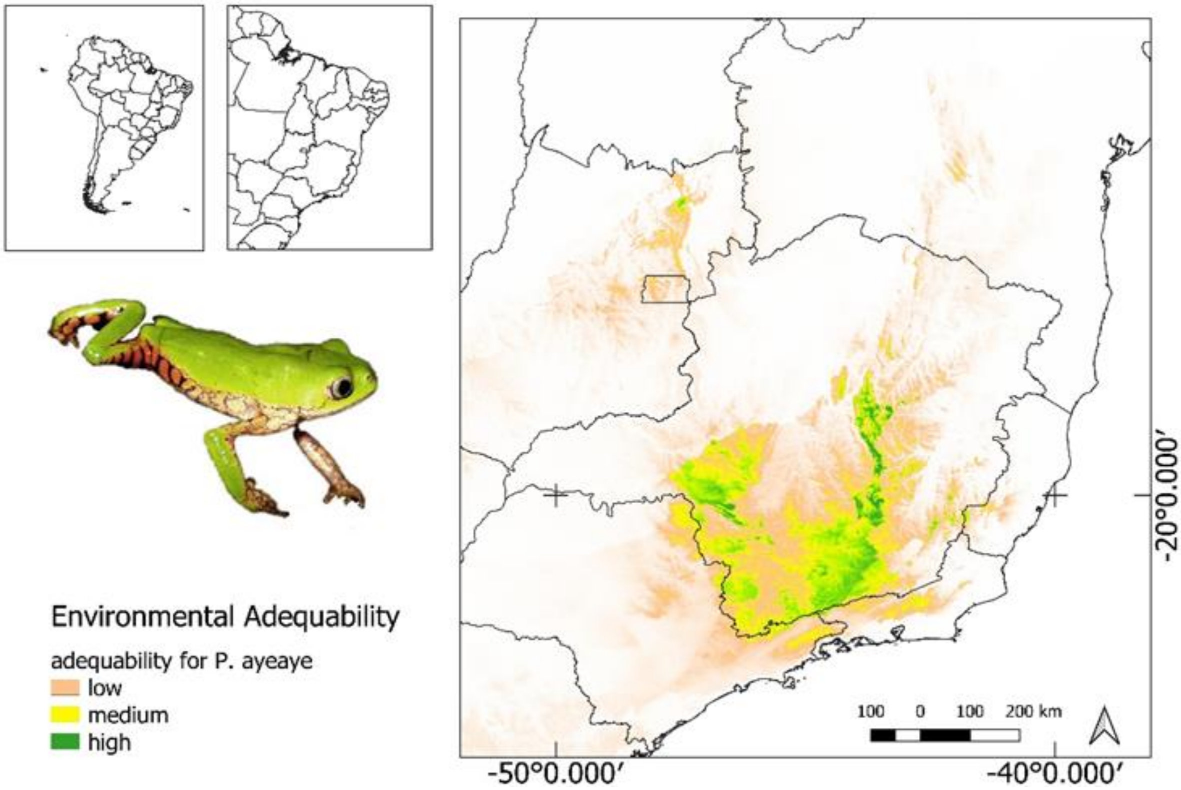


Figure 9: Pithecopus ayeaye´s ecological niche model represented in geographic space as suitability to the environment. (Model created in the analysis and programming environment "R" with the package "KUENM" through the MaxEnt algorithm).

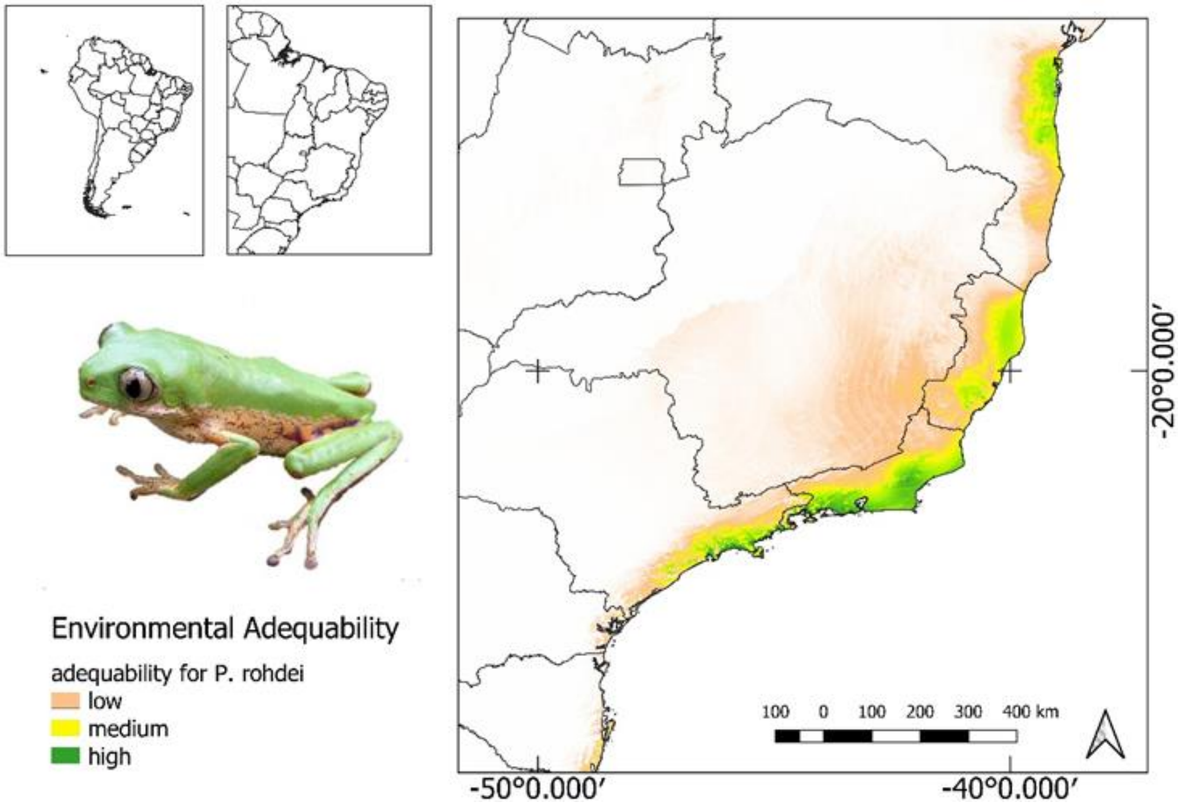


Figure 10: Pithecopus rohdei´s ecological niche model represented in geographic space as suitability to the environment. (Model created in the analysis and programming environment "R" with the package "KUENM" through the MaxEnt algorithm).

6 – Next steps

We concisely present the objectives pursued by this project in the proposal submitted by me to "THE RUFFORD FOUNDATION" as follows:

"This project aims to investigate whether the maintenance or modification of the ecological requirements (i.e. niche) over time had an influence on the diversification of the species belonging to the PIthecopus genus. In addition, we will seek to access the potential effects of climate change on these species. Key working hypotheses to be tested, in summary, include: (1) the ecological speciation mechanism is responsible for the distribution patterns of the species of the group; (2) both conservatism and the evolution of niche are processes responsible for the diversification of Pithecopus and, (3) climate change will lead to the displacement of areas of occurrence. Although some of these hypotheses have been tested at some biogeographic regions, including the Neotropical region, and for some groups of animals (principally birds and mammals), few of these studies include representatives of the group of anurans as the ones that will be considered here. ***For the test of the three conjectures highlighted above we will estimate the ecological niches of each species using modeling techniques****. Ecological Niche Models (ENM´s) now represent a widely used tool in Ecology, Evolution and Biogeography." Finally, we will use brand-new analysis in hypothesis testing. Although recently developed, such analyzes show a high power to resolve issues once they integrate information from several areas of knowledge (e.g. Ecology and Genetics)."*

As highlighted in the fragment above, the base and, in my view, the vital step for the present project is characterised by obtaining the ecological niche models (ENMs) and development of the ecological niche estimates. As shown here in this report, this step was successfully completed to date leaving no doubt as to the effectiveness of the methodology and experimental design chosen by the group for the development of the research. I believe now I can begin the development of the next steps of my PhD. These steps consist in the application of the ecological information obtained (specifically "the niche potential") in specific analyses aimed at solving the conjectures defined for my doctoral project. Specifically, with the ecological data obtained with the modeling, I will perform phylogenetic niche conservatism and niche differentiation tests, thus accessing the potential role of the evolution of the niches in the diversification of the target genus. In addition, we will also use such information in the creation of a lability index that aims to correct predictions of models for the future. Such a correction, theoretically, will enable the obtaining of models more in keeping with reality since it considers the evolutionary history of ecological characters (such as the niche) of the species. As a result of this more refined predictability, access to the impacts of climate change can be more realistic, which automatically leads to more effective planning and action for biodiversity conservation.