Use of Ecological Niche Models to inform fieldwork for the recognition of undescribed mammalian diversity in the Biogeographic Choco

Jhan C. Carrillo-Restrepo, Yehimy X. Rueda-Isaza, Jorge I. Velásquez-Tibatá and Juan F. Díaz-Nieto.

J. C. Carrillo-Restrepo (jcarril4@eafit.edu.co), Y. X. Rueda-Isaza and J. F. Díaz-Nieto, Dept of Biological Sciences and BEC Research group, EAFIT University, Cr 49 No.7sur-50, Bloque 38, Código Postal 050022, Medellín, Colombia. – J. I Velásquez-Tibatá, NASCA Conservation program, The Nature Conservancy, Cll 67 # 7-94, Bogotá, Colombia.

Abstract.

The Neotropics, with about 1701 species of mammals (i.e. 30% of the planet's species), is recognized as one of the regions with the greatest biodiversity. Colombia has 518 species of mammals, corresponding to 10% of the global diversity. Despite these high numbers of richness, recent studies showed that at least 2000 species, mainly distributed in the Neotropical region, are still waiting to be described. Therefore, it is crucial to validate a methodology that optimizes the process of discovering and describing these species in an efficient way. Ecological niche modeling has been proposed as a tool to identify areas that harbor potential undescribed diversity, however, such idea have been vaguely stated in the theory and have hardly been proven in the field. The aim of present work was to build models of ecological niches for 7 species of terrestrial small cis-Andean mammals species with the purpose of projecting them in trans-Andean areas of the Biogeographic Choco. These models identified 4 environmentally suitable zones that prioritize fieldwork for the discovery of undescribed diversity. Parallel to this study, fieldwork was carried out in one of such areas at the northern slopes of Cordillera Central (Anorí municipality), where three individuals of arboreal rats of the genus *Nyctomys* were collected—so far such group was restricted to North America. In addition, based on the high degree of allopatric distribution of our *Nyctomys* populations with respect to those of North America, phylogenetic and species delimitation analyses were conducted to evaluate the species status of these new populations. Our results strongly support the recognition of Anorí population as a new species, but we defer a formal

taxonomic description for future work that includes examining type material of the nominal taxa associated with the genus *Nyctomys*.

Introduction.

One of the most relevant and longstanding questions in the biological sciences is related to the number of species inhabiting our planet. There are several proposals on this topic; some authors have estimated global diversity around 2 million species (Costello et al. 2012), other studies estimate approximately 8.7 million of eukaryotic species (Mora et al. 2011), while other—more extreme—proposals estimate the diversity of the planet in around 100 million species (Erwin 1991). Despite the wide range of values between estimates, which are related to the fact that they all use different taxonomic groups to extrapolate to all other living clades, all these studies agree in the fact that we are still far from recognizing the Earth’s real biodiversity. For instance, Mora et al. (2011) identified that 1.2 million species have been described through taxonomic classification, meaning that 86% of the species present on the planet are waiting to be recognized and described.

Due to a complex interaction between topographic, environmental, and biotic elements, the Neotropics are classified as one of the most important regions in terms of species richness (Myers et al. 2000). In particular for mammals, it has been estimated that in the last 60 years most of the new species described for this group have been concentrated mainly in the Neotropics (Jenkins et al. 2013). In fact, 1701 species of mammals are currently known for this region, representing ca.30% of all species of this group on the planet (Reeder et al. 2007). Colombia, has a privileged location in the Neotropical region, with approximately 528 species of mammals (i.e., 10% of the mammal species of the world: *sensu* Ramírez-Chavez et al. 2016), and two of the most important hotspots of biodiversity: The Northern Andes and the Tumbes-Chocó-Magdalena—herein called Biogeographic Choco (Conservación Internacional 2011)—, which together hold 50.2% of all the mammal species of the tropics in an area that merely corresponds to 1.2% of the planet's surface (Conservación Internacional 2011).

Despite such elevated species richness, several studies have shown that we are still far from recognizing the real diversity of mammals of the world. Reeder et al. (2007) estimated that at least 2000 species (mostly small mammals such as rodents, bats and marsupials), mainly distributed in the neotropical region, are still waiting to be described. It has been more than 250 years since Linneaus started the task of cataloguing the biodiversity of the planet (Linnaeaus 1758) and we still have a long path ahead; it is therefore vital to develop a proposal that could help to expedite such task, if we are to achieve our goal in the near future. Although many of the recently described species have been mainly based using material collected years ago and deposited in museum drawers (Reeder et al. 2007, Funk et al. 2011, Díaz-Nieto et al. 2016, Teta and Díaz-Nieto 2019), some descriptions of new species have been based on recently field collected material (Velazco and Patterson 2014, Velazco and Patterson 2019) and consequently innovative ways that could help us inform field localities to find such species are needed.

The Ecological Niche Modeling (ENM) is an algorithmic methodology that uses associations between environmental variables and localities of occurrence of the species to define the biotic and abiotic conditions where populations can survive (Phillips et al. 2017, Soberón and Nakamura 2009). A particular aspect of the ENM that could be relevant to our purposes in trying to predict areas with undescribed diversity is the over-prediction, which can be understood as a prediction of potentially suitable areas (Gi) according to the biotic (B) and abiotic (A) conditions but which historically have not been accessible (M) for the species (Soberón and Nakamura 2009). In the past, Raxworthy et al. (2003) used an over-prediction of ENM to study the diversity of endemic chameleons to Madagascar and in the process they found 7 new species for science. They conclude that "over-prediction of niche modeling not only provides powerful predictions of species distributions, but also offers an innovative and predictive approach to discover hitherto unknown populations of known or unknown species" (Raxworthy et al. 2003). Consequently, and following Raxworthy et al.'s (2003) rationale, the aim of this work is to use the ENM to inform areas in the Biogeographic Choco with high potential to find undescribed diversity.

Methods.

We constructed Ecological Niche Models (ENMs) of terrestrial small cis-Andean—lowland humid forest east of the Andes—mammal species to identify areas in the Biogeographic Choco with high potential of harboring undescribed biodiversity. To accomplish this, we first trained the models on the native study region for each species (see predictor variables) and then projected the model onto the Biogeographic Choco where we finally overlapped the models to identify areas with high mammal species richness, which in turn could be interpreted as a proxy for areas with undescribed species (Raxworthy et al. 2003).

1. Species selection and occurrence data

We conducted an exhaustive literature review in Patton et al. (2015) and Gardner et al. (2008) in order to identify species of rodents and marsupials that matched the following criteria: (1) Small (< 200 gr) terrestrial or scansorial species with an exclusive cis-Andean distribution, (2) species widely distributed in the Amazon basin or with narrower distributions west of Madeira and Negro rivers, (3) species without any close relatives (i.e., species within the same genus) with known distribution in the Biogeographic Choco, and (4) species included as part of taxonomic/systematics reviews to identify and exclude species complexes. To avoid the incorporation of sampling biases or taxonomic identification errors—which are frequent in records of biological databases such as the Global Biodiversity Information Facility (GBIF) or the Vertebrate Distributed Database Networks (VertNet) (Beck et al. 2014)—we restricted the records used as occurrence data to those exclusively obtained from revisionary work for each species.

2. Predictor variables

We gathered climatic data at 30-arc second resolution (~ 1 km at the equator) from the WorldClim 2.0 database (Fick and Hijmans 2017), specifically the set of 19 bioclimatic variables based on temperature and precipitation, and the mean and coefficient of variation of annual solar radiation. Since it has been shown that rivers function as partial biogeographic barriers for the dispersal of terrestrial vertebrates (including mammals) in the lowlands of the

Amazon (Wallace 1852, Oliveira et al. 2017), we used the polygons delimited by the WWF HydroBASINS level 4 (Lehner and Grill 2013) as an appropriate way to define the study region for the calibration of the model of each species (M). Additionally, we used the polygon of Tumbes-Chocó-Magdalena (Conservación Internacional 2011) to delimit the projection area in the Biogeographic Choco.

3. Ecological Niche Modeling

3.1 Data preparation

For both rodent and marsupial species, we made similar decisions regarding the preparation of data for modeling. To reduce the potential effects of sampling bias and spatial autocorrelation of occurrences, which can inflate the suitability estimates for spatially grouped localities (Veloz 2009; Boria et al. 2014), we employed the R package spThin (Aiello-Lammens et al. 2015) to spatially thin the occurrence records at least 2 km among them (i.e. so that there is no record closer to 2 km than any other record). This distance was chosen according to the ecology of the species, the environmental heterogeneity in the Amazon basin and the spatial resolution of the environmental variables (~ 1 km). Similarly, it has been shown that most ecological niche modeling methods are sensitive to the number of variables used (Braunish et al. 2013, Dorman 2007), and the use of correlated prediction variables can lead to excessive adjustment and, therefore, bias the results (Warren and Seifert 2011). Consequently, to reduce this potential bias, we calculated the variance inflation factors (VIF) (Marquardt 1970), a method that specifically eliminates those variables with VIF values above 10. Finally, for the development of the niche models, 10,000 random points were taken from the M area of each species and used as background data.

3.2 Ecological niche modeling of small terrestrial cis-Andean mammals.

We developed ENMs using the software MaxEnt version 3.4.1 (Phillips et al. 2017) implemented from the statistical environment R version 3.5.1 (R Core Team 2018) using the packages dismo version 1.0-12 (Hijmans et al. 2017) and ENMeval version 0.3.0 (Muscarella et al. 2014). Within MaxEnt, variable selection is controlled by a penalty known as regularization multipliers (RM) and the level of complexity of the response to each environmental variable is controlled by Feature Classes (FCs) (Phillips and Dudík 2008, Merow et al. 2013). To select optimum values of RMs and FCs, we used package ENMEval (Muscarella et al. 2014), and ran MaxEnt models considering feature classes Linear (L), Quadratic (Q), and Linear and Quadratic (LQ) and regularization multiplier values from 0.5 to 4 with a step value of 0.5. We chose to exclude the feature classes "Hinge", "Product" and "Threshold" of the analysis, as their omission results in models that are simpler and easier to interpret (Phillips et al. 2017). ENMeval allows several data-partitioning schemes, for species that had less than 30 occurrences, we used the "jackknife" method following Muscarella et al. (2014). For species with more than 30 occurrences, we used the "block" partition method which consists in 4-fold spatial partitions delineated by the longitude and latitude lines located to balance the number of points in each fold (Muscarella et al. 2014). The “block” method is adequate in our study since it evaluates how well models can be transferred to areas not included in the training data (Muscarella et al. 2014).

The best model for each species was selected using the Akaike Information Criterion corrected for small sample sizes (AICc) (Warren and Seifert, 2011; Warren et al. 2014). As AICc can only select the "best" model from among a set of models generated by ENMeval and does not directly evaluate the performance of the model, we explored the lowest 10 percentile omission rate (OR10) and the area under the receiving operating characteristic curve (AUC) of the models selected as optimal. We constructed the final model with the combination of the regularization multiplier and the feature classes that produced the lowest delta value of AICc, the lowest value of the OR10 and best values of AUC, and using all records (Table 1). The resulting best model for each species were projected onto the Biogeographic Choco area to predict suitable habitats for each species. In projecting models, clamping was allowed (whereas extrapolation was not), so that environmental variables would be kept within those ranges found in the original training area. In addition, as we projected these models onto a region outside their training area, we inspected the Multivariate Environmental Similarity Surfaces (MESS) (Elith et al. 2010) produced by Maxent to determine which environmental variables in the Biogeographic Choco region were outside the range of climatic conditions present in the training area for each species.

3.3 Determination of suitable areas.

For each of the final models of our species, we used the log-log format (cloglog), since this format allows us to intuitively interpret the output of the model as a 0 to 1 habitat suitability index (Phillips et al. 2017). The potential distribution maps of each species were converted into binary maps of "suitable" vs. "unsuitable" habitat using the threshold “10 percent training presence”, since this threshold selects the value above which 10th percentile of the training locations are correctly classified, and it is one of the most common thresholds used in the MaxEnt habitat fitness models (e.g. Anderson et al. 2003). Finally, we added up the binary maps for each species to identify areas with the highest potential for undescribed species (Ferrier and Guisan 2006). Additionally, we added the continuous models for each species following Calabrese et al. (2014) recommendation to derive species richness estimates from ENM.

4. Modelling Testing

4.1 Field Work

Parallel to this study, an independent biotic inventory took place in a locality that corresponded to one of the areas where the models of most of the species overlapped, producing localities for suitable fieldwork to find undescribed diversity (see below, area “d” in results). In particular, fieldwork was developed in the Humid Rain Forest of the northern slopes of Cordillera Central, at a locality known as Anorí (Antioquia, Colombia). The biotic inventory included several taxonomic groups such as: amphibians, birds, insects (emphasis in butterflies, beetles of the family Curculionidae, and true bugs of the families Membracidae, and Aetalionidae), plants (with emphasis in orchids, palms and Cyclanthaceae), reptiles, and mammals. For terrestrial small mammals all the details on the sampling effort and specimens collected are detailed in the results section. However, we collected three specimens of the monotypic arboreal rat genus *Nyctomys*, that until now had a known distribution restricted to North America (Hunt et al., 2004); consequently, this record represents an important extension on its distributional range. Given the elevated degree of allopatric distribution between the north and south American populations, and in the context of our Niche Model results, we conducted a phylogenetic and species delimitation analysis to evaluate the species-level status of this new population (see next section for more details).

4.2 Species delimitation using molecular data.

4.2.1 Taxon and gene sampling.

Methods for DNA extraction followed Giarla et al. (2010). We obtained sequences of two mitochondrial DNA markers, the complete (1140 bp) Cytochrome-b gene (CYTB) and a fragment (657 bp) of the Cytochrome-C Oxidase Subunit I (COI). Both makers were obtained by PCR and primers for CYTB and COI loci were obtained from Percequillo et al. (2011), and Ivanova et al. (2007), respectively. The resulting PCR fragment was Sanger sequenced in MCLAB laboratories, California, USA. For the phylogenetic analyses, our ingroup was comprised by the recently obtained sequences of *Nyctomys* in addition to 10 sequences of CYTB and 6 sequences for COI that were downloaded from the GenBank for this genus (Fig. 1, Table 2). In order to corroborate the genus level identification of our samples, we also downloaded sequences (for each marker) from at least one species of all other genera (*Otonyctomys, Ototylomys,* and *Tylomys*) within the cricetid subfamily Tylomyinae (Musser et al. 2005). Our designated outgroup was comprised by CYTB and COI sequences of *Rhipidomys leucodactylus* (subfamily Sigmodontidae). For assessing sequence quality, COI and CYTB nucleotides were translated to amino acid sequences using Geneious ver R.11 to evaluate that no premature stop codons were present. Subsequently, both CYTB and COI sequences were aligned using default parameters in MUSCLE ver. 7.017 (Edgar 2004) as implemented in Geneious ver R.11.

4.2.2 Phylogenetic analyses.

We performed independent phylogenetic analyses of aligned CYTB and COI sequences using maximum likelihood (ML). Additionally, we performed Bayesian Inference (BI) for the aligned CYTB sequences in order to implement a species delimitation analysis (see below). We partitioned each mitochondrial gene alignment by codon position and estimated the best partition scheme and best-fitted nucleotide substitution model using BIC implemented in PartitionFinder v. 2.0 (Lanfear et al. 2012). We performed four independent ML searches in GARLI 2.0 (Zwickl 2006) and evaluated nodal support based on 1000 bootstrap pseudorreplicate searches also implemented in GARLI (Felsenstein 1981, Felsenstein 1985). For each mitochondrial loci, Bootstrap support values (BS) were summarized in the best ML tree using SUMTREES v. 3.3.1 (Sukumaran et al. 2010). We were not able to concatenate CYTB and COI markers because our taxon sampling was highly uneven.

Our species delimitation analysis requires an ultramateric topology to be implemented and for that purpose we conducted a Bayesian Inference (BI) analysis in BEAST v. 2.5.2 (Bouckaert et al. 2019) using our CYTB alignment. We restricted our analyses to the latter mentioned marker because it contained the largest taxon sampling (compared to COI) and contained haplotypes that could be associated to nominal taxa using literature (Corley et al. 2011). For our BEAST analysis we partitioned CYTB alignment by codon position, used a lognormal relaxed-clock model and implemented a Yule model as tree prior. We performed two independent Markov Chain Monte Carlo (MCMC) analyses for 10 million generations each, sampling every 1000 generations to produce a final dataset of 10000 trees. To ensure convergence, the results of the MCMC runs were inspected in TRACER v. 1.7.1 (Rambaut et al. 2007) and discarded the first 10% of trees of each run as burn-in. The remaining trees were used to summarize tree topology, mean log-likelihood, nodal support (posterior probability, PP), and remaining parameters into a maximum-clade-credibility tree with TreeAnnotator v. 2.5.1 (Bouckaert et al. 2019).

To identify putative species, we used the General Mixed Yule Coalescent Model (GMYC) as implemented in the R package “Splits” (Pons et al. 2006). GMYC method delimits putative species by estimating the point of transition between intra and interspecific evolutionary processes on an ultrametric tree and, unlike other methods, it does not require prior assignment of sequences to taxa or populations (Pons et al. 2006). For the purposes of this study, we recognize putative species as strongly supported mtDNA clades ($PP \geq 0.95$) that fall within the species-delimitation threshold for GMYC. All phylogenetic analyses were implemented in the CIPRES Science Gateway (Miller et al. 2010). We estimated net uncorrected (p-) genetic distances within and between putative species using MEGA X (Kumar et al. 2018).

Results.

Niche models.

After our literature review, seven species, *Euryoryzomys macconnelli* (62 occurrences), *Hylaeamys perenensis* (27), *Hylaeamys yunganus* (88), *Marmosops bishopi* (24), *Mesomys hispidus* (47), *Makalata macrura* (18), and *Scolomys melanops* (11) were selected for the development of ENMs, using a total of 285 occurrence records. The spatial thinning of the occurrences did not reduce our initial data set in any meaningful way, so that we finally obtained a total of 273 occurrence records to be used in the modeling. We calibrated a total of 168 possible models ([8 values of regularization multiplier x 3 combinations of classes of classes] x 7 species), of which the best model for each species was selected. Optimal model settings differed considerably between the small terrestrial cis-Andean mammal species, spanning all combinations of FCs and the full range of RMs. Table 1 shows the different values of the evaluation metrics (see methods) for the best model of each species. The simplest model was for *S. melanops* (LQ1.5 with 1 parameter), and the most complex model was for *H. yunganu*s (LQ1 with 12 parameters). The best AICc value was (216.7602) of the LQ1.5 combination. The average OR10 was highest for *M. hispidus* (0.3318182), and the lowest for *M. bishopi* (0.173913), but was 0.24069 on average for all species. AUC test was highest for *S. melanops* (0.8814), and the lowest for *M. hispidus* (0.6346), but was on average 0.76852 for all species. For *E. macconnelli*, *H. yunganus* and *M. hispidus*, which are the species with the broadest distribution in the Amazon basin, the variable with the highest contribution was bio13 (Precipitation of the wettest month). On the other hand, bio14 (Precipitation of the driest month) was the variable that contributed most to the models of *M. bishopi* and *S. melanops*, two species occurring west of the Madeira and Negro rivers. According to the MESS analysis of MaxEnt, for all species, variables bio19 (Precipitation of the coldest quarter) and bio14 (Precipitation of the driest month) are the variables most dissimilar relative to the calibration area (ie, those that are farther from their training range).

After using the "10 percent training presence" threshold to convert the best model of each species into binary maps of "suitable" vs. "unsuitable" habitat, we overlapped these 7 models in the Biogeographic Choco to find areas with high mammalian richness, which can be interpreted in our case as a proxy of areas with potential undescribed biodiversity (Fig. 2).

Our models predicted a total area of 13370 Km^2 with high richness (areas where 7 and 6 potential distributions of the species overlap, Fig. 3) and these can be group into 4 major areas (see a–d in Fig. 2): a) Southwest of Colombia's Biogeographical Choco, in localities of Mosquera, Olaya Herrera, Roberto Payan, San Francisco and Tumaco; b) the Andean piedmont of the Cordillera Occidental between the Departamentos of Valle del Cauca and Cauca; c) an area north of the biogeographic Choco, in the mid-Atrato river, at municipalities of Vigia del Fuerte and Murindó; and d) a small area in the northern slopes of the Cordillera Central at the locality of Anorí.

Fieldwork.

Small mammals were sampled during 8 nights implementing an effort of 2202 traps/night, including Sherman, Victor, Tomahawk and pitfall traps. This effort resulted in 43 individuals captured, which corresponds to 2.0% trapping success. The captured individuals were identified as 7 species, grouped into 7 genera, 4 families and 2 orders. This diversity is represented in groups such as American marsupials (Didelphidae), small rodents native to the Neotropics (Cricetidae), basket mice (Heteromyidae) and tropical spiny rats (Echimyidae). Particularly, we collected three specimens of the genus *Nyctomys*, which to date, is recognized as a monotypic genus restricted to North America, ranging from Mexico to Central Panama (Hunt et al., 2004). The specimens of the genus *Nyctomys* collected at Anorí (Fig. 2, locality d), correspond to an extension in the range of distribution of the genus of more than 500 km (Fig. 1), it is also the first record for South America and consequently the first record for Colombia.

Phylogenetic analyses and species delimitation.

In our final CYTB matrix, our ingroup contained 2 new sequences from *Nyctomys* (locality d), 10 Genbank sequences of *Nyctomys* corresponding to 8 localities in North America (Fig. 1), 4 *Ototylomys*, 1 *Otonyctomys*, and 4 *Tylomys* sequences; in addition 2 *Rhipidomys leucodactylus* sequences were included as our designated outgroup. CYTB sequences ranged in length from 247 to 1140 bp, representing 78.3% nucleotide coverage overall. For COI, our final matrix contained 3 *Nyctomys* sequences from Anorí, 6 *Nyctomys sumichrasti* sequences from 5 localities of North America (Fig. 1), 4 *Ototylomys*, 1 *Otonyctomys*, and 4 *Tylomys* sequences as ingroup, plus 2 outgroup sequences of *Rhipidomys leucodactylus*. Our COI sequences ranged in length between 597–657 bp, which represents 91.3% nucleotide coverage. The best-fitted partitioning scheme and associated nucleotide substitution model found by PartitionFinder is represented in Table 3.

Maximum Likelihood analyses for CYTB and COI sequences of newly collected material of *Nyctomys* along with available GenBank sequences of this genus confirm the identification of the populations collected in the field as part of the genus *Nyctomys* and they also recovered the genus *Nyctomys* as monophyletic although with low support (See Appendix Fig. S1 and Fig. S2). Our GMYC species delimitation analysis recovered a significant ($p=0.09$) transition from a Yule (inter-specific) to a coalescent (intra-specific) branching process, and a total of 6 lineages cross the GMYC species threshold (dashed line in Fig. 4). The 6 mentioned lineages would correspond to putative species (clades A–F in Fig. 4), however our "intraspecific" taxon sampling is reduced, with 4 "putative species" represented by single haplotypes (clades C, D, E, F), while clades A and B are haplogroups with 2 and 6 haplotypes, respectively. Within the latter mentioned mitochondrial haplogroups, only the clade for Anorí (clade A) is strongly supported. The genetic distances obtained varied from 13.8% (Clade A respect to Clade B) to 16.5%, (Clade A respect to Clase E) (Table 4).

Discussion.

Our work is one of the few studies using ENM to inform areas with high potential to find undescribed biological diversity (see Raxworthy et al., 2003), however it is the first study addressing such question with that particular methodology for mammals. Our work also documents the conservative nature of ecological niche evolution, since through the implementation of modeling of ecological niches of closely related small terrestrial cis-Andean mammal species, we were able to identify an area where fieldwork evidenced a possible new species.

Regarding the developed ENMs, the best models were those that met our criteria for assesing statistical significance, model complexity and performance. However, we did not find agreement between AICc, AUC, or OR10 criteria. Although the AICc metric has been widely adopted in the field of ENMs as a criterion (Shcheglovitova and Anderson 2013; Warren et al. 2013; Muscarella et al. 2014), we caution that it should be used with care since we show that selection of models with better values of AIC can result in models that do not perform well according to the data used. In particular, while the AICc predicted high significance for models, according to the AUC, significantly better than random predictions were recorded for all species except *E. macconnelli*, *M. hispidus*, *H. yunganus* and *H. perenensis*, species widely distributed in the Amazon basin for which the yield was low. This could be attributed to the fact that the performance of the model was affected not only by the magnitude of the calibration area M as different authors suggest (Barve et al. 2011, Velasco and González-Salazar 2019) but also due to the availability high-quality data, since for these species there were very few points that could not be correctly differentiated from the Background data. However, despite the statistical uncertainty of these models of widely distributed species, their biological explanation is relevant since they are predicting the same areas of models with great significance as those of the species *S. melanops*, *M. bishopi* and *M. macrura*.

According to our species delimitation analysis (Fig. 4) the genus *Nyctomys* seems to have much more diversity than the single species currently recognized and the new population at Anorí recorded by our fieldwork represents an apparent new species. Previous authors (Corley et al. 2011) performed sequence analysis of the mitochondrial CYTB gene and showed a possible paraphilia for the genus *Nyctomys*; however, they also presented evidence for an elevated and unusual molecular diversity within the species, which could represent the presence of multiple species not described under the name "*Nyctomys sumichrastii*". Our species delimitation analysis seems to provide additional support for that hypothesis, in particular, 6 clades were recognized as putative species. Our estimates of genetic distances of the examined sequences increase the support of higher diversity within *Nyctomys*. While genetic distances alone are not sufficient to argue species boundaries, there is a generalized consensus (Baker and Bradley 2006) that net mitochondrial distances between clades exceeding 10% must represent interspecific variation (between species) and not intraspecific values (between species). Our data shows that CYTB distance between our putative species from Anorí and all the other 5 recognized putative species ranges between 13.8% and 16.5%, which are values that clearly exceed the average known intraspecific variation and supports the hypothesis that the individuals of the locality of Anorí is a new species not yet described.

Although our data strongly supports the recognition of Anorí population as a new species, we defer a formal taxonomic description for future work that includes examining type material of the nominal taxa associated with the genus *Nyctomys*; nevertheless, the present work is the first step towards the clarification of the systematics of the genus *Nyctomys.*

Acknowledgements.

We are thankful to E. A. Noguera-Urbano for his contributions and comments to initial stages of this work. We want to thank J. L. Parra-Vergara for his comments and reviews at various stages of the project. UNDP, COLCIENCIAS, and Universidad EAFIT funded fieldwork through the Expedicion BIO Anorí 2018.

References.

Aiello-Lammens, M. E. et al. 2015. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. – Ecography 38: 541–545.

Anderson, R. P. et al. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. – Ecol. Modell. 162: 211–232.

Baker, R. J. and Bradley, R. D. 2006. Speciation in mammals and the genetic species concept. – J. Mammalogy 87:643–662.

Barve, N. et al. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. – Ecol Modell. 222:1810–1819.

Beck, J. et al. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. – Ecological informatics 19: 10-15.

Boria, R. A. et al. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. – Ecol. Modell. 275: 73–77.

Bouckaert, R. et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. – PLoS computational biology 15(4).

Bradley, R. D. et al. 2004. Phylogenetic relationships of Neotomine-Peromyscine rodents: based on DNA sequences from the mitochondrial cytochrome b gene. – J. Mammalogy 85:389-395.

Braunish, V. et al. 2013. Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. – Ecography 36: 971–983.

Calabrese, J. M. et al. 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models. – Glob. Ecol. Biogeogr. 23: 99–112.

Conservation Internacional. 2011. Biodiversity Hotspots Revisited.

Costello, C. et al. 2012. Status and solutions for the world's unassessed fisheries. – Science 338: 517–520.

Corley, M. S. et al. 2011. Molecular evidence for paraphyly in *Nyctomys sumichrasti*: Support for a new genus of vesper mice? –Occasional papers Museum of Texas Tech University, 306:1–10.

Díaz-Nieto, J. F. et al. 2016. DNA sequencing reveals unexpected Recent diversity and an ancient dichotomy in the American marsupial genus Marmosops (Didelphidae: Thylamyini). – Zool. J. Linn. Soc**.** 176: 914–940.

Dorman, C. 2007. Promising the future? Global change projections of species distributions. – Basic. Appl. Ecol. 8: 387–397.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. – Nucleic Acids Res. 32: 1792–1797.

Elith, J. 2010. The art of modelling range-shifting species. – Methods Ecol. Evol. 1: 330–342.

Erwin, T. L. 1991. How many species are there?: Revisited. – Conservation Biology 5: 330–333.

Felsenstein, J. Evolutionary trees from DNA sequences: A maximum likelihood approach. – J. Mol. Evol. 17: 368–376.

Felsenstein, J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. – Evolution 39: 783-791.

Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – J. Appl. Ecol. 43: 393–404.

Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. – Int. J. Climatol. 4315: 4302–4315.

Funk, W. C. et al. 2011. High levels of cryptic species diversity uncovered in amazonian frogs. – Proceedings of The Royal Society B: Biological Sciences 279: 1806-1814.

Gardner, A. L. 2008. Mammals of South America, Volume 1: Marsupials, Xenarthrans, Shrews, and Bats. – University of Chicago Press.

Giarla, T. C. et al. 2010. Species limits and phylogenetic relationships in the didelphid marsupial genus Thylamys based on mitochondrial DNA sequences and morphology. – Bulletin of the American Museum of Natural History 346: 1–67.

Hijmans, R. J. et al. 2017. dismo: Species Distribution Modeling. R package version 1.1-4.

Hunt, J. L. et al. 2004. *Nyctomys sumichrasti*. –Mammalian species 754: 1–6.

Ivanova, N. V. et al. 2007. Universal primer cocktails for fish DNA barcoding. – Molecular Ecology Notes 7: 544–548

Jenkins, C. N. et al. 2013. Global patterns of terrestrial vertebrate diversity and conservation. – Proc. Natl. Acad. Sci. U. S. A. 110: 2602-2610.

Kumar, S. et al. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547–1549.

Lanfear, R. et al. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. – Molecular Biology and Evolution 29: 1695–1701.

Linnaeus, C. 1758. Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. – Editio decima, reformata. Holmiae: Laurentius Salvius 1-824.

Lehner, B. and Grill, G. 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. – Hydrol. Process. 27: 2171–2186.

Marquardt, D. 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. – Technometrics 12: 561–612.

Merow, C. et al. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. – Ecography 36: 1058–1069.

Miller, M. A. et al. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proccedings of the Gateway Computing Environments Workshop (GCE), pp 1-8.

Mora, C. et al. 2011. How many species are there on Earth and in the Ocean?. – PLoS Biology 9: e1001127.

Muscarella, R. et al. 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for ecological niche models. – Methods Ecol. Evol. 5: 1198–1205.

Musser, G. M., and Carleton, M. D. 2005. Superfamily Muroidea. – In: Wilson, D. E. and Reeder, D. M. (eds.), Mammal Species of the World: A Taxonomic and Geographic Reference. Third edition. pp. 894–1531.

Myers, N. et al. 2000. Biodiversity Hotspots for conservation priorities. – Nature 403: 853–858.

Oliveira, U. et al. 2017. Biogeography of Amazon birds: rivers limit species composition, but not areas of endemism. – Scientific Reports 7: 2992.

Patton, J. L. et al. 2015. Mammals of South America, Volume 2: Rodents – University of Chicago Press.

Percequillo, A. R. et al. 2011. A new genus and species of rodent from the brazilian atlantic forest (Rodentia: Cricetidae: Sigmodontinae: Oryzomyini), with comments on Oryzomyine biogeography. – Zool. J. Linn. Soc. 161: 357-390.

Phillips, S. J. and Dudík, M. 2008. Modeling of species distributions with maxent: new extensions and a comprehensive evaluation. – Ecography 31: 161–175.

Phillips, S. J. et al. 2017. Opening the black box: an open-source release of maxent. – Ecography 40: 887–893.

Pons, J. et al. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. – Syst. Biol. 55: 595– 609.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambaut, A. and Drummond, A. J. 2007. Tracer v. 1.5. Available at: http://beast.bio.ed.ac.uk/Tracer.

Ramírez-Chaves, H. et al. 2016. Cambios recientes a la lista de mamíferos de Colombia. – Mammalogy Notes 31:1–20.

Raxworthy, C. J. et al. 2003. Predicting distributions of known and unknown reptile species in Madagascar. – Nature 426: 837–841.

Reeder, D. M. et al. 2007. Global trends and biases in new mammal species discoveries. – Occasional Papers, Museum of Texas Tech University 269: 1–35.

Shcheglovitova, M. and Anderson, R. P. 2013. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. –Ecol Model 269:9–17.

Soberón, J. and Nakamura, M. 2009. Niches and distributional areas: Concepts, methods, and assumptions. – Proc. Natl. Acad. Sci. U. S. A. 106: 19644–19650.

Sukumaran, J. and Holder, M. T. 2010. DendroPy: a Python library for phylogenetic computing. – Bioinformatics 26: 1569–1571.

Teta, P. and Díaz-Nieto J. F. 2019. How integrative taxonomy can save a species from extinction: The supposedly extinct mouse opossum Cryptonanus ignites (Diaz, Flores and Barquez, 2000) is a synonym of the living C. chacoensis (Tate, 1931). – Mammalian Biology.

Velasco, J. A. and González-Salazar, C. 2019. Akaike information criterion should not be a “test” of geographical prediction accuracy in ecological niche modelling. – Ecological Informatics 51: 25–32.

Velazco, P. M. and Patterson, B. 2014. Two new species of yellow-shouldered bats, genus Sturnira Gray, 1842 (Chiroptera, Phyllostomidae) from Costa Rica, Panama and western Ecuador. – Zookeys 402: 43-66.

Velazco, P. M. and Patterson, B. 2019. Small mammals of the Mayo River Basin in northern Peru, with the description of a new species of Sturnira (Chiroptera, Phyllostomidae). – Bulletin of the American Museum of Natural History 429: 67pp.

Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. – J. Biogeogr. 36: 2290–2299.

Wallace, A. R. 1854. On The Monkeys Of The Amazon. – Annals And Magazine Of Natural History 84: 451–454.

Warren, D. L. and Seifert, S. 2011. Ecological niche modelling in Maxent: the importance of model complexity and the performance of model selection criteria. – Ecol. Appl. 21(2): 335–342.

Warren, D. L. et al. 2014. Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. – Divers. Distrib. 20: 334–343.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD thesis, University of Texas at Austin, Austin (TX).

FIGURES.

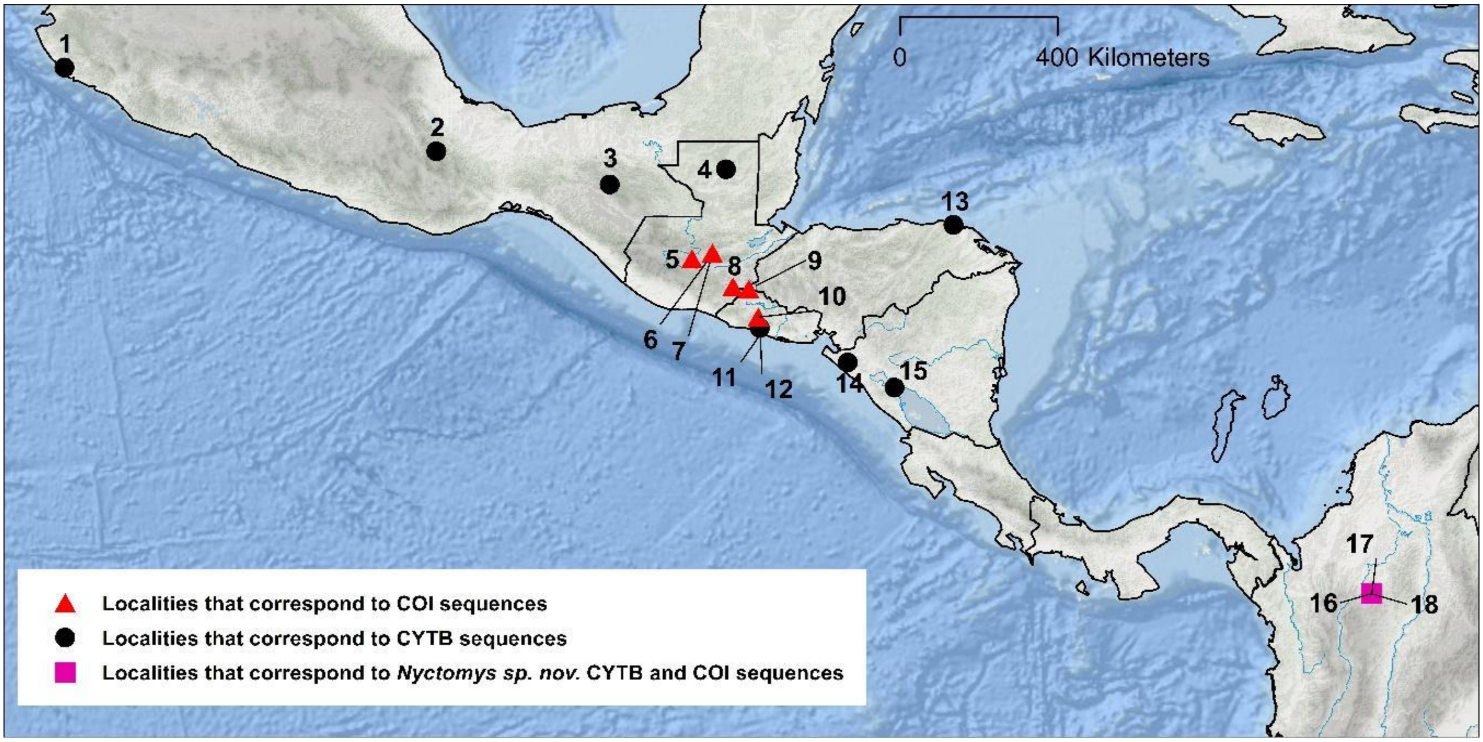

Figure 1. Collection localities for sequenced specimens of *Nyctomys*.

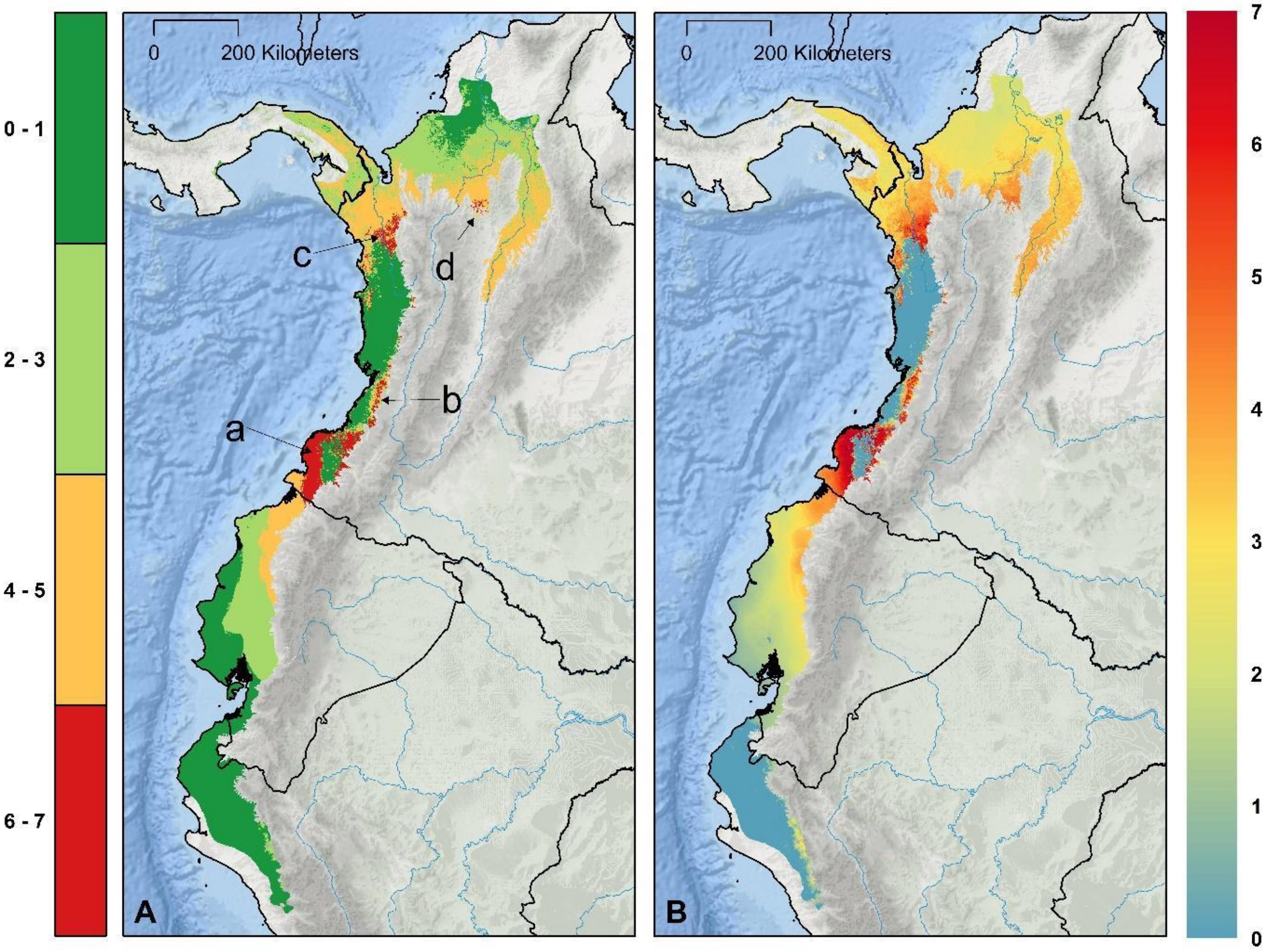

Figure 2. A) Predicted species richness in the Biogeographic Choco from stacked ENMs of 7 terrestrial small cis-Andean mammals. Areas with highest richness are shown in red and areas with lowest richness are shown in green. B) Predicted potential distribution of terrestrial small cis-Andean mammals species, areas with highest suitability are shown in red and areas with lowest suitability are shown in blue following Calabrese et al. (2014). Red areas on both maps could be interpreted as a proxy for areas with undescribed mammalian species.

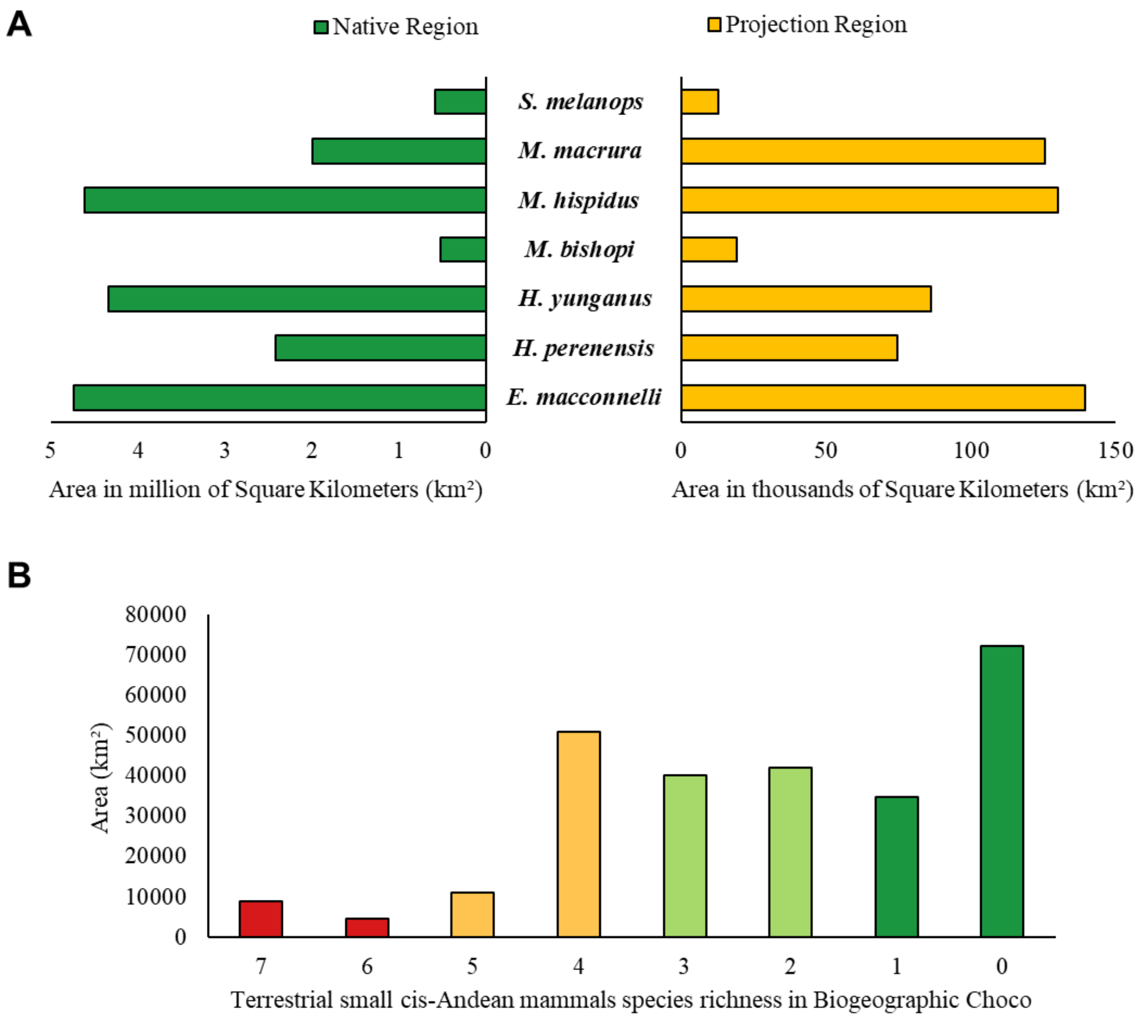

Figure 3. A) Appropriate area (km^2) estimated for each terrestrial small cis-Andean mammal species both in its native area (Amazon Basin) and in the projection area (Biogeographic Choco). B) Estimated total area for each region where models overlap in the Biogeographic Choco.

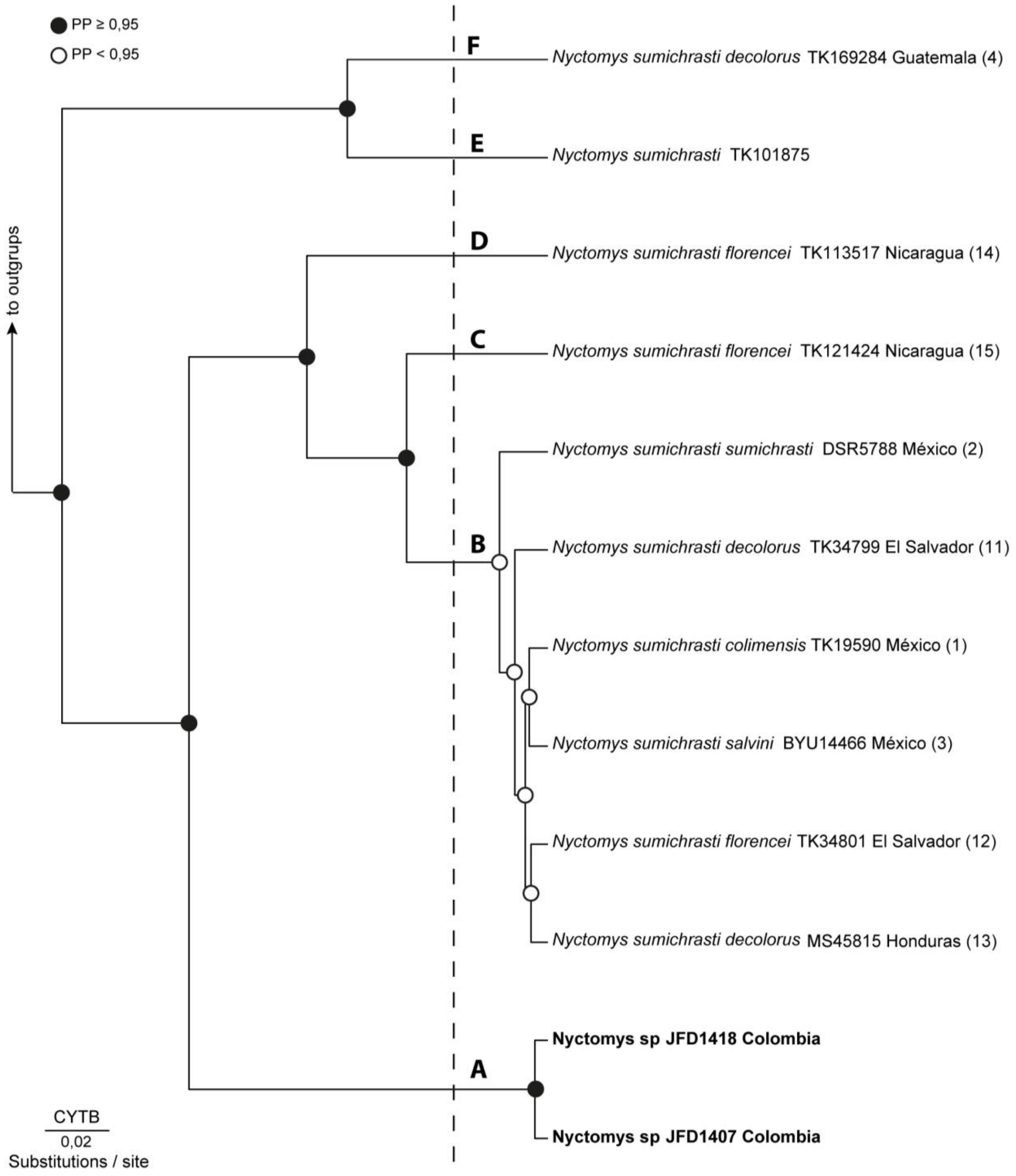

Figure 4. BEAST ultrametric tree based on CYTB sequences of *Nyctomys* sequences. Dashed vertical line shows the threshold between Yule and coalescent branching processes as estimated by implementation of the GMYC model. Filled circles at internal nodes represent high support values, posterior probability (PP) ≥ 0.95.

Table 1. Parameters and Evaluation metrics of the best model (for each of our 7 mammal species) used to evaluate its behavior within the training area (Amazon basin). For all of these models ΔAIC was zero.

Specie	Feature Classes	Regularization multiplier	train AUC	avg.diff.AUC	avg.test.or10pct	AICc	nparam	10^{th} Percentile training presence threshold
Euryoryzomys macconnelli	Q	1	0.6748	0.1696	0.28229	1953.879	8	0.3648
Hylaeamys perenensis	L	2	0.7368	0.1263	0.28	773.1143	3	0.4405
Hylaeamys yunganus	LQ	1	0.7711	0.1299	0.14464	2603.163	12	0.3453
Marmosops bishopi	L	2	0.8634	0.0655	0.1739	634.4879	4	0.37
Mesomys hispidus	L	4.5	0.6346	0.1685	0.33181	1350.664	3	0.5029
Makalata macrura	L	0.5	0.8176	0.1207	0.22222	550.0298	5	0.3954
Scolomys melanops	Q	1.5	0.8814	0.05611	0.25	216.7602	1	0.3288

Feature Classes: L–Linear, Q–Quadratic, LQ–Linear and Quadratic.

Table 2. Sequences of *Nyctomys*. In the Locality column, numbers in parentheses correspond to localities mapped in Figure 1.

Taxon	Gene Sequence	Locality	Specimen Voucher	GenBank Accession #	Source
Nyctomys sumichrasti colimensis	CYTB	México: Jalisco (1)	TK19590	JQ183066	Corley, M. S. et al. 2011
Nyctomys sumichrasti sumichrasti	CYTB	México: Oaxaca (2)	DSR5788	JQ183067	Corley, M. S. et al. 2011
Nyctomys sumichrasti salvini	CYTB	México: Chiapas (3)	BYU14466	JQ183065	Corley, M. S. et al. 2011
Nyctomys sumichrasti decolorus	CYTB	Guatemala: Petén (4)	TK 169284	JN851816	Corley, M. S. et al. 2011
Nyctomys sumichrasti decolorus	CYTB	Honduras: Gracias a Dios (5)	MSB:45815	KY754074	Steppan and Schenk, 2017
Nyctomys sumichrasti decolorus	CYTB	El Salvador: La Paz (6)	TK34799	JQ183063	Corley, M. S. et al. 2011
Nyctomys sumichrasti florencei	CYTB	El Salvador: La Paz (7)	TK34801	JQ183064	Corley, M. S. et al. 2011
Nyctomys sumichrasti florencei	CYTB	Nicaragua: Chinandega (8)	TK113517	JQ183061	Corley, M. S. et al. 2011
Nyctomys sumichrasti florencei	CYTB	Nicaragua: Atlántico Norte (9)	TK121424	JQ183062	Corley, M. S. et al. 2011
Nyctomys sumichrasti	CYTB		TK101875	AY195801	Bradley et al. 2004
Nyctomys sumichrasti	COI	Guatemala: Baja Verapaz (10)	ROM:98520	JF444684	
Nyctomys sumichrasti	COI	Guatemala: Baja Verapaz (11)	ROM:31505	JF444683	

Nyctomys sumichrasti	COI	Guatemala: Sacatepequez (12)	ROM:98385	JF444682
Nyctomys sumichrasti	COI	Guatemala: Zacapa (13)	ROM:99899	JF444680
Nyctomys sumichrasti	COI	El Salvador: San Salvador (14)	ROM:101222	JF444681
Nyctomys sumichrasti	COI	El Salvador: Santa Ana (15)	ROM:101531	JF444679

Table 3. Optimal partitioning scheme and substitution models for each gene data set.

Gen	Partition	Best Model	Characters
CYTB	1	SYM + Γ	CYTB position 1
	2	HKY + I	CYTB position 2
	3	GTR + I + Γ	CYTB position 3
COI	1	GTR + Γ	COI position 1
	2	F81 + I	COI position 2
	3	GTR	COI position 3

GTR, Generalized Time Reversible model; HKY, Hasegawa-Kishino-Yano model; F81, Felsenstein (1981); SYM, Symmetrical model; I, proportion of invariant sites; Γ, gamma-distributed rate variation.

Table 4. Net uncorrected (p-) genetic distances between putative species on the upper right diagonal and Net uncorrected (p-) distance within putative species of *Nyctomys* in diagonal.

	Clade A	Clade B	Clade C	Clade D	Clade E	Clade F
Clade A	**0.4**					
Clade B	13.8	**1.2**				
Clade C	15.8	8	-			
Clade D	16	11.1	12.8	-		
Clade E	16.5	16.1	16.1	16.7	-	
Clade F	16.2	15	16.5	16.6	9.4	-

Appendix or Supplementary material section

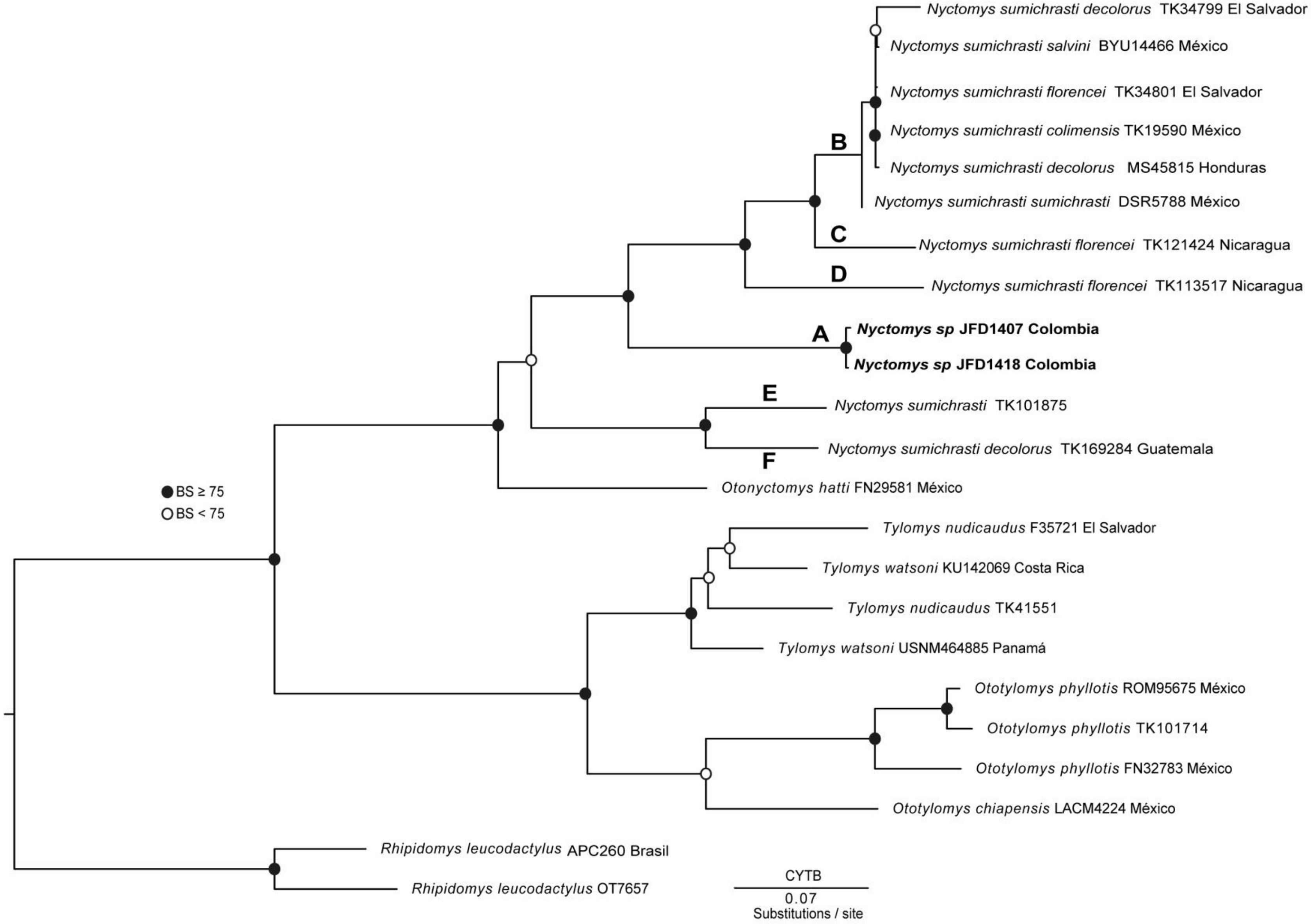

Fig. S1. Maximum-likelihood topology of CYTB sequences of *Nyctomys* species and all other species of genera within subfamily Tylominae. Filled circles at internal nodes represent high bootstrap support values. Bolded terminal corresponds to the haplogroup of the Anorí collected specimens.

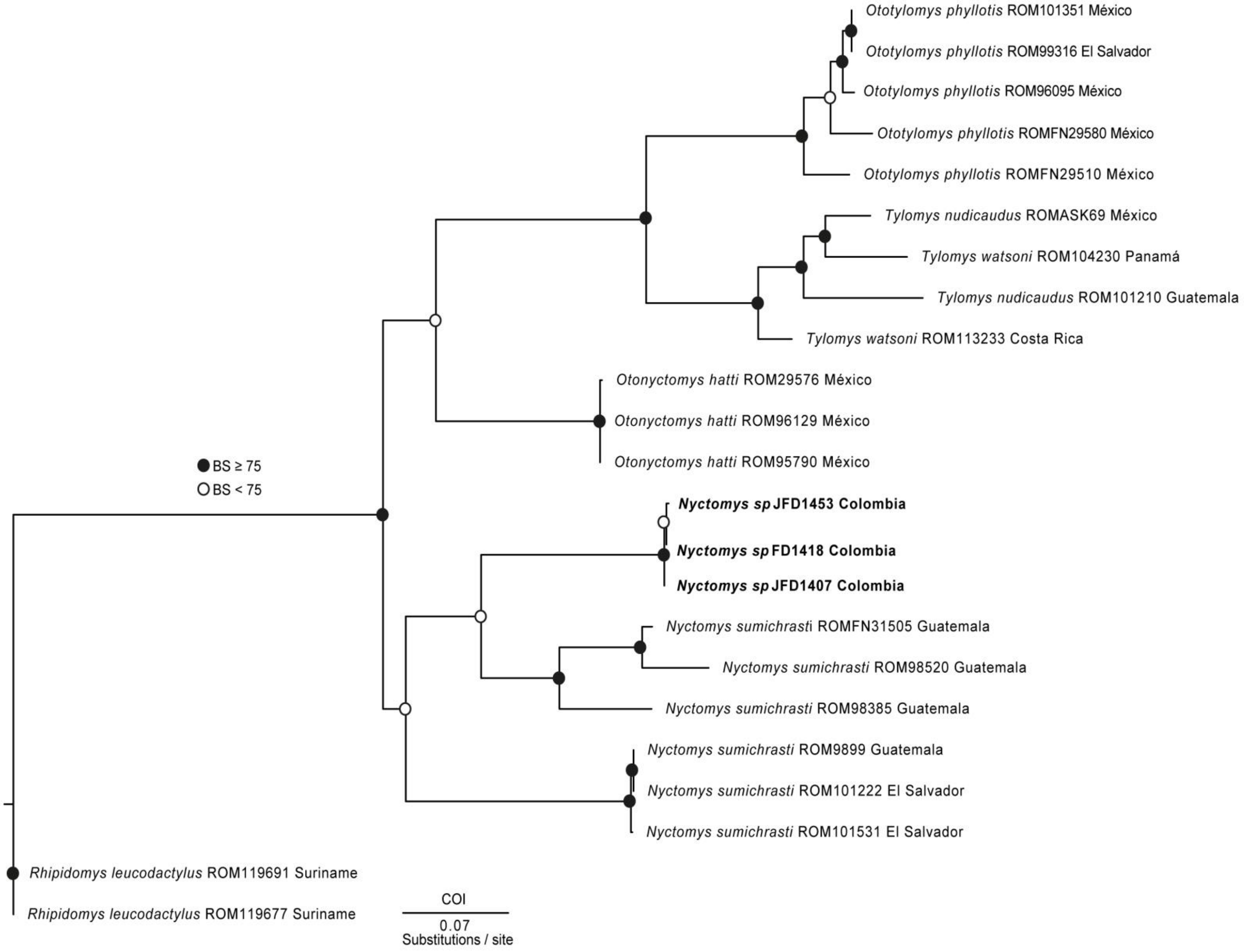

Fig. S2. Maximum-likelihood topology of COI sequences of *Nyctomys* species and all other species of genera within subfamily Tylominae. Filled circles at internal nodes represent high bootstrap support values. Bolded terminal corresponds to the haplogroup of the Anorí collected specimens.