Suitability prevalence area index in late quaternary explains genetic diversity in Tassel eared Squirrels

Norma Hernandez, Angel Robles Nathan Upham

2023-06-27

Abstract

As a case study we consider squirrels (*Sciurus aberti*), a species currently distributed in disjunct patches from the southern Rocky Mountains in the United States to the northern Sierra Madre Occidental in Mexico for which it is possible to find reliable information on both the genetic structure of its populations and its current distribution (Bono et al. 2018; Burgin et al. 2018).

Our results reveal that fitness prevalence corresponds to the fixation index of *S. aberti* populations with respect to a source population. Populations closer to the historical endemic area present a higher genetic diversity and a lower F_{st} value. Finally, this study allows us to add a biogeographic explanation to the results obtained with population genetic methods and to generate maps of this structure as tools to support conservation with a perspective that integrates both population genetics and historical patterns of species distribution.

Introduction

The current distributions of species do not always correspond to their historical distributions over evolutionarily significant periods. This is because environmental conditions are not static over time; species tend to distribute where conditions are most favorable.

It follows from this relationship between environment and species distribution that population size also varies with time. This variation in population size is related to the effective population size N_e (Karlin 1968), so that an index reflecting changes in environmental conditions in geography might be expected to be related to N_e and thus to indicators of population structure such as the fixation index F_{st}. Thus, it is possible to relate patterns of changes in the distribution of species to the genetic structure of their populations using a statistical model that explains this relationship.

With this approach we can predict the geographic pattern of population structure from environmental information. This approach is strongly driven by advances in currently available climate simulations (Leonardi et al. 2023; Krapp et al. 2021), as well as next generation sequencing data, and supported by both ecological niche (Thorup et al. 2021; Nogués-Bravo 2009) and population genetics theories (Lira-Noriega and Manthey 2014).

In this work we propose a method to find the Suitability Prevalence Area (SPA) as an index with a double purpose: 1) to find endemic areas to delimit the historical distribution of the species and 2) to explain the patterns of genetic diversity.

To obtain the SPA, we performed a historical reconstruction of the geographical range back to 120 000 BC at 2 000 year intervals and recorded the environmental suitability at each site.

Subsequently, we delimited historical endemic areas to locations where the prevalence of suitability remained at 90% during this period. From the fixation index calculated with respect to populations in the historical endemic areas, a statistical model of the fixation index as a function of SPA was performed. With this

statistical model, the fixation index values were projected to the current distribution to obtain a map with the geographic pattern of this index.

Theoretical basis

Environment, fitness and suitability   For a given region $\mathbf{G}$ in te geography, at any point in time, we have a set of existing environmental conditions $\mathbf{E}(\mathbf{G}, t) = \left\{ \vec{e}_i ; \mathbf{G}, t \right\}$. We define $\vec{e}_i = (e_1, e_2, ..., e_v)_i$ as a vector of environmental variables, that is, a point in the space of v environmental variables $\mathbf{E}(\mathbf{G}, t)$. Then $R_0(\vec{e}_i)$ is a function that relates fitness to environment, that is, for a species with non-overlapping generations, the population's net reproductive rate R_0 is a measure of fitness that depends on environmental conditions. If we take the set of places in environmental space where these values are greater than a threshold, we can define a volume in multidimensional space such that $R(\vec{e})$ is always positive and greater than this threshold, that is $\mathbf{N}_F = \{\vec{e} | R_0(\vec{e}) \geq k\}$. i. e. the set of environmental conditions for which values of fitness are capable of sustaining populations. We call $\mathbf{N}_F \subseteq \mathbf{E}(\mathbf{G}, t)$. This is a way of expressing the fundamental niche as a function of environmental conditions. So we can map each point of geography to ambient space. and, for each site in the geography there will correspond a site in the environmental space with different environmental conditions and different growth rates. However all sites within the region defined by $\mathbf{N}_F$ will have viable populations and positive rates depending on time because we assume that the environment is dynamic and each geographic point is static, so for each site in geographic space this growth rate depends on the environment and therefore is different at each time: $R_0 = R_0(\vec{e}(t))\}$. Thus locally the growth rate will be related to the local population size as long as it is within the boundary defined by $\mathbf{N}_F$.

This equation was proposed by Soberon and Peterson Soberón and Peterson (2020) to relate demographics to a function of fitness and environment. In that study they tested the validity with monthly variables and empirical values on a short time scale. relating life history parameters and the environmental suitability of niche models.

This set can be seen as a convex volume in environmental space and we relate this growth rate to an environmental suitability function bounded between 0 and 1, for example, for a unimodal multivariate pdf, it would be equivalent to the distance from the maximum of the function. That is, the fitness-related environmental suitability can be any function such that it is 0 if $R_0(\vec{e}_i) > 1$ and 1 when R_0 is evaluated at the expected value of $\vec{e}(t)$ for all values of the set $\mathbf{N}_F$ where $R_0(\vec{e}_i) > 1$. In a practical way, instead of $\mathbf{N}_F$, we approximated $\mathbf{N}_F^*(\mathbf{G}, \mathbf{t}) = \mathbf{E} \cap \mathbf{N}_F$ (the existing niche) from known presence points in the geography $\mathbf{G}$ obtained from emprical observations stored in database, and map to $\mathbf{E}$ in order to numerically fitting a convex volume of the environmental space that enclose the known information.

The objective of our approach is not to know the function of R_0 but rather to relate this local variation and the function $\mathbf{N}_F$ with a function $S_i(\vec{e}_i)$ of environmental suitability that can be estimated from known methods of applied mathematics and machine learning (for example the case of Maxent, Phillips, Anderson, and Schapire (2006)).

Population size locally is a function of environmental suitability, where if suitability is 0 the population is 0 and if suitability is 1 the population has a maximum growth rate $R_i(\vec{e}_i)$

$$N_i \propto S_i(\vec{e}_i)$$

Once we have an approximation of this volume, we consider that each point in the geography is static and has a trajectory in the environmental space and for each trajectory the fitness function changes and therefore the suitability function changes but not in a homogeneous way, so each site presents different displacement rates in enviromental spaces and as a consequence, different demographic rates. So we can establish that for each site i in geographic space $\mathbf{G}$, there is a population size N_i that changes with respect to a fitness function dependent on time-varying environmental conditions:

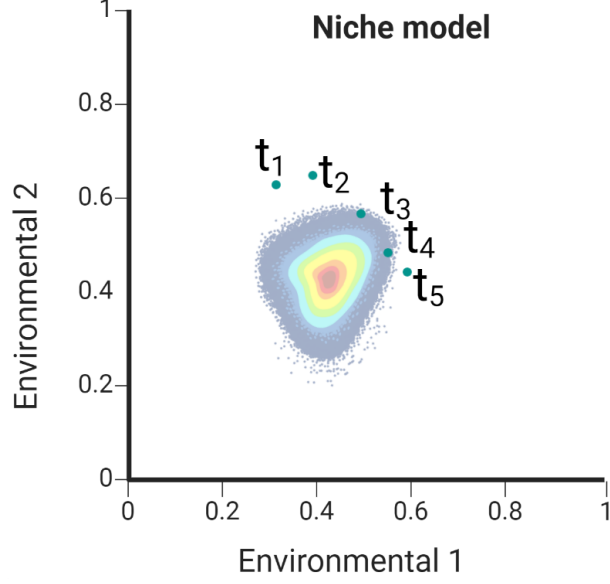$$N_i(\vec{e}_i(t)) \propto S_i(\vec{e}_i(t))$$

Figure 1: Hypothetical niche in environmental space. An arbitrary point moves within the ambient space entering and exiting the volume (niche) where the intrinsic growth rate is positive.

Genetic variation and effective population size On the other hand, loss of heterozygosity may be related to the amount of genetic variation present in the absence of mutation and selection, so theoretically, one would expect a strong correlation between effective population size and heterozygosity on the basis of population genetics. We start from the assumption that in the absence of gene flow between subpopulations, the rate of fixation by isolation depends on the effective population size N_e when subpopulations diverge from a common ancestor in t generations. Wright (1943):

$$F_{st}(t) = 1 - e^{-kt/N_e}$$

where we assume, according to Nei (Nei 1986), that genetic drift within each subpopulation causes the average heterozygosity between populations (H_s) to approach zero ($F_{st} = 1 - H_s/H_t$), i.e., for $t = 0$ $F_{st} = 0$ and for very large t the index is 1. that is, for $t = 0$ $F_{st} = 0$ and for very large t the index is 1. We have included a constant k without loss of generalization.

Using $e(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + ...$ and leaving to first order for N_e large enough

$$F_{st} \approx \frac{kt}{N_e} + C$$

Moreover, the effective population can be estimated as the harmonic mean of the population in a demographic time series (Karlin 1968) for τ generations $\frac{1}{N_e} \approx \frac{1}{\tau}\left(\frac{1}{N_1} + \frac{1}{N_2} + ... + \frac{1}{N_\tau}\right)$. We substitute the harmonic mean into the above equation, considering that although the number of generations is not exactly equal to the time of the fixation index, they are expected to be very close.

$$F_{st} \approx \left(kt\right)\frac{1}{\tau}\sum_{j=1}^{\tau}\frac{1}{N_j}$$

Relationship between the fixation index and environmental suitability We can infer locally for each population N_i in the geography expect to be proportional to the environmental suitability for each time

3

considered, For simplification in the index we select a N_i population in the geography and infer that locally, this population is proportional to the environmental suitability in each time considered, i. e. $N_j \propto S_j(\vec{e}_j)$ (each index j is for the generation time), as we describe above, so we can asume an harmonic mean of suitabilities $\frac{1}{\tau} \sum_{j=1}^{\tau} \frac{1}{S_j} = \frac{1}{S}$ and substitue in the previous equation:

$$F_{st} \approx \left(kt\right) \frac{1}{\tau} \sum_{j=1}^{\tau} \frac{1}{S_j(\vec{e}_j)}$$

We call this harmonic mean of environmental suitability *Suitability prevalence index* (S) given between 0 and 1, since it reflects the rate of population growth as a function of environment over a period of time. Thus we have a way to check that the loss of heterozygosity (through the fixation index F_{st}) is inversely proportional to the fitness prevalence for an i site in the geography:

$$F_i^{st} = \frac{\beta_0}{S_i} + \beta_1$$

If we consider all sites in the available geography, we call this harmonic mean of fitness *Suitability prevalence Area* $\mathbf{S} = \{S_i | S_i > k\}$ since we expect that areas where the F_{st} is lower are in areas where the prevalence over time is higher, i.e. environmental conditions remain suitable for a significant period of time and therefore no effects of genetic drift of populations have occurred. Conversely, it is expected that in populations with a high fixation index value, fitness will not have prevailed over the same time interval even though conditions currently exist and populations are present.

Methods

We summarize our 5 steps predictive approach in the figure 2.
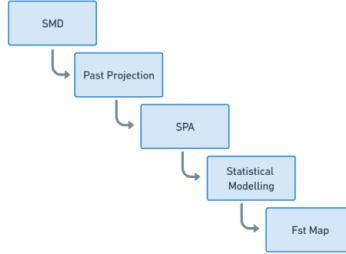

Figure 2: Realized niche in environmental space. A point dynamically moves along different time slices

We propose a method to find the Suitability Prevalence Area (SPA) as an index with a dual purpose: 1) to find endemic areas to delimit the historical distribution of species and 2) to explain patterns of genetic diversity.

The method for estimating SPA consists of a series of 4 steps:

1. Model the potential distribution of a taxon with present-day environmental conditions and obtain an environmental suitability map.
2. Project the potential distribution model to various environmental scenarios in the past with constant time steps and obtain the environmental suitability for each time scenario.
3. Characterize past environments and retain projections only for sites with analogous environmental conditions to the present.
4. Calculate the harmonic mean of the suitability for all the sites all time slices.

Subsequently, one application of the SPA is to explain genetic diversity (the fixation index) as a function of the SPA and project it to the geography. For this it is necessary to perform a series of steps, which consist of:

- Obtain georeferenced points of genetically structured populations with fixation index values.
- Extract values from the SPA map with these points.
- A linear model of the fixation index value with respect to the SPA is performed.
- With the linear model all SPA values of suitability are interpolated.
- A map is generated with all the values interpolated by the model.

To test our model we use the Aberti squirrel (*Sciurus aberti*) as a case study because:

- Since it is a medium-sized mammal (0.5 kg) its identification is easy.
- The subspecies are geographically isolated, since they are found in pine forests surrounded by desert. Therefore we can consider the assumption of isolation by distance (Wright 1943).
- There are sufficient occurrence records for the entire range of the known distribution.
- It has low speciation rate (0.25) (Upham, Esselstyn, and Jetz 2019), therefore, we assume that its fundamental niche is stable over time.
- There is fixation index information (Bono et al. 2018) for different populations of the subspecies, which is useful for using SPA as a way to explain genetic differentiation between populations.

Unique occurrence records

We have downloaded geographic information on Sciurus aberti from the open access platform Global Biodiversity Information Facility (GBIF). Subsequently, we eliminated duplicate records, those without precise coordinates and those with uncorroborated collection information. Once the unique occurrence records were obtained, we filtered the geographic space in order to avoid overpredictions in the spatial distribution models caused by the agglomeration of occurrence points. The filtering consisted of superimposing the records in the geographic space, then we made 5 latitudinal slices of 5 degrees of separation between each one, from latitude 20 degrees south to latitude 45 degrees north, taking as criteria the known distribution of the species and the ecoregions of North America, level 1 (downloaded from the US Environmental Protection Agency, 2010). Once the geographic space was segmented, we left the same number of records for each band (30 records per band) obtaining a total of 150 unique records of presence.

Environmental characterization

The environmental information to generate the models, from the present and the reconstruction to the past, was obtained from Pastclim 1.2, an R statistical software package designed to download and manipulate paleoclimatic datasets. In our particular case we chose the Beyer at al. Beyer, Krapp, and Manica (2020) set as it has environmental information available up to 120 000 BE, significantly evolutionary time, at 2000 year time intervals based on Global Circulation Models HadCM3 (Singarayer and Valdes 2010) and HadAM3H (Valdes et al. 2017). The environmental dataset has 17 bias-corrected bioclimatic variables, reduced spatial scale and spatial resolution of 0.5°square cells.

Accessibility region (M).

Species distribution models according to the BAM diagram (Soberon and Peterson 2005) states that the geographic range occupied by a species ($\mathbf{G}_o$) is the region of appropriate assemblages in terms of abiotic conditions. The region $\mathbf{M}$ represents in geography areas in which the species has access due to its colonization capabilities and the structure of geographic barriers within a specific period of time (Soberón and Nakamura 2009).

In this study we selected a set of North American ecoregions for modeling to delimit our Accessibility region. We took into account ecoregions that overlap with the known distribution of the species and also took into account neighboring ecoregions to increase the potential accessibility in projection scenarios to the past. Thus, regions 5, 6, 9, 10, 12, 13 of the level I ecoregions (https://www.epa.gov/) were selected.

Niche models

Current distribution   Estimating niche boundaries for species occurrence is called ecological niche modeling (ENM), and when the emphasis is on geographic distribution, it is known as species distribution modeling (SDM) (Guisan and Thuiller 2005; Peterson and Soberón 2012; Saupe et al. 2012). We model the current distribution of S. aberti using the MAXENT 3.4.4 algorithm (Phillips, Anderson, and Schapire 2006), which is a correlative model based on the maximum entropy principle used to estimate species distributions. The output of Maxent is a relative occurrence rate (ROR) interpreted as a probability of habitat suitability given observed environmental conditions by presence-only points (Merow, Smith, and Silander Jr 2013).

The parameterization was performed with the default values of the program with the exception of the "extrapolate" and "clamping" options to avoid artificial extrapolations in the extreme values of the climatic variables used in the models. To calibrate the models we used 70% of the records and the remaining 30% to validate them.

In order to give statistical certainty and to consider uncertainty, as well as to reduce overfitting and to have a suitable model to extrapolate to past scenarios, 10 replications were performed for each model using a cross-validation with the training data. In the end, the average model of these replications was considered as the result.

We evaluated the statistical significance of the resulting models for the present using the partial ROC test (Peterson, Papeş, and Soberón 2008), which is a modification of the ROC (receiver Operation Characteristic) test. The results of this test are proportions (ratios) of the area under the curve of the model with respect to a null model, product of repetitions that allow to statistically evaluate the areas under the curve (AUC) in relation to that expected by chance (Peterson, Papeş, and Soberón 2008), where a value derived by chance would be 0 and an acceptable value, according to the proportion of minimum omission errors tolerated in the model, would be greater than 1. A different set of test data was used on the average model of the replicates and in this way the reduction of overfitting was guaranteed.

Reconstruction to the past   The average model obtained from Maxent was extrapolated to environmental conditions into the past. Environmental layers were taken from pastclim (Leonardi et al. 2023) in 2000 year time slices to cover a total of 120 000 years. For each scenario into the past, 10 replicates were cross-validated and averaged.

Analogy of environmental conditions   To be certain that areas in the past are climatically analogous to the present we implemented the mobility-orient parity (MOP) test. This method generates a value from 0 to 1 according to the environmental analogy of past conditions to the present to eliminate geographic areas in the projections to the past outside of climates analogous to the present (Owens et al. 2013).

Suitability Prevalence Area (SPA)

Finally, to generate the SPA, both the average projection of the present and all the average projections of the past are considered. The output files are read in raster format.

A harmonic mean is performed as follows: For each layer the inverse of each cell value is obtained. Then the inverse values are summed by cells and divided by the number of scenarios. Finally, the inverse of each cell is obtained. It is important to consider that the Maxent outputs may contain very small values close to zero in many cells, for this reason we added a small amount of 0.001 to all the values in all the maps so that when obtaining the harmonic mean we could have an inverse value for values close to zero without altering the biological sense of environmental suitability.

Endemic historic Area

To calculate the historic endemic area, the SPA result is taken and a threshold is applied to generate a binary map according to the **S** equation. This threshold is the proportion of time that pixel has had ideal conditions. For our study, we consider the value of 90% since for a period of 120 000 years it is a coverage of 100 000 years, which is a significantly evolutionary time to fix genetic patterns at the population level.

Fixation index as a functino of SPA

To test our hypothesis of genetic diversity as a function of environmental suitability prevalence, from a set of genetically structured populations, a matrix of fixation indices between populations is considered. For this work we considerer the data from @Bono et al. (2018). It consideres 10 populations along the known distribution and provides values of F_st between this populations in matrix form. The matrix was transformed into a table as pairs between populations with index values between each pair. The values of the population with itself, i.e. 0, were included. Subsequently, georeferences were obtained for each population considered in the fixation index matrix. From this geo-referencing the SPA value for each population was extracted. Finally, the SPA value was aligned in the table of fixation index pairs. In total the table has 4 columns: source population, target population, fixation index and SPA of the target population. This table was separated by source population groups to generate a set of linear models between SPA values and fixation indexes F_{st}. An analysis was performed for each source population with the hypothesis that the correlation between SPA and fixation index exists and is non-zero if a dispersal process has occurred from the historical endemic area of the taxon.

Fixation index projected in geography

The linear model that presents the highest value of R^2 is considered suitable to make a geographic projection of the fixation index and observe genetic diversity in geographic space. To generate this map, the SPA values are taken and a table of coordinates and SPA values is generated for each cell of the raster. The SPA values are then extrapolated to obtain fixation index values from the linear model obtained previously. This column is added to the coordinate table and a raster is generated with the projected F_{st} values. Finally, a cutoff of this raster is performed using a map of the distribution of the species at present with a cutoff threshold of 90% in binary format (where 1 =presence and 0 = absence), since the records of the presence localities were obtained from bases that might contain some errors in the data (Phillips, Anderson, and Schapire 2006).

Results

Current and past distribution

We show the species distribution model for *Sciurus aberti* along the time line of the projected area per time slice. We show 5 notable moments in time line. and 5 notable points

SPA and Endemic Historic Area

We get the suitability presence area (figure 2) delmited by a 0.5 and 0.9 threshold (Figure 4). The endemic region where the enviornmental conditions 90 % del tiempo is the Sierra Madre Occidental, in the limits of Sonora, Chihuahua, Sinaloa and Durango. We correspond this Area to the *Sciurus aberti barbieri* subspecies. Also is notable that at 0.5 threshold, other regions seems to have stable conditions. We observe the area close to Coronado National Forest as well as Gila National Forest and Mescalero Reservation as suitable areas along time in the United States and some Areas of the Volcanic Belt in Mexico (Puebla, north of Guerrero State and Jalisco.

Fst as a funciton of SPA and Fixation index projected in geography

For all the linear regressions performed (Table 1), the only one that had statistical significance is the one for the Sciurus aberti barberi population with an R-squared value of 0.794.
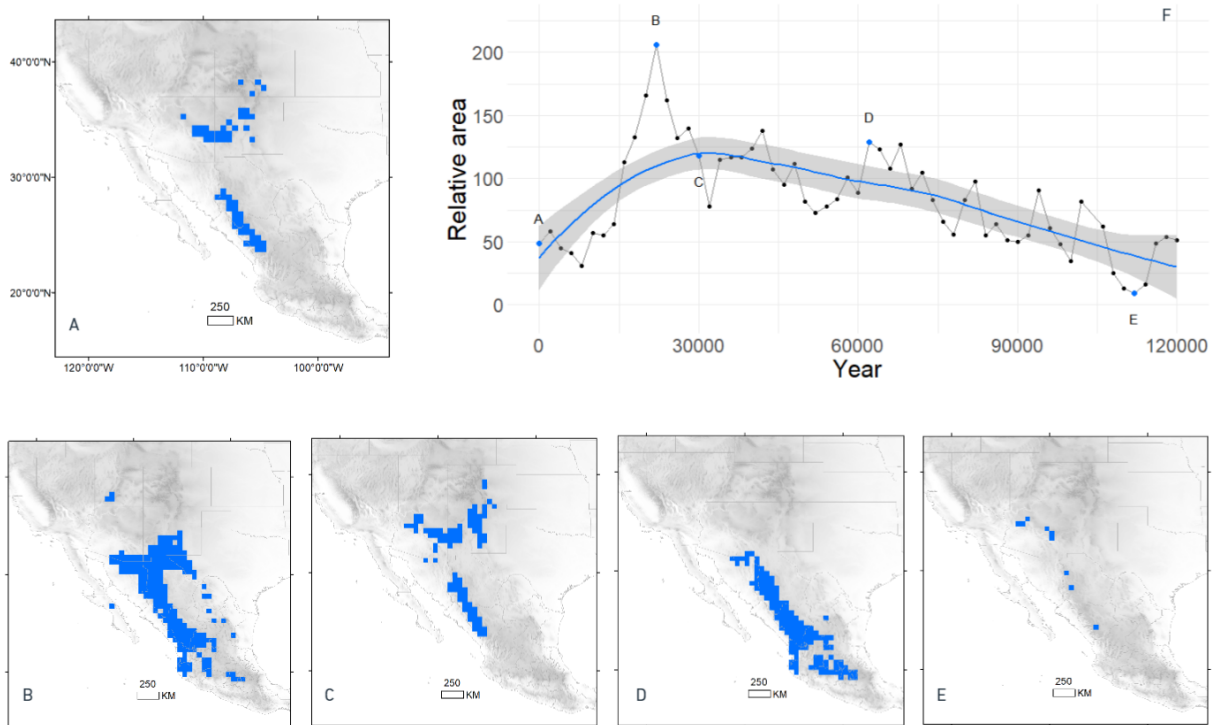
Figure 3: A) Species distribution model for *Sciurus aberti* present distribution. B) Relative projected area for SDM in pixels acoording a threshold of 0.9 suitability for each past time-sliced environmental conditions. We can observe a maximum area corresponding to the last interglacial era and an inverted U-shape in the trend of the area along the time considered. Also It can be noted that the actual distribution has similar area to the coditions 120 000 years ago. B) and D) corresponds to distribution peaks along the time scale and C) and E) corresponds to minimum
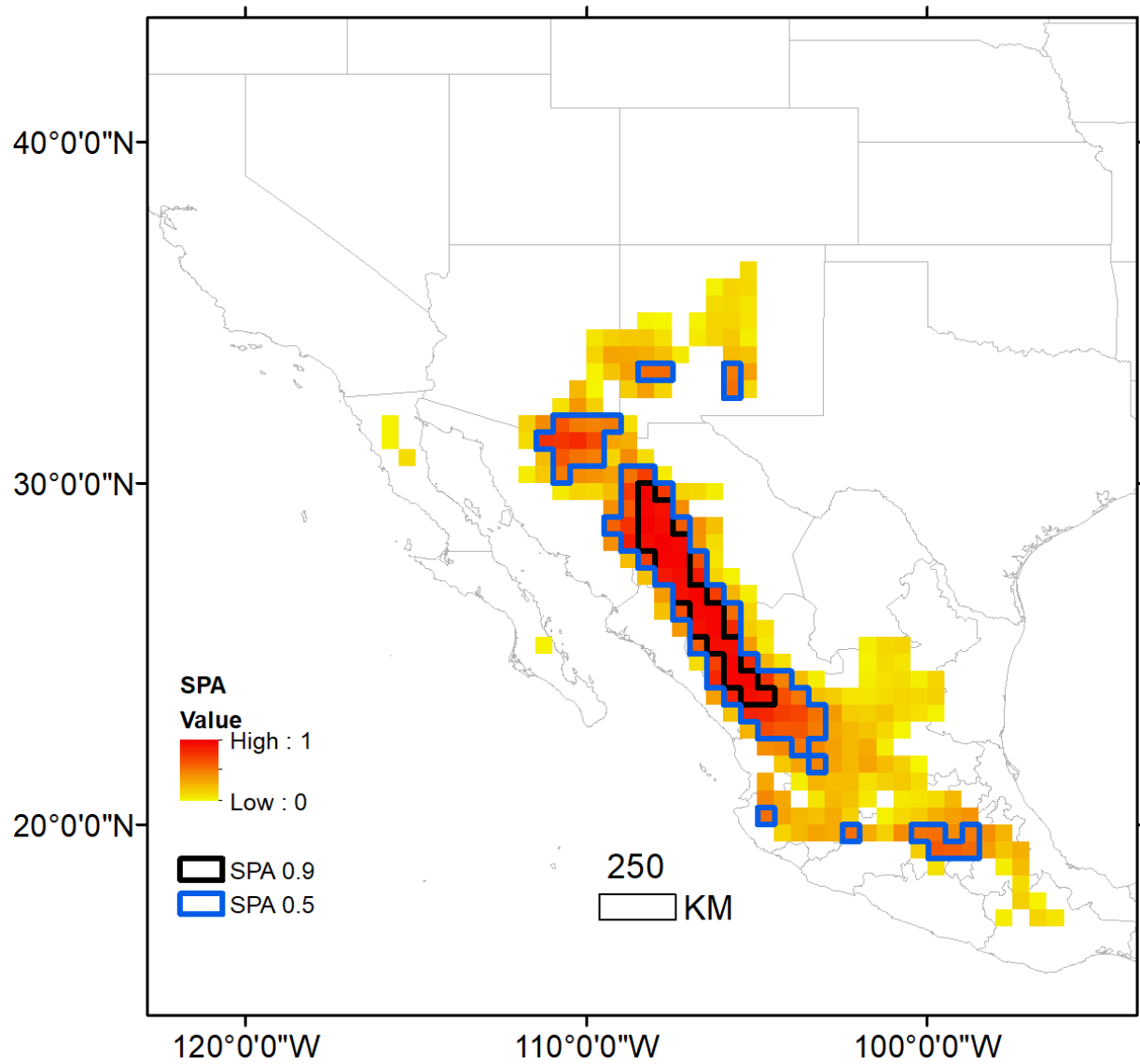
Figure 4: Suitability prevalence area. Here we observe the harmonic mean of suitability for each pixel along the distribucion from 0 to 1. The bold lines shows the the area when the harmonic mean es above 0.5 threshold (blud line) and 0.9 threshold (black line)

Table 1: R squared from linear models (fixation index as a function of suitability prevalence area). The only R squared with statistical significance is barberi

population	r.squared	p.value
S. a. aberti Carson-SFW	0.000	0.962
S. a. aberti Coconino-Gila	0.131	0.304
S. a. aberti MT-Zuni	0.002	0.894
S. a. aberti San Juan	0.000	0.952
S. a. barberi	0.794	0.001
S. a. chuscensis E	0.002	0.892
S. a. chuscensis W	0.018	0.713
S. a. ferreus Carson E	0.040	0.581
S. a. ferreus Pike	0.000	0.967
S. a. kaibabensis	0.005	0.841

It can be noted that S. a. aberti Coconino-Gila had an R-squared value greater than 0 (0.131) but it is not statistically significant. Furthermore, in our approach we imputed the value of the origin with an Fst of zero. To show that the model hypothesis still holds, we performed a regression for barberi without the point of origin imposed (Figure 5b), in addition to eliminating the S. aberti chuscensis populations due to their low uncertainty in georeference, obtaining an R-squared of 0.72 (p value = 0.016).
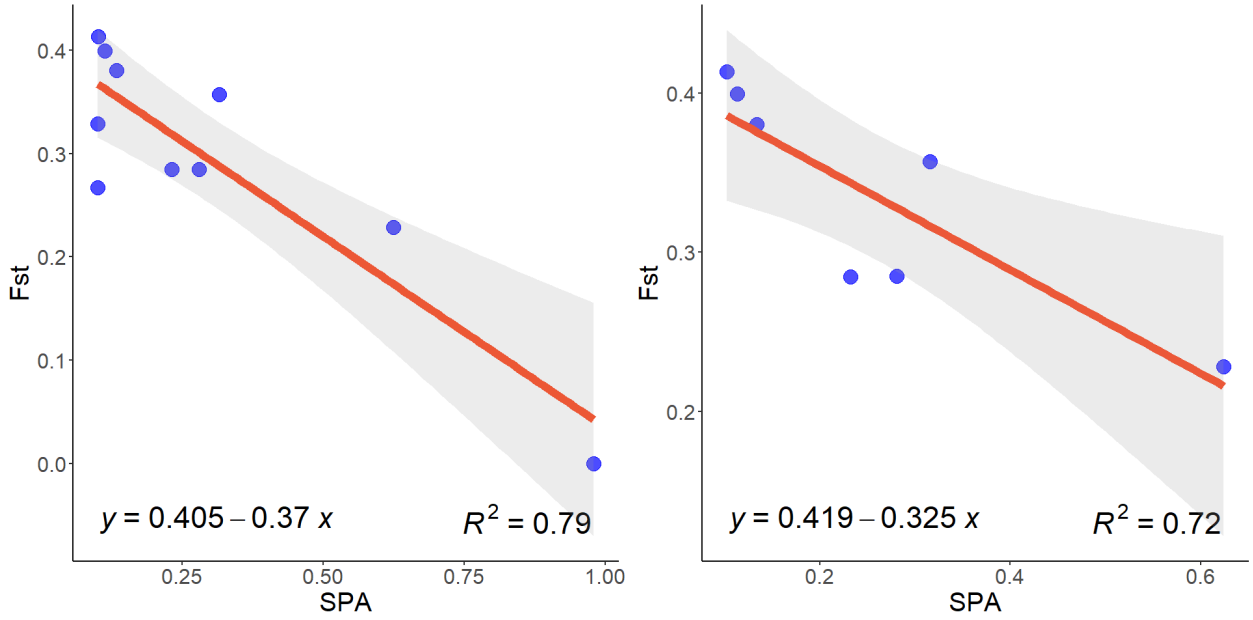

Figure 5: Linear regression for S a. barberi population A) SPA in all populations B) withouth barberi and chusensis populations. We can observe that in both cases the fixation index is explained with the SPA and the hypothesis holds without the imputation of the origin of the population (this means to add a 0 F_{st} value for barberi population)

From the model we extrapolated the values from SPA along the geography to new F_{st} values to get a geographical projection of fixation index from *S. a. barberi* population. (Figure 6)
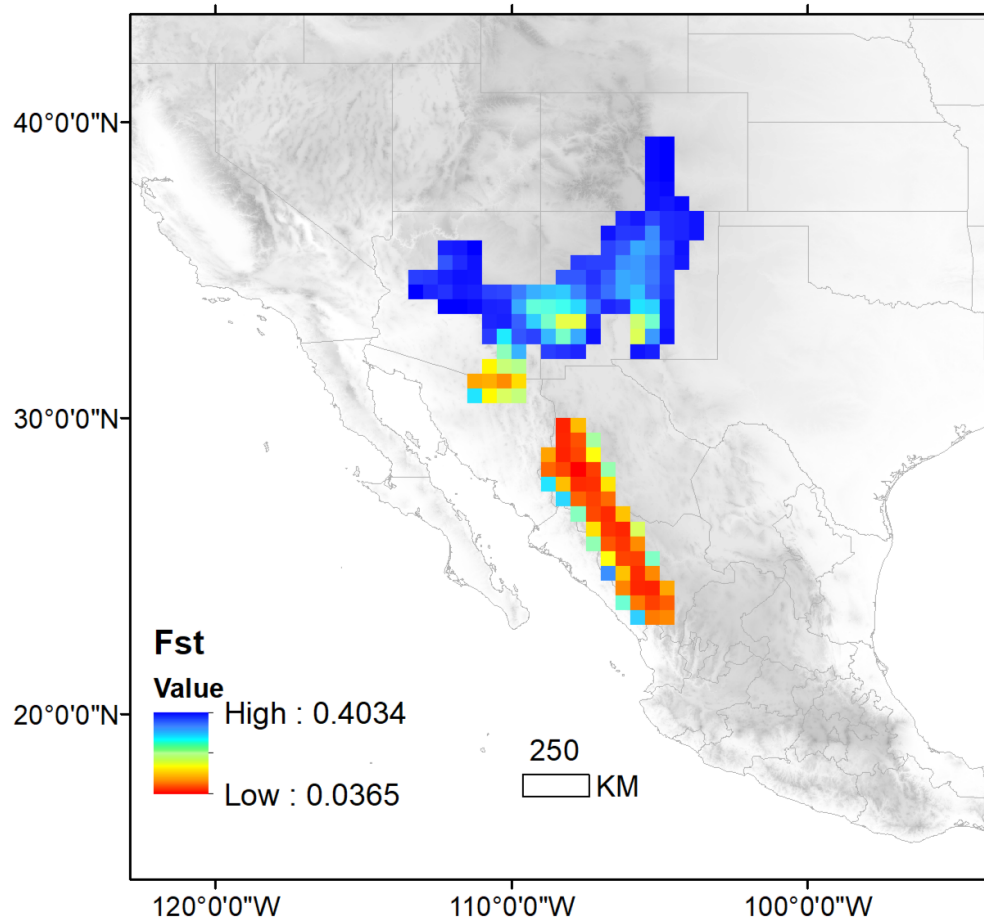
Figure 6: Projection of the F_{st} model to the geography in the actual distribution of *S aberti*

References

Beyer, Robert M, Mario Krapp, and Andrea Manica. 2020. "High-Resolution Terrestrial Climate, Bioclimate and Vegetation for the Last 120,000 Years." *Scientific Data* 7 (1): 236.

Bono, Jeremy M, Helen K Pigage, Peter J Wettstein, Stephanie A Prosser, and Jon C Pigage. 2018. "Genome-Wide Markers Reveal a Complex Evolutionary History Involving Divergence and Introgression in the Abert's Squirrel (Sciurus Aberti) Species Group." *BMC Evolutionary Biology* 18 (1): 1–17.

Burgin, Connor J, Jocelyn P Colella, Philip L Kahn, and Nathan S Upham. 2018. "How Many Species of Mammals Are There?" *Journal of Mammalogy.* Oxford University Press US.

Guisan, Antoine, and Wilfried Thuiller. 2005. "Predicting Species Distribution: Offering More Than Simple Habitat Models." *Ecology Letters* 8 (9): 993–1009.

Hedrick, Philip W. 2005. "A Standardized Genetic Differentiation Measure." *Evolution* 59 (8): 1633–38.

Karlin, Samuel. 1968. "Rates of Approach to Homozygosity for Finite Stochastic Models with Variable Population Size." *The American Naturalist* 102 (927): 443–55.

Krapp, Mario, Robert M Beyer, Stephen L Edmundson, Paul J Valdes, and Andrea Manica. 2021. "A Statistics-Based Reconstruction of High-Resolution Global Terrestrial Climate for the Last 800,000 Years." *Scientific Data* 8 (1): 228.

Leonardi, Michela, Emily Y Hallett, Robert Beyer, Mario Krapp, and Andrea Manica. 2023. "Pastclim 1.2: An r Package to Easily Access and Use Paleoclimatic Reconstructions." *Ecography*, e06481.

Lira-Noriega, Andrés, and Joseph D Manthey. 2014. "Relationship of Genetic Diversity and Niche Centrality: A Survey and Analysis." *Evolution* 68 (4): 1082–93.

Merow, Cory, Matthew J Smith, and John A Silander Jr. 2013. "A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter." *Ecography* 36 (10): 1058–69.

Nei, Masatoshi. 1986. "Definition and Estimation of Fixation Indices." *Evolution* 40 (3): 643–45.

Nogués-Bravo, David. 2009. "Predicting the Past Distribution of Species Climatic Niches." *Global Ecology and Biogeography* 18 (5): 521–31.

Owens, Hannah L, Lindsay P Campbell, L Lynnette Dornak, Erin E Saupe, Narayani Barve, Jorge Soberón, Kate Ingenloff, et al. 2013. "Constraints on Interpretation of Ecological Niche Models by Limited Environmental Ranges on Calibration Areas." *Ecological Modelling* 263: 10–18.

Peterson, A Townsend, Monica Papeş, and Jorge Soberón. 2008. "Rethinking Receiver Operating Characteristic Analysis Applications in Ecological Niche Modeling." *Ecological Modelling* 213 (1): 63–72.

Peterson, A Townsend, and Jorge Soberón. 2012. "Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right." *Natureza & Conservação* 10 (2): 102–7.

Phillips, Steven J, Robert P Anderson, and Robert E Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190 (3-4): 231–59.

Saupe, EE, V Barve, CE Myers, J Soberón, N Barve, CM Hensz, AT Peterson, H Lc Owens, and A Lira-Noriega. 2012. "Variation in Niche and Distribution Model Performance: The Need for a Priori Assessment of Key Causal Factors." *Ecological Modelling* 237: 11–22.

Singarayer, Joy S, and Paul J Valdes. 2010. "High-Latitude Climate Sensitivity to Ice-Sheet Forcing over the Last 120 Kyr." *Quaternary Science Reviews* 29 (1-2): 43–55.

Soberon, Jorge, and A Townsend Peterson. 2005. "Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas."

Soberón, Jorge, and Miguel Nakamura. 2009. "Niches and Distributional Areas: Concepts, Methods, and Assumptions." *Proceedings of the National Academy of Sciences* 106 (supplement_2): 19644–50.

Soberón, Jorge, and A Townsend Peterson. 2020. "What Is the Shape of the Fundamental Grinnellian Niche?" *Theoretical Ecology* 13 (1): 105–15.

Thorup, Kasper, Lykke Pedersen, Rute R Da Fonseca, Babak Naimi, David Nogués-Bravo, Mario Krapp, Andrea Manica, et al. 2021. "Response of an Afro-Palearctic Bird Migrant to Glaciation Cycles." *Proceedings of the National Academy of Sciences* 118 (52): e2023836118.

Upham, Nathan S, Jacob A Esselstyn, and Walter Jetz. 2019. "Inferring the Mammal Tree: Species-Level Sets of Phylogenies for Questions in Ecology, Evolution, and Conservation." *PLoS Biology* 17 (12): e3000494.

Valdes, Paul J, Edward Armstrong, Marcus PS Badger, Catherine D Bradshaw, Fran Bragg, Michel Crucifix, Taraka Davies-Barnard, et al. 2017. "The BRIDGE HadCM3 Family of Climate Models: HadCM3@ Bristol V1. 0." *Geoscientific Model Development* 10 (10): 3715–43.

Wright, Sewall. 1943. "Isolation by Distance." *Genetics* 28 (2): 114.