**Landscape changes decreases genetic diversity in** **the Pallas’ long-tongued bat**

Rosane G. Collevatti, Luciana C. Vitorino, Thiago B. Vieira, Monik Oprea, Mariana P. C. Telles

**Supplementary File S1**

**Materials and Methods**

**Populations and tissues sampling**

We aimed to sample at least 10 individuals per populations, but in four populations we could only capture less than 10. The number of individuals sampled per population ranged from seven to 22 (Table S1 in Appendix S1). Bats were captured and examined following methods described in Sikes et al. (2011) and wing membrane sampling followed the instructions for Wing Punch and Hair Sampling Protocols of American Natural History Museum (<http://research.amnh.org/vz/mammalogy>), and were stored in 95% ethanol.

**Genetic data**

DNA was extracted following the protocol of Miller et al. (1988). For microsatellite genotyping, forward primers were labeled with fluorescent dyes (6-FAM, HEX and NED, Applied Biosystems, CA), and amplifications were performed in a 15 µl reaction volume containing 0.13 µM of each primer, 1U Taq DNA polymerase (Phoneutria, BR), 0.21 mM of each dNTP, 1X reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCL2), 0.21 mg of BSA and 12.5 ng of template DNA. Amplifications were performed using a Verity® thermal cycler (Applied Biosystems, CA) with an initial denaturation of 94 oC for 5 min, followed by 45 cycles of 1 min at 94 oC, 1 min at the annealing temperature (according to each primer) and 1 min at 72 oC, and a final extension step of 30 min at 72 oC to control for Taq polymerase adenylation. The PCR products were subjected to electrophoresis on an ABI Prism 3100 automated DNA sequencer (Applied Biosystems, CA) and were sized by comparison to the GeneScan ROX 500 size standard (Applied Biosystems, CA).

Electropherograms were analyzed using GeneMapper® v4.1 software (Applied Biosystems, CA) and were visually checked to minimize genotyping errors due to stutter bands and drop out. Then, genotypes were analyzed using Micro-Checker 2.2.3 software (Van Oosterhout et al., 2004) to detect errors due to null alleles, stutter and dropout. The raw data using MicroChecker showed no significant evidence of genotyping errors or null alleles for microsatellite loci.

For sequencing, fragments were amplified by PCR in a 20 uL volume containing 3.0 uM of each primer, 1.5 unit Taq DNA polymerase (Phoneutria, BR), 250 uM of each dNTP, 1X reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl2), 2.0 uL of DMSO 50% and 15.0 ng of template DNA. Amplifications were performed using a GeneAmp PCR System 9700 (Applied Biosystems, CA, USA) with the following conditions: 96 °C for 2 min (1 cycle); 94 °C for 1min, 56°C for 1min, 72°C for 1min (30 cycles) and 72 °C for 10 min (1 cycle). PCR fragments were sequenced using the BigDye® Terminator v3.1 kit (Applied Biosystems, CA), and sequenced in forward and reverse directions using the GS 3500 Genetic Analyzer (Applied Biosystems, CA), according to the manufacturer’s instructions. Sequences were filtered using the software SeqScape v2.7 (Applied Biosystems, CA) with the following conditions: Phred quality score > 30, mixed base identification >= 15% (identification of heterozygous positions); maximum of mixed bases = 20%, maximum Ns = 10%, minimum clear length = 100 bp. Consensus sequences were obtained using SeqScape and aligned (see Appendix S3) with the software ClustalΩ (Sievers et al*.* 2011). A subset of individuals (one per population) was genotyped and sequenced twice, using duplicated PCRs, to verify the consistency of amplification.

**Genetic diversity and differentiation at microsatellite loci**

Genetic diversity parameters were estimated following Nei (1978). We also estimated mean number of alleles per locus (*A*) and allelic richness (*AR*, Mousadik and Petit, 1996). All analyses of genetic diversity and randomization-based tests for deviation from Hardy-Weinberg equilibrium were performed with the software FSTAT 2.9.3.2 (Goudet, 2002).

Wright’s *FST* (Wright, 1951) was estimated based on an analysis of variance of allele frequencies (Weir and Cockerham, 1984) implemented in the software FSTAT 2.9.3.2 (Goudet, 2002). Although *FST* is influenced by heterozygosity, it is widely used to quantify genetic differentiation and should appear in every study for comparison with other studies (Meirmans and Hedrick, 2011). In addition to *FST*, we estimated *G’ST* (Hedrick, 2005) and *D* (Jost, 2008). *G’ST* is similar to *FST* but takes into account the observed genetic diversity within population and the number of subpopulations (Hedrick, 2005), and might be accurate to compare measures of genetic differentiation among different landscapes. Jost’s *D* might also be advantageous in a study dealing with different populations at different landscapes, since it uses a multiplicative partitioning of diversity, based on the effective number of alleles rather than on the expected heterozygosity (Jost, 2008), and it is insensitive to the population size.

For Bayesian clustering we used the Structure Harvester software (Earl and von Holdt, 2012) to detect the number of K that best fits to the data, using Evanno's method (Evanno et al., 2005). We estimated a posteriori probability of K (number of groups) ranging from 1 to 18 (corresponding to each sampled locality), assuming uniform prior values, admixture model of ancestry and correlated allele frequencies. The analyses were based on 20 independent runs with 1,000,000 iterations following a burn-in-period of 250,000 iterations.

**Genetic diversity and differentiation at mitochondrial *CYB***

Population differentiation (*FST*) was assessed using analysis of molecular variance (AMOVA, Excoffier et al., 1992)and statistical significance was tested by a non-parametric permutation test (10,000 permutations). BAPS v6.0 analysis was performed with ‘not fixed number of clusters (*K*)’ (Corander and Marttinen, 2006; Corander et al., 2008) with an upper limit of *K* = 18.

Intraspecific phylogeny was inferred using a median-joining network (Bandelt et al., 1999). Characters were equally weighted and due to the high number of haplotypes and the high network complexity, we performed the analysis using different values of epsilon () of 0, 10, 20 and 30, and kept  = 10. Finally, we reduced the network complexity, collapsing haplotypes with frequency equal to 1 and number of mutated positions equal to 1 (tip haplotypes).

**Effective population sizes and population connectivity**

We estimated the mutation or coalescent parameter theta (*θ* = 2*μNe,* for mitochondrial haploid genomes*, θ* = 4*μNe,* for nuclear microsatellite diploid genomes), and the immigration rate among all population pairs (*M* = 2*Nem*/*θ,* for haploid genome*,* *M* = 4*Nem*/*θ,* for diploid genomes). Estimations were based on a Bayesian model using the Markov Chain Monte Carlo (MCMC) approach implemented in Lamarc 2.1.10 software (Kuhner, 2006). To set the priors, the GTR evolutionary model was selected for *CYB* using Akaike Information Criterion implemented in the software jModelTest2 (Darriba et al., 2012). Each analysis was run with 20 initial chains of 4,000 steps and three final chains of 50,000 steps. The chains were sampled every 100 steps. We used the default settings for the initial estimate of theta. The program was run four times for each parameter to certify convergence among runs and validate the analyses using Tracer 1.7.1 (Rambaut et al., 2018), and combined results were then generated using LogCombiner 1.10 (Rambaut and Drummond, 2017). Results were considered when ESS ≥ 200 (effective sample size) and when marginal posterior probability densities were unimodal and converged among runs. Effective population size was obtained from *θ* (Kingman, 1982).

**Ecological niche modeling**

We obtained 13,069 occurrence records of *G. soricina* across the Neotropics
(Figure S1 in Appendix S2) from the online databases GBIF (Global Biodiversity Information Facility http://www.gbif.org/) and Species Link (<http://splink.cria.org.br/>). All records were examined for probable errors (such as taxonomic identification) and duplicates. Records with spatial autocorrelation in climatic data were removed. The remaining 2,353 occurrence records (Table S2 in Appendix S1) were mapped in a grid of cells of 2.5’ x 2.5’ (longitude x latitude), ~4.6 x 4.6 km at the equator, encompassing the Neotropics, to generate the matrix of presence used to calibrate the ENMs.

For each algorithm and AOGCM, models were built using 75% of occurrences and tested against the remaining records (25%). To avoid modeling bias, this procedure was repeated 50 times for each combination of ENM and AOGCM by randomly splitting training (75%) and test data (25%). After eliminating the models with poor performance (True Skill Statistics, TSS < 0.7, Allouche et al., 2006), we computed binary maps using thresholds established by the AUC, the area under the receiver operating characteristic (ROC) curve (maximizing specificity + sensitivity). The frequency of predicted presence was used as a measure of consensus suitability from 50 initial models.

We applied a hierarchical ANOVA using the predicted suitability from all 75 models (5 ENMs x 5 AOGCMs x 3 Times) as the response variable to disentangle the effects of climate change on species distribution through time from predictive uncertainties in the potential distribution due to modeling components (i.e. ENMs, AOGCMs). For this, the ENM and AOGCM components were nested into the time component, but crossed by a two-way factorial design within each time period (Terribile et al., 2012). Because we do not have repetitions, the residual term from ANOVA represents the interaction between AOGCM and ENM (AOGCM x ENM; Sokal and Rohlf, 1995).

The full ensemble was obtained for each time period using the predictive performances (TSS) to compute a weighted mean of suitabilities, from which historical refugia were mapped (i.e. all grid cells with suitability values >= 0.5 during the three time periods). The threshold of 0.5 corresponds to the suitability value after excluding the lowest 10th percentile. We performed sensitivity analyses using different thresholds, 0.4, 0.6 and 0.7, corresponding to the 5th, 12th, and 17th percentiles, respectively, and the results differed only for a threshold > 0.7, resulting in a small refugium in Northeast Brazil.

**Landscape and climatic effects on genetic diversity**

We identified six different land cover types: natural vegetation remnants (forest and savanna), urban area, agriculture (mainly corn and soybean), water body, mining and outcrops, pasture and mixed use (see Figure 1, Table S5).

For Generalized Linear Mixed Model (GLMM) to analyze the effects of landscape structure and paleodistribution on the genetic parameters: genetic diversity (*He*), allelic richness (*AR*), inbreeding (*f*), haplotype diversity (*h*), number of haplotypes (*Nh*) and nucleotide diversity (π). We used as explanatory variables matrix type and percentage of natural vegetation remnants, suitability at present day derived from ENM and climatic stability. Climatic stability was defined as the difference in suitability between present-day and the LGM. We used suitability at present day due to high correlation among suitabilities at 21 ka, 6ka and 0 ka because of the quasi-stability through time in geographical range. In addition, because contemporary landscape structure and historical processes, such as population size fluctuations (Zellmer and Knowles, 2009; Carvalho et al., 2015; Carvalho et al., 2017), may affect genetic diversity and differentiation, we also used historical (based on coalescence) and contemporary (based on coancestry) effective population sizes (*Ne*) in each locality as explanatory variables.

We built several models with combinations of explanatory variables from landscape, climatic niche and effective population size. We also built a null model by randomly sampling data in the matrix keeping the b equal to zero (constant variables) for all explanatory variables (absence of specific landscape processes), to check whether random effects could also account for some of the variation in the genetic response variables. Landscape was fitted as a random factor whereas explanatory variables were fitted as fixed factors. Analyses were carried out using the MCMCglmm package (Haldfield, 2019) implemented in R version 3.6.1. (R Core team, 2019), using a Bayesian framework. We used a total of 80,000 iterations chains with 20,000 chains of burn-in and a Gaussian distribution. To select which model best explains the observed variation in genetic parameters among landscapes we calculated AIC corrected for small samples (AICc) and ΔAICc*i* , i.e. the difference between each model and the best models (where *i* represents each model). We also estimated Akaike’s weight of evidence (wAICc) as the relative contribution of model *i* to explain the observed pattern, given a set of competing models (Burnhan and Anderson, 2002). Models with ΔAICc < 2 were considered equally likely to explain the observed pattern.

**Landscape and climatic effects on genetic differentiation**

ResistanceGA package (Peterman, 2018) uses a genetic algorithm in a linear mixed effects model to iteratively optimize the adjustment of the resistance distance to the genetic distance. We extracted the land cover between all pairs of populations using the map from the Brazilian Ministry of Environment database (<http://mapas.mma.gov.br/mapas/aplic/> probio/datadownload.htm). We converted land use in resistance surface using resistance values based on information available in the literature (*e.g.* Estrada et al., 1993) and bat experts’ opinion. Resistance values were: vegetation remnants and water bodies = 1, agriculture and urban areas = 4, pastures = 3 and *Eucalyptus* spp. = 2.5. We used suitability at present day due to high correlation among suitabilities at 21 ka, 6ka and 0 ka because of the quasi-stability through time in geographical range. All surfaces were resampled to a resolution of 240 m. Maps processing was performed using ArcGIS v9.3. Optimization was performed with the function gdistance/commuteDistance that uses Circuitspace to calculate resistance distance.

To analyze genetic discontinuity among populations using the Delaunay network we linked the 18 sampled populations and we considered a discontinuity between pairs of populations when the ratio between the genetic distance (pairwise *FST*) and the geographical distance between them was higher than 10% of the highest ratios (Legendre and Legendre, 1998; Manel et al., 2003). Then, we mapped populations and discontinuity on the land use map.

**Results**

**Genetic diversity and population differentiation**

Population differentiation for microsatellite loci was low but significant (*FST* = 0.056, p < 0.001; D = 0.082, p < 0.001; *G’ST* = 0.146, p < 0.001). Most pairs of populations presented low values of pairwise *FST* (Table S6), pairwise Jost’s *D* and pairwise *G’ST* (Tables S7 and S8); the only exception was population 18, which showed high and significant values for all comparisons.

Delta K was also high for K = 5 (Appendix S2 Figures S2) and showed similar clustering patterns compared to K = 3, with population 18 assigned to a single cluster.

For CYB, populations showed significant genetic differentiation (*FST* = 0.631; *p < 0.001*), and high pairwise *FST* (Table S9). Almost all populations were clustered together in one group (cluster I, in blue), but individuals from populations 2, 4, 17 and 18 from Central Brazil, 8 from Northeast, and 14, from Southeast, were grouped in cluster II (brown).

Populations 1, 6, 12, 16 and 18 had very low historical effective population sizes based on coalescent analysis (Table S10). Population 18 received less than 1.0 migrant per generation from all populations, for microsatellite loci, and populations 2, 5, 6, 11 17 also had low *Nem* for *CYB*.

**Ecological niche modeling**

Ensembled ENM models were well evaluated. MRI: accuracy (0.7551), sensitivity (0.881), specificity (0.9496), kappa (0.770) and TSS (0.8377). MPI: accuracy (0.7559), sensitivity (0.7488), specificity (0.9938), kappa (0.703) and TSS (0.7426). MIROC accuracy (0.8075), sensitivity (0.8967), specificity (0.9847), kappa (0.766) and TSS (0.8814). CNRM: accuracy (0.7274), sensitivity (0.8843), specificity (0.9495), kappa (0.702) and TSS (0.8338). CCSM: accuracy (0.7588), sensitivity (0.8925), specificity (0.9624), kappa (0.809) and TSS (0.8549). The analysis of uncertainty using hierarchical ANOVA showed high proportional variance from time and AOGCM (Table S15, Figure S4), but variation was spatially structured and higher outside of the *G. soricina* occurrence range, indicating that the ENMs were able to detect the effects of climate changes on the distribution dynamics through the last glaciation, despite the AOGCM variation.

**References**

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

Bandelt, H.J., Forster, P., Rohl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Burnham, K.P., Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach*,* 2nd edn. Springer: New York.

Carvalho, C.S., Ribeiro, M.C., Côrtes, M.C., Galetti, M., Collevatti, R.G. 2015. Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm. Heredity 115, 216–224.

Carvalho, C.S., Ballesteros-Mejia, L., Ribeiro, M.C., Côrtes, M.C., Santos, A.S., Collevatti, R.G. 2017. Climatic stability and contemporary human impacts affect the genetic diversity and conservation status of a tropical palm in the Atlantic Forest of Brazil. Conserv. Genet. 18, 467–478.

Corander, J., Marttinen, P. 2006. Bayesian identification of admixture events using multi-locus molecular markers. Mol. Ecol. 15,2833–2843.

Corander, J., Marttinen, P., Sirén, J., Tang, J. 2008. Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9,1–14.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9,772.

Earl, D.A., vonHoldt, B.M. 2012. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. Conser. Genet. Resour. 4, 359-361.

Excoffier, L., Smouse, P., Quattro, J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131,479–491.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14*:* 2611-2620.

Forster, P., Bandelt, H.J., Rohl, A., et al. 2004. Network 4.2.0.1. Software available free at www.fluxus-engineering.com. Fluxus Technology Ltda.

Goudet, J. 2002. FSTAT 2.9.3.2.: A Program to estimate and test gene diversities and fixation indices. Available at http://www.unil.ch/izea/softwares/fstat.html.

Gruber, B., Adamack, A. 2019. Package ‘PopGenReport. Free available at ttp://www.popgenreport.org/>.

Haldfield, J. 2019. Package MSMCglmm: MCMC Generalised Linear Mixed Models. Free available at https://cran.rproject.org/web/packages/ MCMCglmm/ index.html.

Hedrick, P.W. 2005. A standardized genetic differentiation measure. Evolution 59,1633–1638.

Irwin, D.M., Kocher, T.D., Wilson, A.C. 1991. Evolution of the cytochrome b gene of mammals. J. Mol. Evol. 32, 128–144.

Jost, L. 2008. *GST* and its relatives do not measure differentiation. Mol. Ecol. 17, 4015–4026.

Kingman, J.F.C. 1982. The coalescent. Stoch. Proc. Appl.13, 235–248.

Kuhner, M.K. 2006. Lamarc 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22,768–770.

Legendre, P., Legendre, L. 1998. Numerical ecology. Oxford: Elsevier.

Manel, S., Schwartz, M.K., Luikart, G., Taberlet, P. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18, 189 -197.

Meirmans, P.G., Hedrick, P.W. 2011. Assessing population structure: FST and related measures. Mol. Ecol. Resour. 11, 5-18.

Miller, S.A., Dykes, D.D., Polesky, H.F. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nuc. Acids Resear. 16, 1215.

Mousadik, A., Petit, R.J. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* (L.) Skeels) endemic to Morocco. Theor. Appl. Genet. 92, 832–839.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics89, 583-590.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press: New York.

Peterman, W.E. 2018. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. Methods Ecol. Evol. 9, 1638–1647.

Rambaut, A., Drummond, A.J. 2017. LogCombiner 1.10. Available at: http://beast.bio.ed.ac.uk/Tracer.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. syy032.

R Core Team. 2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org>.

Sikes, R.J., Gannon, W.L., The Animal Care and use Committee of the American Society of Mammalogists. 2011. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. J. Mammal. 92, 235–253.

Sievers, F., Wilm, A., Dineen, D.G., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7,1–6.

Sokal, R.R., Rohlf, F.J. 1995. Biometry: The Principles and Practice of Statistics in Biological Research, 3rd edn. W. H. Freeman and Company: New York.

Terribile, L.C., Lima-Ribeiro, M.S., Araújo, M.B., Bizão, N., Collevatti ,R.G., Dobrovolski, R., Franco, A.A., et al. 2012. Areas of climate stability of species ranges in the Brazilian Cerrado: disentangling uncertainties through time. Natureza & Conservação 10,152–159.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P. 2004. MicroChecker: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4, 535–538.

Weir, B.S., Cockerham, C. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358-1370.

Wright, S. 1951.The genetic structure of populations. Annals Eugenics 15, 323–354.

[Zellmer, A.J](https://www.ncbi.nlm.nih.gov/pubmed/?term=Zellmer%20AJ%5BAuthor%5D&cauthor=true&cauthor_uid=19674302)., [Knowles, L.L](https://www.ncbi.nlm.nih.gov/pubmed/?term=Knowles%20LL%5BAuthor%5D&cauthor=true&cauthor_uid=19674302). 2009. Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. [Mol. Ecol.](https://www.ncbi.nlm.nih.gov/pubmed/19674302) 18, 3593-602.