Habitat loss does not always entail negative genetic consequences

Carolina S. Carvalho[1], Éder C. M. Lanes[1], Amanda R. Silva[1,2], Cecilio F. Caldeira[1], Nelson Carvalho-Filho[1], Markus Gastauer[1], Vera L. Imperatriz-Fonseca[1], Wilson Nascimento Júnior[1], Guilherme Oliveira[1], José O. Siqueira[1], Pedro L. Viana[2], Rodolfo Jaffé[1,3*]

[1] Instituto Tecnológico Vale, Belém-PA, Brazil.

[2]Museu Paraense Emílio Goeldi, Belém-PA, Brazil.

[3] Departamento de Ecologia, Universidade de São Paulo, São Paulo-SP, Brazil.

[*] **Correspondence:**

Rodolfo Jaffé

email: r.jaffe@ib.usp.br

Keywords: Gene flow, genetic diversity, isolation by resistance, landscape genomics, open-pit mining, RAD-sequencing, SNPs.

Abstract

Although habitat loss has large, consistently negative effects on biodiversity, its genetic consequences are not yet fully understood. This is because measuring the genetic consequences of habitat loss requires accounting for major methodological limitations like the confounding effect of habitat fragmentation, historical processes underpinning genetic differentiation, time-lags between the onset of disturbances and genetic outcomes, and the need for large numbers of samples, genetic markers and replicated landscapes to ensure sufficient statistical power. In this paper we overcame all these challenges to assess the genetic consequences of extreme habitat loss driven by mining in two herbs endemic to Amazonian Savannas. Relying on genotyping-by-sequencing of hundreds of individuals collected across two mining landscapes, we identified thousands of neutral and independent single-nucleotide polymorphisms (SNPs) in each species and used these to evaluate population structure, genetic diversity and gene flow. Since open-pit mining in our study region rarely involves habitat fragmentation, we were able to assess the independent effect of habitat loss. We also accounted for the underlying population structure when assessing landscape effects on genetic diversity and gene flow, assessed the sensitivity of our analyses to the resolution of spatial data, and used annual species and cross-year analyses to minimize and quantify possible time-lag effects. We found that both species are remarkably resilient, as genetic diversity and gene flow patterns were unaffected by habitat loss. Whereas historical habitat amount was found to influence inbreeding; heterozygosity and inbreeding were not affected by habitat loss in either species, and gene flow was mainly influenced by geographic distance, pre-mining land cover and local climate. Our study demonstrates that it is not possible to generalize about the genetic consequences of habitat loss, and imply that future conservation efforts need to consider species-specific genetic information.

Introduction

In spite of ample evidence showing that habitat loss has large, consistently negative effects on biodiversity (Fahrig, 2003), very few studies have assessed the consequences of habitat amount on genetic variation (DiLeo and Wagner, 2016; Monteiro et al., 2019). Habitat loss can potentially impact the demographics of natural populations, reducing population size, gene flow and genetic diversity, and thereby increasing inbreeding and extinction risk (Allendorf et al., 2013). Understanding the genetic consequences of habitat loss is therefore essential to safeguard biological diversity and fulfill Aichi Biodiversity Targets and Sustainable Development Goals (Tittensor et al., 2014).

Important limitations constrain the quantification of habitat amount effects on genetic variation. Firstly, habitat loss and fragmentation are often confounded, so disentangling the relative contribution of habitat amount requires controlling for fragmentation (Fahrig, 2003). Secondly, landscape effects can also be easily confounded with historical demographic processes that underlying population structure (Llorens et al., 2018). Thirdly, a coarse resolution of spatial data and time-lags between the onset of disturbances and genetic responses may mask the effects of recent landscape modification (Anderson et al., 2010; Balkenhol et al., 2016). Finally, large numbers of samples and genetic markers, and replicated sampling designs that capture enough landscape heterogeneity are needed to detect or rule out possible landscape effects with sufficient statistical power (McCartney-Melstad et al., 2018; Storfer et al., 2010). Failure in overcoming any of these limitations may hide important detrimental effects to the maintenance of genetic variability, or reveal spurious patterns unrelated to habitat loss.

Few studies have attempted to quantify the impact of habitat loss on both genetic diversity and gene flow, and neither has yet accounted for all the methodological limitations outlined above (Balkenhol et al., 2016; DiLeo and Wagner, 2016). Here we fill this important knowledge gap assessing the genetic consequences of extreme habitat loss driven by open-pit mining in two endemic plants from the Eastern Amazon. Firstly, we were able to assess the independent effect of habitat loss, as open-pit mining in our study region rarely involves habitat fragmentation (See Fig. S1 in Supporting Information). We also controlled for historical demographic processes by accounting for the underlying population structure when assessing landscape effects on genetic diversity and gene flow; assessed the sensitivity of our analyses to the resolution of spatial data; and used annual species (which complete a full reproductive cycle and die within one year) to minimize possible time-lag effects. Finally, we sampled hundreds of individuals scattered across two separate regions exposed to mining, and genotyped them at thousands of single nucleotide polymorphisms (SNPs) distributed across their genomes to assure high statistical power.

The banded iron formations known as Cangas from the Carajás Mineral Province in the Eastern Amazon harbor one of the world's largest deposits of high-grade iron ore (Skirycz et al., 2014), which has attracted substantial attention from mining companies. In fact two of the world's largest iron-ore mines are located in the region (Fig. 1), with operations in Serra Norte dating back to the 1980s, while Serra Sul only began activities in 2014. Based on a curated inventory of Canga plants from this region (Viana et al., 2016), we selected annual herbs (to minimize time-lag effects), occurring exclusively in Canga ecosystems (where mining activities are concentrated), and endemic to our study region. From the few available species meeting these criteria, *Brasilianthus carajensis* (Melastomataceae) and *Monogereion carajensis* (Asteraceae), were among the easiest to find and identify in the field. Since both species seem to be pollinated by insects and their seeds dispersed by the wind (Cruz et al., 2016; Rocha et al., 2017), we expected them to be susceptible to habitat loss,

given that the progeny of insect- and wind-pollinated plants has been shown to be strongly negatively affected by habitat fragmentation (Aguilar et al., 2019). Relying on genotyping-by-sequencing of hundreds of individuals from both species collected across these two mining landscapes, we assessed the influence of habitat loss on genetic diversity and gene flow. We also performed a suit of germination experiments to determine if these plants are able to successfully colonize iron-ore mines. We predicted that: i) Individuals surrounded by undisturbed habitats would show higher genetic diversity and lower inbreeding than those exposed to habitat loss driven by mining; ii) Gene flow would be best explained by recent landscape modifications, and mining areas would represent barriers to gene flow; iii) Mining waste substrates would hinder germination.

Materials and Methods

Sampling, DNA extraction, and genome size estimation

We collected leaf tissue samples of 150 individuals of *B. carajensis* and 207 individuals of *M. carajensis* between February and June 2017 (SISBIO collection permit N. 48272-4). Samples were collected in the main Canga plateaus of our study area, comprising the entire occurrence range of both species. Care was taken to sample individuals at or around iron ore mines and separated by at least 20 m from each other to avoid collecting siblings (Fig. 1). Both species exhibit a patchy distribution, with *B. carajensis* usually occurring in aggregates of up to hundreds of individuals occupying rocky exposed soils, while isolated individuals of *M. carajensis* were often found in shaded areas and soils with a larger organic mater content (personal observation). Our study area harbors some of the world's largest iron-ore mines, one located in Serra Norte, comprised of an archipelago of Canga plateaus, and one in Serra Sul formed by a large continuous plateau. While in Serra Norte we collected 103 and 120 individuals, in Serra Sul we sampled 17 and 149 individuals of *B. carajensis* and *M. carajensis* respectively. To ensure high DNA quality and concentration, we preserved *B. carajensis* samples in silica and *M. carajensis* samples in 10 mL of a NaCl-saturated solution of 2% CTAB (Rogstad, 1992), and stored them at -80 °C until analysis. Total DNA of *B. carajensis* was extracted using a CTAB 2% protocol (Doyle and Doyle, 1987) followed by a DNA purification protocol (Michaels et al., 1994); whereas the DNeasy Plant Mini Kit (Qiagen, EUA) was used for *M. carajensis.* DNA concentration for both species was quantified using the Qubit High SensitivityAssay kit (Invitrogen), and DNA integrity assessed through 1.2% agarose gel electrophoresis. All DNA samples were adjusted to a final concentration of 5 ng/µL in a final volume of 30 µL. We used flow cytometry to estimate genome size in both species. Nuclei were obtained from fresh leaf tissues chopped along with references in general purpose buffer with 1% Triton X-100 and 1% PVP-30 (Loureiro et al., 2007). The whole sample preparation was conducted on ice until the events acquisition on a PI fluorescence mean under a 575/26 bandpass filter. Triplicates of 1000 PI stained nuclei were analyzed under a 488 nm laser on BD FACS Aria II cytometer. The internal standard used was tomato (*Lycopersicon esculentum*; 2C = 1.98 pg, (Dolezel et al., 1992)).

RAD sequencing and SNP discovery

DNA samples were shipped to SNPSaurus (http://snpsaurus.com/) for sequencing and bioinformatic analyses of raw reads (trimming and variant calling). Briefly, nextRAD genotyping-by-sequencing libraries were prepared (Russello et al., 2015) using Nextera DNA Flex reagent (Illumina, Inc) and considering the estimated genome size of each species (2C DNA content was 508 Mbp in *B. carajensis* and 6,284 Mbp in *M. carajensis*). Fragmented DNA was then amplified for 25 cycles at 75

degrees, with one of the primers matching the adapter and extending 8 nucleotides into the genomic DNA with the selective sequences GTGTAGAA (*B. carajensis*) and TGCAGGAG (*M. carajensis*). Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer will be efficiently amplified. The nextRAD libraries were sequenced on a HiSeq 4000 with four lanes of 150 bp reads (University of Oregon). Genotyping analysis used custom scripts (SNPsaurus, LLC) that trimmed the reads using bbduk (BBMap tools, http://sourceforge.net/projects/bbmap/): bbmap/bbduk.sh in=$file out=$outfile ktrim=r k=17 hdist=1 mink=8 ref=bbmap/resources/nextera.fa.gz minlen=100 ow=t qtrim=r trimq=10. A de novo reference was created by collecting 10 million reads in total, evenly from the samples, and excluding reads that had counts fewer than 5 or more than 700 for *B. carajensis*; and fewer than 6 or more than 1,000 for *M. carajensis*. The remaining loci were then aligned to each other to identify allelic loci and collapse allelic haplotypes to a single representative. All reads were mapped to the reference with an alignment identity threshold of 90% using bbmap (BBMap tools). Genotype calling was done using Samtools and bcftools (samtools mpileup -gu -Q 12 -t DP, DPR -f ref.fasta -b samples.txt | bcftools call -cv - > genotypes.vcf). The vcf was filtered to remove alleles with a population frequency of less than 3%. Loci were removed that were heterozygous in all samples or had more than 2 alleles in a sample (suggesting collapsed paralogs). The absence of artifacts was checked by counting SNPs at each read nucleotide position and determining that SNP number did not increase with reduced base quality at the end of the read. A total of 43,887 contigs were generated for *B. carajensis* (sequencing depth ranged between 18 and 239), and 36,040 for *M. carajensis* (depth ranging between 14 and 246). Geographic coordinates in decimal degrees, genotypes in Variant Call Format and sequences in FASTA format for both species are provided here: https://figshare.com/articles/Habitat_loss_does_not_always_entail_negative_genetic_consequences/8224175.

Neutral datasets

The R package *r2vcftools* (https://github.com/nspope/r2vcftools) - a wrapper for VCFtools (Danecek et al., 2011) - was used to perform final quality control on the genotype data (see detailed script here: https://github.com/rojaff/r2vcftools_basics. We excluded loci containing more than 30% missing data, and filtered them for quality (Phred score > 50 both species), read depth (30 – 240 both species), linkage disequilibrium (LD, $r^2 < 0.6$ and $r^2 < 0.4$ for *B. carajensis* and *M. carajensis*, respectively), and strong deviations from the Hardy Weinberg Equilibrium (HWE, $p < 0.0001$ both species) (O'Leary et al., 2018). Additionally, we removed any potential loci under selection detected through genome scans, whereby F_{ST} outlier tests were applied after assessing population structure with the *snmf* function from the LEA package (Frichot et al., 2014). The genomic inflation factor and a trial-and-error approach were used to calibrate *p*-values, and the Benjamini-Hochberg algorithm ($q = 0.05$) was used to correct for false discovery rates (François et al., 2016), see example script here: http://membres-timc.imag.fr/Olivier.Francois/LEA/files/LEA_snmf.html). The resulting sets of neutral and independent loci were then used in all subsequent analyses.

Genetic structure

We used two complementary genetic clustering method to assess population structure: Admixture *(Alexander et al., 2009)* and Discriminant Analysis of Principal Components - DAPC from the *adegenet* package (Jombart et al., 2010; Jombart and Ahmed, 2011). Admixture puts individuals into groups to maximize HWE while that DAPC minimizes allelic differences. For the former analysis, the number of ancestral populations (*k*) was allowed to vary between 1 and 10, and the best *k* was chosen based on cross-validation errors (Frichot et al., 2014). For the second analyses,

the number of clusters was assessed using the function "*find.cluster*", which runs successive *k*-means clustering with an increasing number of clusters, and then determined the best-supported number of genetic clusters using the Bayesian Information Criterion (BIC). Considering the ancestry coefficients assigned by Admixture, we then estimated expected heterozygosity (H_E) and inbreeding coefficients (F) for each genetic cluster using the "het" option in VCFtools implemented in *r2vcftools* (Danecek et al., 2011). Additionally, we assessed fine-scale spatial genetic structure within each genetic cluster by quantifying spatial autocorrelation in Yang's genetic relatedness between pairs of individuals *(Yang et al., 2010)*. To do so we used local polynomial fitting (LOESS) of pairwise relatedness and pairwise geographic distance. In order to test if the average observed relatedness predicted by LOESS at a given distance differs from the null model, row and column indices for the relatedness matrix were permuted 999 times. The smoothing parameter was fixed as the default for loess() in R (span = 0.75). At each permutation a LOESS model was re-fitted using the permuted relatedness and geographic distance matrix, and 95 % percentiles of the permutation-derived LOESS predictions were used to generate confidence envelopes around the null expectation (see example script here: https://github.com/rojaff/Lplot).

Land cover maps

To account for time-lag effects when assessing the genetic consequences of habitat loss, we built land cover maps for different years (2016, 2014, 2011 and 1979), comprising pre-mining maps (1979). *Landsat* 2 images (spatial resolution of 80 meters in 7 spectral bands) were used for year 1979, *Landsat* 7 images (spatial resolution of 30 meters in 7 spectral bands) for year 2011, and *Sentinel* images (spatial resolution of 10 meters in 4 spectral bands) for years 2014 and 2016. Images were downloaded from the Earth Explorer Server (https://earthexplorer.usgs.gov/), selecting scenes from the month of July to minimize clouds. All images were converted to ground reflectance in percentage using the ATCOR algorithm of the PCI Geomatica 2016 software. The scenes were joined to create a mosaic of the study area and derive the Normalized Difference Vegetation Index - NDVI (Tarpley et al., 1984). We then employed the eCognition 9 software using a Geographic Object-Based Image Analysis (GEOBIA) to classify land cover types. The Multi-resolution classification algorithm was selected, given that it allows obtaining segments with different sizes due to brightness, shape, smoothness and compactness. Montane Savanna (Canga), Water, Forest, Mine, Pasture and Urban classes were identified.

Genetic diversity

To assess the effect of habitat loss on genetic diversity, we regressed individual-level diversity metrics (H and f), estimated with "het" option in VCFtools (Danecek et al., 2011), on historical habitat amount and habitat loss (measured in area) driven by mining in different years, using high resolution land cover maps (10 x 10m). By so doing we explicitly evaluated the effect of habitat loss within a buffer accounting for historical habitat amount. Historical habitat amount was calculated by extracting the proportion of Canga habitat in a buffer surrounding each individual using pre-mining maps (1979). Habitat loss in different years (2011, 2014, and 2016) was calculated by subtracting habitat amount for a given year from historical habitat amount. To select an optimal buffer size we first ran uni-variate models using habitat amount extracted from the most recent land cover maps (2016) with buffers varying in size between 100 and 900m, and then compared all models using AIC. As habitat amount calculated with the largest buffers (900m) was always among the best models ($\Delta AIC \leq 2$), we chose this buffer size to encompass a greater portion of lost areas (Table S1). Additionally, the small flowers of our study species suggest their pollinators do not forage beyond 1Km (Greenleaf et al., 2007). In Serra Norte, 44% of individuals of *B. carajensis* and 41% of *M.*

carajensis experienced habitat loss, whereas 55% and 57% of individuals of *B. carajensis* and *M. carajensis,* respectively, faced habitat loss in Serra Sul.

Because genetic parameters (*H* and *f*) are affected by demographic processes, we explicitly accounted for demography either by analyzing individuals belonging to the same cluster or by including genetic cluster identity as a random effect in our models. In Serra Norte, which comprises an archipelago of Canga plateaus, we fit linear mixed-effect models, using each plateau as a random effect to account for site-specific characteristics and spatial autocorrelation. In the case of *B. carajensis* from Serra Norte, we also included a random effect specifying the genetic cluster containing each individual (see genetic structure results using the Admixture software). In Serra Sul, which comprises a single large plateau, we used generalized least-squares models (GLS) fitted with different correlation structures (linear, exponential, Gaussian, and spherical) to explicitly model spatial autocorrelation. The "*weight*" argument was used in some cases to account for heteroscedasticity. Raw *f* and logit-transformed *H* were used as response variables and models fitted using the *nmle* R package (Pinheiro et al., 2009). For each model, we calculated AIC, the difference of each model and the best model (ΔAIC) and the Akaike's weight of evidence (wAIC). Models with $\Delta AIC \leq 2$ were considered as equally plausible (Zuur et al., 2009). The set of best models ($\Delta AIC \leq 2$) were compared to reduced models without each predictor variable, using likelihood ratio tests (LRT, $\alpha = 0.05$), and all models were validated by plotting residual *vs.* fitted values and by checking for residual autocorrelation. Relative variable importance was calculated summing the Akaike weights of the best-fitting models in which the variable of interest was present (https://github.com/carolinacarvalho/Importance_plot). Model averaging across the set of best models was used to compute parameters estimates that account for uncertainty in model selection (Burnham and Anderson, 2002). We also estimated confidence intervals for parameter estimates to assess the statistical power of our models (Hoenig and Heisey, 2001).

Gene flow

To assess the effect of habitat loss on gene flow, we first optimized gene flow hypotheses and then tested them by modeling isolation by resistance (IBR, (McRae, 2006)). Yang's genetic relatedness between pairs of individuals (Yang et al., 2010) was used as a proxy for recent gene flow, given that it represents the number of common ancestors in the recent past (Wang, 2017). Although relatedness is not a direct measure of gene flow, it has been widely used to describe genetic connectivity among individuals (Balkenhol et al., 2016; Monteiro et al., 2019; Storfer et al., 2010) and a recent simulation study showed that it is among the most accurate individual-based genetic distance metric for landscape genetic studies (Shirk et al., 2017). Resistance to gene flow due to mining was modeled using land cover maps for different years (2016, 2014, 2011 and 1979) containing only the major land cover classes of our study region: Montane Savanna (Canga), Forest (evergreen forest) and Mine. Water bodies, Pasture and Urban areas were excluded because they occurred outside the extent of our samples (Fig. 1). By so doing we were able to evaluate the permeability to gene flow of each land cover class; and test whether habitat loss driven by mining hindered gene flow across our replicated landscapes. Additional variables found to be important predictors of gene flow in other plants (Dyer, 2016; Lanes et al., 2018) were modeled along with land cover, including geographic distance, elevation (DEM retrieved from the USGS Earth Explorer), terrain roughness (generated from the DEM using the Terrain Analysis plug-in from QGIS), and bioclimatic variables (retrieved from WorldClim). To select a set of orthogonal variables explaining most climatic variation across our study area, we first ran separate principal component analyses (PCA) for each species using the extracted values from all 19 WorldClim bioclimatic layers plus elevation (scaled) (Fig. S2). We then selected the three variables showing the strongest correlation with the first, second and third PCA axis (which explained more than 85% of total variance in both

B. carajensis and *M. carajensis*). These were minimum temperature of coldest month (bio06) and precipitation of wettest (bio16) and coldest quarter (bio19) for *B. carajensis*; and minimum temperature of coldest month (bio06), precipitation of wettest quarter (bio16) and temperature seasonality (bio04) for *M. carajensis*.

A genetic algorithm (unrelated to the genetic data) was implemented through the *ResistanceGA* package to generate optimized resistance surfaces for each one of these variables (Peterman, 2018). The advantage of this optimization procedure is that it relies on empirical genetic data, ensuring that resistance values attributed to resistance surfaces will relate meaningfully to the movement of genes across the landscape (Peterman, 2018). Moreover, since resistance values are not defined *a priori*, the optimized resistance surfaces can be considered unbiased by any existing knowledge or human preferences. In the case of land cover maps, random initial resistance values were assigned for each class; then pairwise resistance distances were measured using random-walk commute times; and finally pairwise genetic distance was regressed on resistance distance using maximum likelihood population effect models (MLPE, see below). The whole process was iterated until no significant change was found in the objective function (Peterman, 2018). We then performed the same steps for the remaining continuous predictors, but instead of assigning random initial resistance values, eight types of transformations were applied to the raw values. In this case, two parameters controlling Ricker and Monomolecular functions were iteratively varied during the optimization (Peterman, 2018). Ten independent runs of optimization were conducted for each surface to assess the convergence in parameter estimates (Khimoun et al., 2017). All rasters were set to Universal Transverse Mercator (UTM) projection, and cropped to the extent of sampling locations plus a buffer area of 5 km to minimize border effects (Lanes et al., 2018). Land cover resistance surfaces and terrain roughness were optimized using 250 x 250 m resolution maps, while 900 x 900 m resolution maps were used for WorldClim layers as this is the highest available. Serra Norte and Serra Sul were analyzed separately aiming to replicate IBR analyses in two separate areas exposed to open-pit mining.

Using the program Circuitscape V4.0 (McRae, 2006), we then calculated pairwise resistance distances between all samples, employing the optimized resistance surfaces described above plus a surface where all pixels were set to 1 to create a null model of isolation by geographic distance (IBD). To assess isolation by resistance (IBR), defined as the correlation between genetic and resistance distances (McRae, 2006), we fitted mixed-effects regression models using penalized least squares and a correlation structure designed to account for the non-independence of pairwise distances (maximum-likelihood population effects - MLPE: https://github.com/nspope/corMLPE; (Clarke et al., 2002)). Yang's genetic relatedness between individuals was used as the response variable and the different resistance distances (contemporary and historical land cover, elevation, terrain roughness, temperature, precipitation, and geographic distance) as predictors. All MLPE models accounted for the underlying population structure, either by considering only individuals belonging to the same genetic cluster (most cases), or by including an additional random effect specifying if pairwise distances represented individuals from the same or from different genetic clusters (the case of *B. carajensis* from Serra Norte, see genetic structure results using the Admixture software). To evaluate the incidence of time-lag effects potentially masking mining effects on gene flow, we first fitted uni-variate models for each species and region using resistance distances from land cover surfaces from all years, plus those from geographic distance surfaces. The best models were selected using the Akaike Information Criterion ($\Delta AIC < 2$), and whenever geographic distance was found among the best models we considered IBD as the most parsimonious gene flow model (Balkenhol et al., 2016; Burnham and Anderson, 2002). To evaluate the sensitivity of our analysis to the resolution (grain size) of spatial data, we also compared uni-variate land cover models containing resistance distances computed from surfaces with different grain sizes (100 x 100 m, 300 x 300 m,

600 x 600 m and 900 x 900 m). Results were consistent across the different resolutions (Table S2), so we ran all subsequent analysis using a grain size of 900 x 900 m. We then fitted multiple regression models containing resistance distances from the best uni-variate land cover models selected in the previous step and resistance distance from all other optimized surfaces for each species and region. Models containing all possible combinations of non-collinear predictors ($r < 0.6$, Fig. S3) were compared using the *dredge* function from the package *MuMIn* (https://github.com/rojaff/dredge_mc; (Barton and Barton, 2015)), and best models were selected using AIC. Likelihood ratio tests (LRT) were performed to assess the influence of each predictor variable on the best model's log-likelihood (Jaffé et al., 2016), and relative variable importance, model-averaged parameter estimates and confidence intervals were calculated as described above. Finally, we carried out a barrier analysis to identify genetic discontinuities between individuals by using Monmonier's algorithm and Gabriel's graph implemented in package *adegenet* (Jombart and Ahmed, 2011).

Germination experiments

To evaluate if seeds from both study species are able to germinate inside iron ore mines, we ran a set of germination experiments. Seeds from both species were sown over four different substrates (Whatman® paper, Canga topsoil, forest topsoil, and mining waste substrate) placed in plastic boxes (Gerbox – 11 x 11 x 4 cm) and kept in a growth chamber (Fitotron SGC 120, Weiss Technik, UK) under continuous darkness, constant temperature (20°C) and air humidity (60%) for 33 consecutive days, from September 4th to October 7th 2018. Substrates received distilled water until the retention capacity, and water losses by evaporation were replaced daily. All treatments were carried out with five replicates for each substrate in each species. Each replicate contained 25 seeds from *B. carajensis* and 12 seeds from *M. carajensis*. The number of germinated seeds was recorded daily, with germination defined as the emission of 2 mm of primary root.

Results

Neutral dataset

We collected leaf tissue samples of 150 individuals of *B. carajensis* and 207 individuals of *M. carajensis* distributed across the entire occurrence range of both species and surrounding two large iron ore mines (Fig. 1). Samples were frozen and their DNA later extracted and shipped for genotyping-by-sequencing (RAD-sequencing) and bioinformatic processing. We identified a total of 10,016 SNPs in *B. carajensis* and 20,464 SNPs in *M. carajensis,* but after filtering these for missing data, quality, depth, linkage disequilibrium, deviations from the Hardy-Wenberg Equilibrium and F_{ST} outlier loci, we obtained sets of neutral and independent markers containing 1,411 and 6,052 loci for each species, respectively.

Genetic structure

Two complementary genetic clustering approaches used to assess population structure (Admixture and DAPC) indicated the presence of three clusters in *B. carajensis* and two in *M. carajensis* (Fig. 2, Fig. S4-S6). Cluster-level heterozygosity was slightly higher in *B. carajensis* than in *M. carajensis,* and significant albeit low inbreeding was found in one genetic cluster of each

species (Fig. 2). Both species showed spatial autocorrelation in genetic relatedness in Serra Norte but not in Serra Sul, and the strength of spatial autocorrelation was higher in *B. carajensis* (Fig. 2).

Genetic diversity

To assess the effect of habitat loss on genetic diversity, we regressed individual-level diversity metrics on historical habitat amount and habitat loss driven by mining in different years. Heterozygosity (*H*) and inbreeding (*f*) were not influenced by habitat loss, either in Serra Norte nor in Serra Sul, as the set of best-fitting models always included null models or historical (pre-mining) habitat amount (Fig. 3, Table S3). Although confidence intervals for the effect of habitat loss on heterozygosity were usually narrow, those describing the effect of habitat loss on inbreeding were very broad in *B. carajensis* and moderately so in *M. carajensis* (Table S4). Historical habitat amount was found to be associated with inbreeding in both species, although the direction of the effect varied (Fig. 4, Table S4).

Gene flow

To assess the effect of habitat loss on gene flow, we first employed a genetic algorithm to optimize gene flow hypotheses, then calculated resistance distances between individual samples, and finally modeled isolation by resistance (IBR) regressing pairwise genetic relatedness on resistance distances through Maximum Likelihood Population Effects (MLPE) models. While resistance to gene flow due to mining was modeled using land cover maps for different years (2016, 2014, 2011 and 1979), additional covariates modeled along with land cover included geographic distance, terrain roughness, elevation, and bioclimatic variables. The optimization of resistance surfaces revealed that Canga was the land cover class representing lowest resistance to gene flow in both species, whereas mining areas and evergreen forests imposed higher resistance (Fig. S7-S10). However, univariate MLPE regression models revealed that geographic distance usually explained relatedness patterns as well as land cover (Table S2), and only pre-mining land cover (1979) was found to explain relatedness patterns better than geographic distance in *M. carajensis* from Serra Norte. Our results thus reveal that mining neither hinders nor facilitates gene flow in these two endemic annual plants. While these results hold across different resolutions (Table S2), an independent barrier analysis also failed to identify barriers between individuals separated by mining areas (Fig. S11). Multiple MLPE regression models showed that isolation by geographic distance (IBD) explained genetic relatedness patterns in *B. carajensis*, whereas IBR was more important in *M. carajensis* (Fig. 3, Table S5). In all cases, genetic relatedness decreased with increasing resistance (Fig. 4, Table S6).

Germination experiments

Germination experiments revealed that seeds from both species are able to germinate in mining waste substrates. Whereas *M. carajensis* showed similar germination in Canga and mining substrates, germination rates of *B. carajensis* were higher in Canga topsoil (Fig S12).

Discussion

Our study is the first to assess the genetic consequences of habitat loss while accounting for all the major limitations constraining the quantification of habitat amount effects on genetic variation.

Our results reveal that habitat loss driven by mining did not affect genetic diversity or gene flow in two endemic herbs from Amazonian Savannas. Whereas historical habitat amount was found to influence inbreeding, heterozygosity and inbreeding were not affected by habitat loss in either species. Finally, gene flow was mainly influenced by geographic distance in *B. carajensis* and by pre-mining land cover and local climate in *M. carajensis*.

The genetic structure in *B. carajensis* mirrored that from the co-occurring perennial morning glory *Ipomoea maurandioides* (Lanes et al., 2018), showing two differentiated genetic clusters in Serra Norte, while *M. carajensis* only presented one cluster in Serra Norte and another one across the remaining distribution range. This genetic structure was considered when assessing landscape effects on genetic diversity and gene flow in order to control for historical demographic processes. Additionally, since open-pit mining in our study region did not usually result in decreased structural connectivity between habitat patches (since these were already separated, Fig. S1), we were able to assess habitat loss effects that were not heavily influenced by habitat fragmentation (Fahrig, 2003).

The maintenance of genetic diversity in spite of extreme habitat loss suggests that our study plants are able to colonize mining environments and maintain gene flow across open-pit mines. Germination experiments revealed that seeds from both species can indeed germinate in mining waste substrates. Additionally, both plant species showed extensive gene flow across mining areas, and mining neither enhanced nor hindered gene flow. Similar results were found for a threatened orchid and the American pika, which showed analogous levels of genetic diversity in mining and natural habitats (Esfeld et al., 2008; Waterhouse et al., 2017), although neither gene flow nor historical effects were assessed. Inbreeding levels in our focus species are comparable to those observed in the widespread *I. maurandioides* (Lanes et al., 2018), and since they were associated with historical habitat amount they seem to reflect density-dependent selfing (Leimu et al., 2006). Our results thus reveal that some insect-pollinated and wind-dispersed plants do not experience genetic erosion due to habitat loss (Aguilar et al., 2008, 2019). A possible mechanism explaining the maintenance of genetic diversity is seed dormancy over long periods of time, which would result in multiple overlapping generations being represented in the seed bank (Honnay et al., 2008).

Both species presented spatial autocorrelation in genetic relatedness in Serra Norte but not in Serra Sul, indicating a more restricted gene flow in the Canga archipelago of Serra Norte than in the large continuous plateau of Serra Sul. Additionally, geographic distance was weakly correlated with recent land cover resistance in Serra Norte but not in Serra Sul, where it was strongly correlated with land cover resistance from all years (Fig. S3). We thus expected that isolation by resistance (IBR) would be easier to disentangle from isolation by distance (IBD) in Serra Norte than in Serra Sul. In Serra Norte, however, geographic distance and pre-mining land cover (highly correlated with geographic distance) were the best predictors of current gene flow in *B. carajensis* and *M. carajensis* populations, respectively. Considering the strong winds characterizing Montane Savanna ecosystems from the Carajás Mineral Province (Skirycz et al., 2014), and the fact that wind currents in open landscapes are known to facilitate long-distance dispersal of plant propagules (Heydel et al., 2014; Soons et al., 2004), we posit that wind-mediated dispersal is driving gene flow across Montane Savannas and open-pit mines. High levels of gene flow have also been detected in others wind-dispersed species in open anthropogenic landscapes like agricultural areas (Aavik et al., 2013; Heydel et al., 2014; Kamm et al., 2010), suggesting that open areas promote genetic connectivity in wind-dispersed plants. On the other hand, local climate differences also appear to explain gene flow patterns in *M. carajensis* populations from Serra Sul better than IBD, suggesting mismatches in flowering periods (Dick et al., 2008) or different local adaptations (Hoffmann and Sgrò, 2011; Lenormand, 2002). We nevertheless caution that our study design and the little available knowledge

on the natural history of these plants do not allow disentangling the relative contribution of pollen and seed dispersal on gene flow.

The absence of an effect of habitat loss on genetic variation can be attributed to time-lags between the onset of disturbances and genetic responses (Schlaepfer et al., 2018). We overcame this methodological limitation by focusing on species with a short generation time (i.e. completing their life cycle within one year), and by explicitly incorporating time scale into our analyses (evaluating land cover maps from different years). Moreover, historical demographic processes are unlikely to have biased our results, since our isolation by resistance models explicitly accounted for the underlying population structure. Mining operations began in the 1980s in Serra Norte, allowing enough time (~40 generations) to assess genetic responses to mining. On the other hand, Serra Sul was still pristine by 2013, so only three plant generations were exposed to mining before our samples were collected in 2017. This could explain why in Serra Sul recent land cover did not explain relatedness patterns better than geographic distance alone (Table S2). However, the fact that recent land cover did not explain relatedness patterns in either species in Serra Norte, strongly suggests that gene flow has been maintained across mines. In contrast, land cover in existence two decades ago was found to explain gene flow in a perennial narrow endemic morning glory occurring in Serra Norte (Lanes et al., 2018), indicating that our methods should be sufficient to detect an effect of mining should there be one, although differences in reproductive systems and dispersal modes could also underlie these different results (Aguilar et al., 2008; Vranckx et al., 2012). Additionally, our findings were unaffected by the resolution of spatial data and were supported by an independent barrier analysis, so they strongly indicate that gene flow in our two annual herbs is unaffected by habitat loss driven by mining.

The incidence of time-lag effects on the response of genetic diversity to habitat loss is nevertheless more difficult to assess, since different metrics respond at different rates. Empirical and simulations studies have shown changes in inbreeding and allelic richness immediately after the onset of disturbances, whereas heterozygosity is usually lost more slowly, over subsequent generations (Keyghobadi et al., 2005; Lowe et al., 2005). We therefore caution that longer time lags would be needed to rule out an effect of habitat loss on the observed heterozigosity of our study species. On the other hand, ~40 generations should be enough to detect a response in the levels of inbreeding, and confidence intervals for the effect of habitat loss on inbreeding (Table S4) suggest that our models had sufficient power to identify non-significant effects. Our results thus indicate, with moderate confidence, that habitat loss did not result in increased inbreeding in our study plants.

Conclusions

Using thousands of genetic markers to study two annual endemic plants in replicated landscapes, we found that extreme habitat loss driven by mining did not result in any detectable genetic consequences. Since our results are largely unbiased by the effect of habitat fragmentation, the underlying genetic structure of plant populations, the resolution of spatial data, or time-lag effects, they reveal that habitat loss does not always entail negative genetic consequences. Although habitat fragmentation has been shown to disrupt gene flow and increase inbreeding across plants species, regardless of their characteristics (Aguilar et al., 2019), our study unveils remarkably resilient species to extreme habitat loss as similar levels of genetic diversity and gene flow were found in mining and natural habitats. These findings imply that it is not possible to generalize about

the genetic consequences of habitat loss, so future conservation efforts need to consider species individually.

Acknowledgments

Funding was provided by Instituto Tecnológico Vale, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grants 301616/2017-5 (RJ), 307479/2016-1 and 402756/2018-5 (GO), and 300714/2017-3 (EL), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) grant 88887.156652/2017-00 (CSC). We thank Alexandre Castilho, Cesar Neto and Waléria Monteiro for assistance in the field, Gleiciane Salvador and Manoel Lopes for help in the laboratory, Prof. Ing. Jaroslay Dolezel for provinding the standards used in flow cytometry analyses, and Nathaniel Pope for advice on the statistical analyses. This manuscript has been released as a Pre-Print at bioRxiv (Carvalho et al., 2019).

Author contributions

RJ conceived, designed and coordinated the project. RJ and PLV coordinated the field work and sampling. ECML, ARS, CFC and MG performed laboratory work. RJ, ECML, CSC, ARS and CFC performed the data analysis. The first draft of the paper was written by CSC and ECML with input from RJ. All authors contributed to discussing the results and editing the paper.

Conflict of interest

Instituto Tecnológico Vale is a non-profit and independent research institute, and the choice of questions, study organisms and methodological approaches were exclusively defined by the authors, who declare no conflicting interests.

Data Availability Statement

Geographic coordinates in decimal degrees, genotypes in Variant Call Format and sequences in FASTA format for both species are provided here: https://figshare.com/articles/Habitat_loss_does_not_always_entail_negative_genetic_consequences/8224175. All the mentioned R scripts have been deposited in GitHub and their url addresses provided in the text.

References

Aavik, T., Holderegger, R., Edwards, P. J., and Billeter, R. (2013). Patterns of contemporary gene flow suggest low functional connectivity of grasslands in a fragmented agricultural landscape. *J. Appl. Ecol.* 50, 395–403. doi:10.1111/1365-2664.12053.

Aguilar, R., Cristóbal□Pérez, E. J., Balvino□Olvera, F. J., Aguilar□Aguilar, M., Aguirre□Acosta, N., Ashworth, L., et al. (2019). Habitat fragmentation reduces plant progeny quality: a global synthesis. *Ecol. Lett.* 0, ele.13272. doi:10.1111/ele.13272.

Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y., and Lobo, J. (2008). Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Mol. Ecol.* 17, 5177–88. doi:10.1111/j.1365-294X.2008.03971.x.

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109.

Allendorf, F. W., Luikart, G. H., and Atiken, S. N. (2013). *Conservation and the genetics of populations*. Second. Wiley-Blackwell.

Anderson, C. D., Epperson, B. K., Fortin, M.-J., Holdergger, R., James, P. M. A., Rosenberg, M. S., et al. (2010). Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol. Ecol.* 19, 3565–3575. doi:10.1111/j.1365-294X.2010.04757.x.

Balkenhol, N., Cushman, S., Storfer, A., and Waits, L. eds. (2016). *Landscape Genetics: Concepts, Methods, Applications*. Hoboken, NJ: John Wiley & Sons.

Barton, K., and Barton, M. K. (2015). Package ‘MuMIn’Version 1.

Burnham, K. K. P., and Anderson, D. R. D. (2002). *Model selection and multimodel inference: a practical information-theoretic approach (2nd ed)*.

Carvalho, C. da S., Lanes, E. C. M., Silva, A. R., Caldeira, C. F., Carvalho-Filho, N., Gastauer, M., et al. (2019). Habitat loss does not always entail negative genetic consequences. *bioRxiv*, 528430. doi:10.1101/528430.

Clarke, R. T., Rothery, P., and Raybould, A. F. (2002). Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *J. Agric. Biol. Environ. Stat.* 7, 361–372. doi:10.1198/108571102320.

Cruz, A. P. O., Viana, P. L., and Santos, J. U. (2016). Flora das cangas da Serra dos Carajás, Pará, Brasil: Asteraceae. *Rodriguesia* 67, 1211–1242. doi:10.1590/2175-7860201869118.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330.

Dick, C. W., Hardy, O. J., Jones, F. A., and Petit, R. J. (2008). Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. *Trop. Plant Biol.* 1, 20–33. doi:10.1007/s12042-007-9006-6.

DiLeo, M. F., and Wagner, H. H. (2016). A landscape ecologist's agenda for landscape genetics. *Curr. Landsc. Ecol. Reports* 1, 115–126. doi:10.1007/s40823-016-0013-x.

Dolezel, J., Sgorbati, S., and Lucretti, S. (1992). Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol. Plant.* 85, 625–631. doi:10.1111/j.1399-3054.1992.tb04764.x.

Doyle, J., and Doyle, J. L. (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull* 19, 11–15.

Dyer, R. J. (2016). "Landscapes and Plant Population Genetics," in *Landscape Genetics: Concepts, Methods, Applications*, eds. N. Balkenhol, S. Cushman, A. Storfer, and L. Waits (Hoboken, NJ: Wiley Online Library), 181–198.

Esfeld, K., Hensen, I., Wesche, K., Jakob, S. S., Tischew, S., and Blattner, F. R. (2008). Molecular data indicate multiple independent colonizations of former lignite mining areas in Eastern Germany by Epipactis palustris (Orchidaceae). *Biodivers. Conserv.* 17, 2441–2453. doi:10.1007/s10531-008-9391-7.

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annu. Rev. Ecol. Evol. Syst.* 34, 487–515. doi:10.1146/annurev.ecolsys.34.011802.132419.

François, O., Martins, H., Caye, K., and Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Mol. Ecol.* 25, 454–469. doi:10.1111/mec.13513.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196, 973–983. doi:10.1534/genetics.113.160572.

Greenleaf, S. S., Williams, N. M., Winfree, R., and Kremen, C. (2007). Bee foraging ranges and their relationship to body size. *Oecologia* 153, 589–596.

Heydel, F., Cunze, S., Bernhardt-Römermann, M., and Tackenberg, O. (2014). Long-distance seed dispersal by wind: disentangling the effects of species traits, vegetation types, vertical turbulence and wind speed. *Ecol. Res.* 29, 641–651. doi:10.1007/s11284-014-1142-5.

Hoenig, J. M., and Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am. Stat.* 55, 19–24.

Hoffmann, A. A., and Sgrò, C. M. (2011). Climate change and evolutionary adaptation. *Nature* 470, 479–485. Available at: https://doi.org/10.1038/nature09670.

Honnay, O., Bossuyt, B., Jacquemyn, H., Shimono, A., and Uchiyama, K. (2008). Can a seed bank maintain the genetic variation in the above ground plant population? *Oikos* 117, 1–5. doi:10.1111/j.2007.0030-1299.16188.x.

Jaffé, R., Pope, N., Acosta, A. L., Alves, D. A., Arias, M. C., De la Rúa, P., et al. (2016). Beekeeping practices and geographic distance, not land use, drive gene flow across tropical bees. *Mol. Ecol.* 25, 5345–5358. doi:10.1111/mec.13852.

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi:10.1093/bioinformatics/btr521.

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11, 94. doi:10.1186/1471-2156-11-94.

Kamm, U., Gugerli, F., Rotach, P., Edwards, P., and Holderegger, R. (2010). Open areas in a landscape enhance pollen-mediated gene flow of a tree species: evidence from northern Switzerland. *Landsc. Ecol.* 25, 903–911.

Keyghobadi, N., Roland, J., Matter, S. F., and Strobeck, C. (2005). Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. *Proc. R. Soc. B Biol. Sci.* 272, 553–560. doi:10.1098/rspb.2004.2976.

Khimoun, A., Peterman, W., Eraud, C., Faivre, B., Navarro, N., and Garnier, S. (2017). Landscape genetic analyses reveal fine-scale effects of forest fragmentation in an insular tropical bird. *Mol. Ecol.* 26, 4906–4919. doi:10.1111/mec.14233.

Lanes, É. C., Pope, N. S., Alves, R., Carvalho Filho, N. M., Giannini, T. C., Giulietti, A. M., et al. (2018). Landscape genomic conservation assessment of a narrow-endemic and a widespread morning glory from Amazonian Savannas. *Front. Plant Sci.* 9, 532. doi:10.3389/fpls.2018.00532.

Leimu, R., Mutikainen, P., Koricheva, J., and Fischer, M. (2006). How general are positive relationships between plant population size, fitness and genetic variation? *J. Ecol.* 94, 942–952. doi:10.1111/j.1365-2745.2006.01150.x.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends Ecol. Evol.* 17, 183–189. doi:https://doi.org/10.1016/S0169-5347(02)02497-7.

Llorens, T. M., Ayre, D. J., and Whelan, R. J. (2018). Anthropogenic fragmentation may not alter pre-existing patterns of genetic diversity and differentiation in perennial shrubs. *Mol. Ecol.* 27, 1541–1555. doi:10.1111/mec.14552.

Loureiro, J., Rodriguez, E., Dolezel, J., and Santos, C. (2007). Two new nuclear isolation buffers for plant DNA flow cytometry: A test with 37 species. *Ann. Bot.* 100, 875–888. doi:10.1093/aob/mcm152.

Lowe, A. J., Boshier, D., Ward, M., Bacles, C. F. E., and Navarro, C. (2005). Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity (Edinb).* 95, 255–273. doi:10.1038/sj.hdy.6800725.

McCartney-Melstad, E., Vu, J. K., and Shaffer, H. B. (2018). Genomic data recover previously undetectable fragmentation effects in an endangered amphibian. *Mol. Ecol.* 27, 4430–4443. doi:10.1111/mec.14892.

McRae, B. H. (2006). Isolation by Resistance. *Evolution (N. Y).* 60, 1551. doi:10.1554/05-321.1.

Michaels, S. D., John, M. C., and Amasino, R. M. (1994). Removal of polysaccharides from plant DNA by ethanol precipitation. *Biotechniques* 17, 274–276.

Monteiro, W. P., Veiga, J. C., Silva, A. R., Carvalho, C. da S., Lanes, É. C. M., Rico, Y., et al. (2019). Everything you always wanted to know about gene flow in tropical landscapes (but were afraid to ask). *PeerJ* 7, e6446. doi:10.7717/peerj.6446.

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., and Portnoy, D. S. (2018). These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Mol. Ecol.* 27, 3193–3206. doi:10.1111/mec.14792.

Peterman, W. E. (2018). ResistanceGA : An R package for the optimization of resistance surfaces using genetic algorithms. *Methods Ecol. Evol.* 9, 1638–1647. doi:10.1111/2041-210X.12984.

Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. (2009). Linear and nonlinear mixed effects models. *R Packag. version*. Available at: http://ftp.auckland.ac.nz/software/CRAN/doc/packages/nlme.pdf.

Rocha, K. C. de J., Goldenberg, R., Meirelles, J., and Viana, P. L. (2017). Flora das cangas da Serra dos Carajás, Pará, Brasil: Melastomataceae. *Rodriguesia* 68, 997–1034. doi:10.1590/2175-7860201869118.

Rogstad, S. H. (1992). Saturated NaCl-CTAB solution as a means of field preservation of leaves for DNA analyses. *Taxon* 41, 701. doi:10.2307/1222395.

Russello, M. A., Waterhouse, M. D., Etter, P. D., and Johnson, E. A. (2015). From promise to practice: pairing non-invasive sampling with genomics in conservation. *PeerJ* 3, e1106. doi:10.7717/peerj.1106.

Schlaepfer, D. R., Braschler, B., Rusterholz, H.-P., and Baur, B. (2018). Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. *Ecosphere* 9, e02488. doi:10.1002/ecs2.2488.

Shirk, A. J., Landguth, E. L., and Cushman, S. A. (2017). A comparison of individual-based genetic distance metrics for landscape genetics. *Mol. Ecol. Resour.* 17, 1308–1317. doi:10.1111/1755-0998.12684.

Skirycz, A., Castilho, A., Chaparro, C., Carvalho, N., Tzotzos, G., and Siqueira, J. O. (2014). Canga biodiversity, a matter of mining. *Front. Plant Sci.* 5, 653. doi:10.3389/fpls.2014.00653.

Soons, M. B., Heil, G. W., Nathan, R., and Katul, G. G. (2004). Determinants of long-distance seed dispersal by wind in grasslands. *Ecology* 85, 3056–3068. doi:10.1890/03-0522.

Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., and Waits, L. P. (2010). Landscape genetics: where are we now? *Mol. Ecol.* 19, 3496–3514. doi:10.1111/j.1365-294X.2010.04691.x.

Tarpley, J. D., Schneider, S. R., and Money, R. L. (1984). Global vegetation indices from the NOAA-7 meteorological satellite. *J. Clim. Appl. Meteorol.* 23, 491–494. doi:10.1175/1520-0450(1984)023<0491:GVIFTN>2.0.CO;2.

Tittensor, D. P., Walpole, M., Hill, S. L. L., Boyce, D. G., Britten, G. L., Burgess, N. D., et al. (2014). A mid-term analysis of progress toward international biodiversity targets. *Science (80-.).* 346, 241–244. doi:10.1126/science.1257484.

Viana, P. L., Mota, N. F. D. O., Gil, A. D. S. B., Salino, A., Zappi, D. C., Harley, R. M., et al. (2016). Flora das cangas da Serra dos Carajás, Pará, Brasil: História, área de estudos e metodologia. *Rodriguesia* 67, 1107–1124. doi:10.1590/2175-7860201667501.

Vranckx, G., Jacquemyn, H., Muys, B., and Honnay, O. (2012). Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. *Conserv. Biol.* 26, 228–237. doi:10.1111/j.1523-1739.2011.01778.x.

Wang, J. (2017). Estimating pairwise relatedness in a small sample of individuals. *Heredity (Edinb).* 119, 302–313. doi:10.1038/hdy.2017.52.

Waterhouse, M. D., Blair, C., Larsen, K. W., and Russello, M. A. (2017). Genetic variation and fine-scale population structure in American pikas across a human-modified landscape. *Conserv. Genet.* 18, 825–835. doi:10.1007/s10592-017-0930-1.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569. doi:10.1038/ng.608.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. a, Smith, G. M., and Ebooks Corporation. (2009). *Mixed effects models and extensions in ecology with R*. doi:10.1007/978-0-387-87458-6.

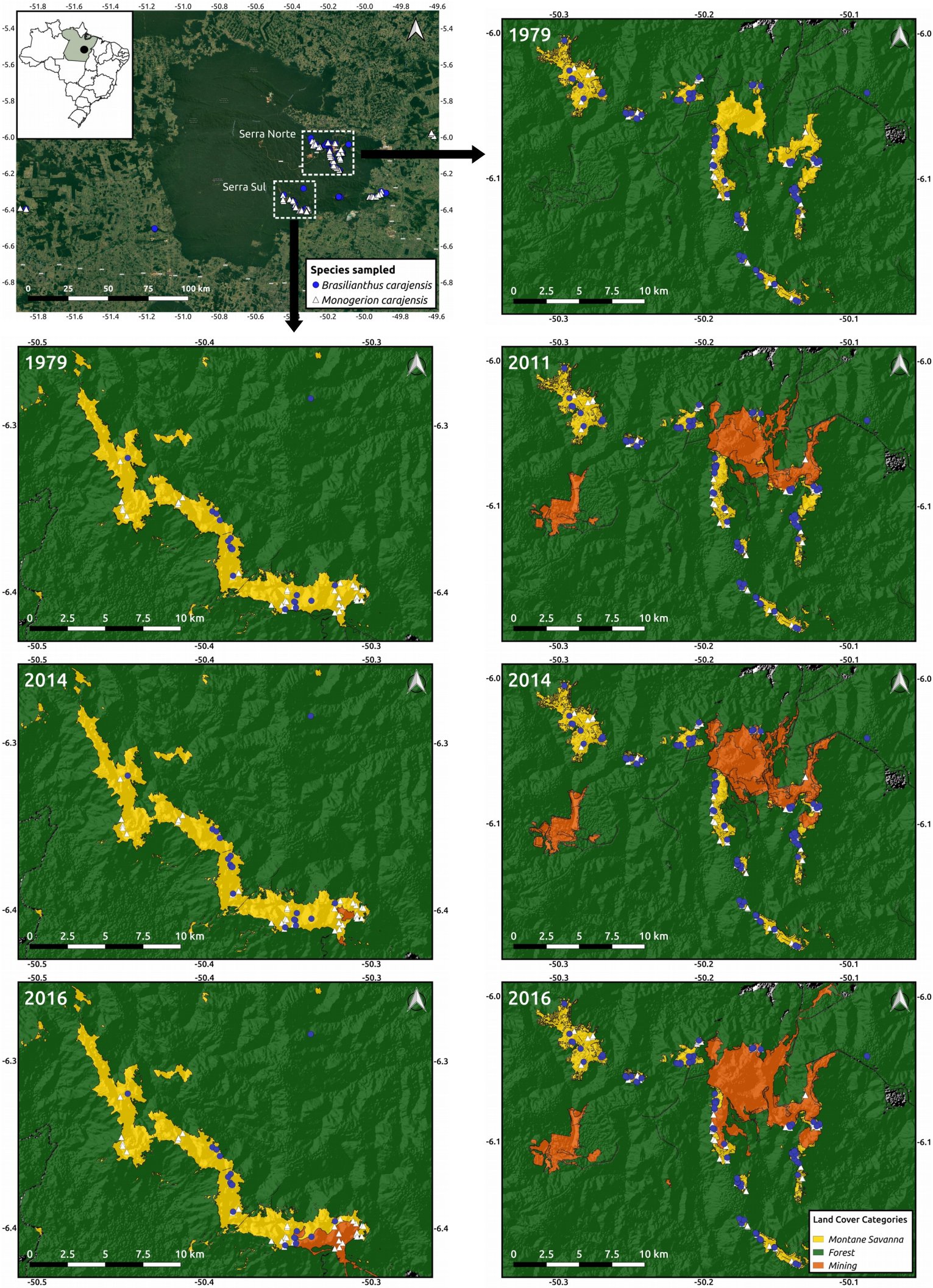
Figure legends

Fig. 1. Map of the study region depicting the location of the collected samples from *Brasilianthus carajensis* (blue circles) and *Monogereion carajensis* (white triangles) in Serra Norte (right panels) and Serra Sul (left panels). Hill shade maps are shown overlaid with land cover color maps for the different years analyzed. The location of the Carajás Mineral Province within Brazil is shown on the upper left corner.

Fig. 2. Map showing the ancestry coefficients from *Brasilianthus carajensis* (A and C) and *Monogereion carajensis* (B and D) in Serra Norte (upper panels) and Serra Sul (lower panels) determined using the Admixture software. Montane Savanna areas are shown in green against hill shade layers. Smaller lower-left corner plots show spatial autocorrelation in genetic relatedness, where black solid lines are the LOESS fit to the observed relatedness, gray shaded regions are 95% confidence bounds around the null expectation (black dotted lines), and short vertical lines at the bottom of the figure are observed pairwise distances. Genetic diversity measures for each genetic cluster are shown in the upper tables. Sample sizes (N) are followed by mean expected heterozygosity (H_E) and mean inbreeding coefficient (F), and values represent 95% confidence intervals.

Fig. 3. Relative variable importance in the set of best-fitting models (ΔAIC $\leq$ 2) for *Brasilianthus carajensis* and *Monogereion carajensis* in Serra Norte and Serra Sul (see methods and Supplementary Tables S3 and S5 for details). Individual-level genetic diversity metrics (H and f) were response variables and habitat amount in 1979 and habitat loss in 2011, 2014 and 2016 were predictors in genetic diversity models. Pairwise inter-individual genetic relatedness was the response variable and resistance distances computed from optimized surfaces were predictors in IBR models. Likelihood Ratio Test (LRT) were performed to assess if each predictor variable significantly improved the model's log-likelihood (significance levels are highlighted with: *$p < 0.05$; **$p < 0.01$; and ***$p < 0.001$).

Fig. 4. Coefficient plots for the set of best-fitting models (ΔAIC $\leq$ 2) for *Brasilianthus carajensis* and *Monogereion carajensis* in Serra Norte and Serra Sul (see methods and Supplementary Tables S4 and S6 for details). Points represent model-averaged regression coefficients and lines the 95% confidence intervals.

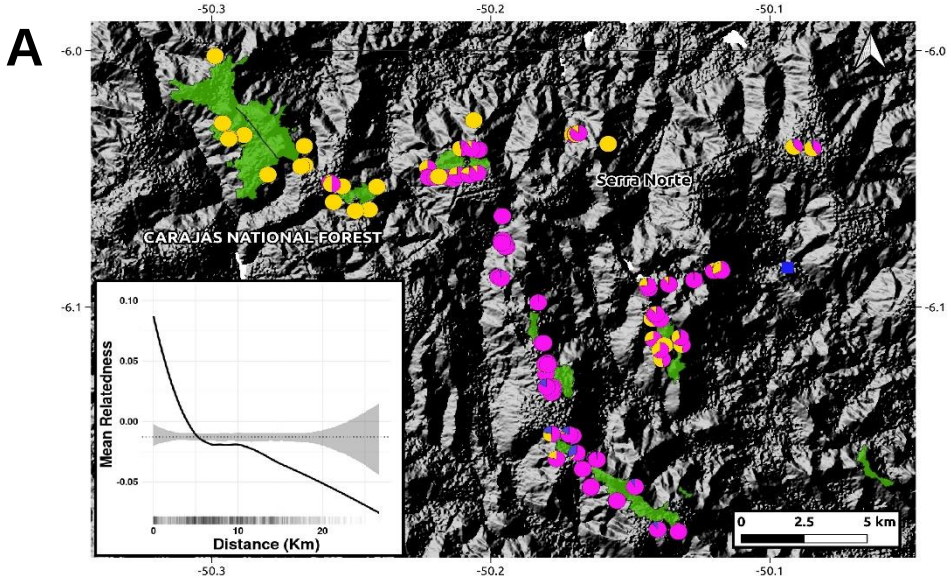

Serra Norte
Serra Sul
Species sampled
Brasilianthus carajensis
Monogerion carajensis
0 25 50 75 100 km
1979
2011
2014
2016
0 2.5 5 7.5 10 km
Land Cover Categories
Montane Savanna
Forest
Mining

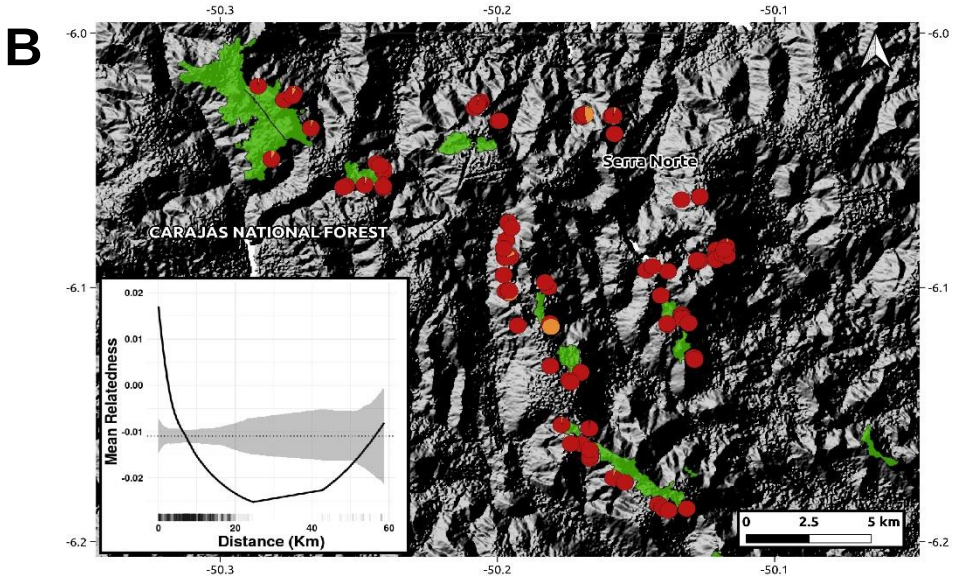
Brasilianthus carajensis

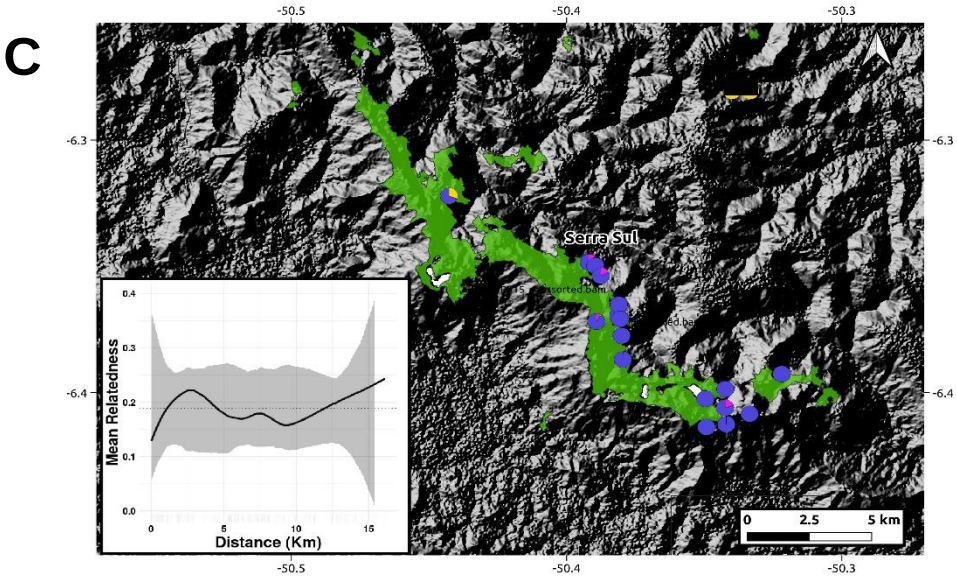
	N	H_E	*F*
Cluster I	28	[0.22/0.23]	[-0.03/0.11]
Cluster II	85	[0.21/0.21]	[-0.07/0.02]
Cluster III	37	[0.23/0.24]	[0.02/0.16]

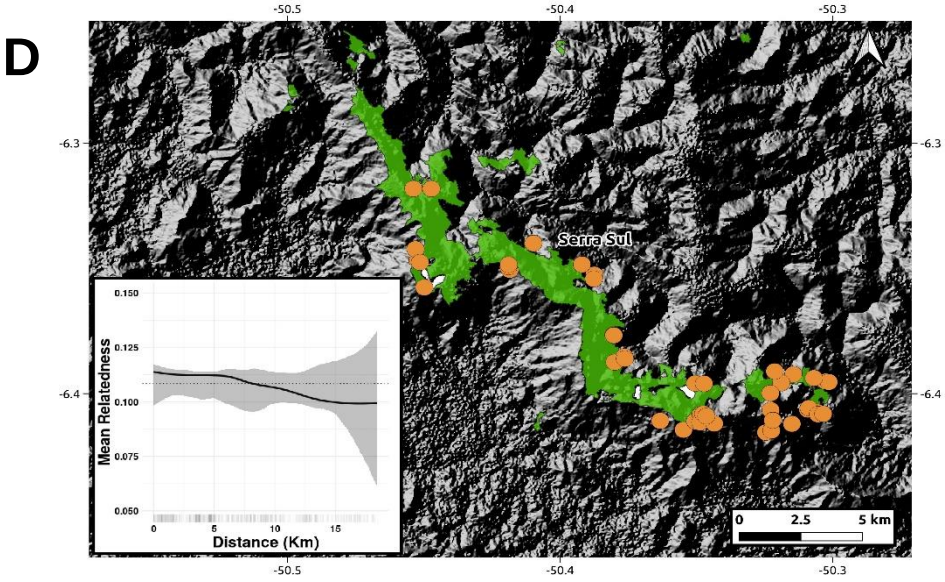
Monogereion carajensis

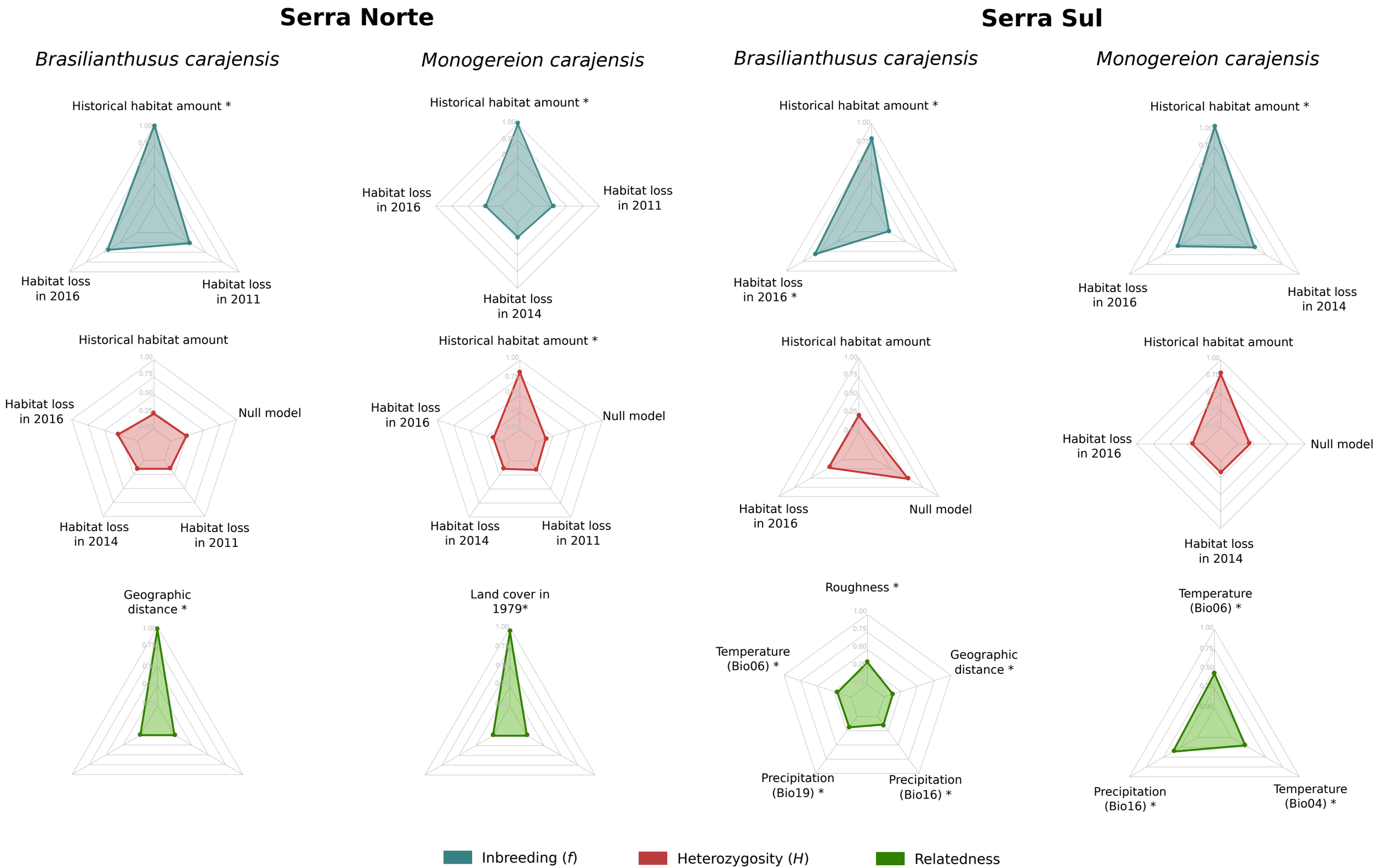
	N	H_E	*F*
Cluster I	116	[0.18/0.18]	[-0.02/0.01]
Cluster II	91	[0.16/0.16]	[0.003/0.06]

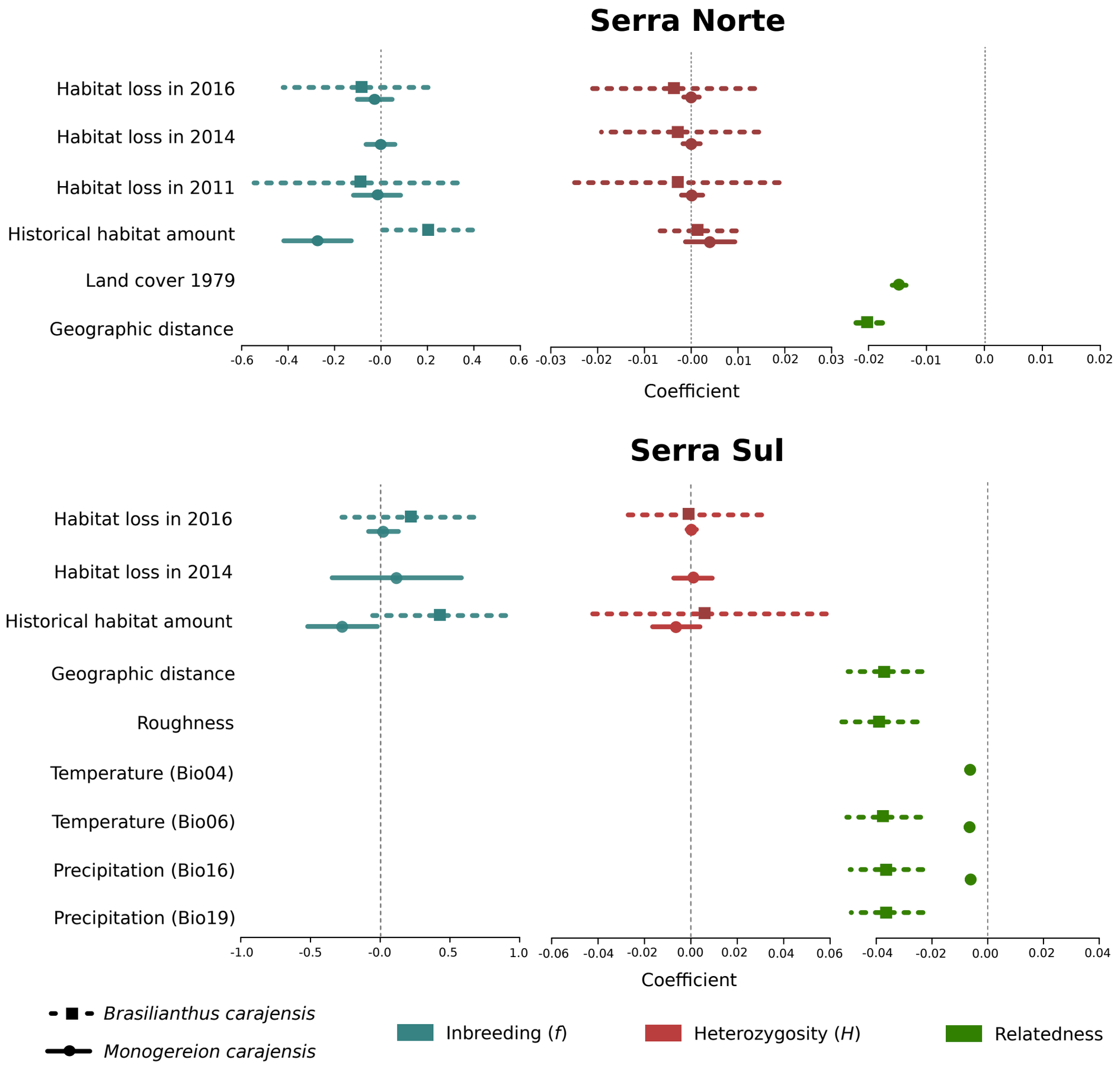









Serra Norte
Brasilianthusus carajensis
Monogereion carajensis
Serra Sul
Brasilianthusus carajensis
Monogereion carajensis
Historical habitat amount *
Habitat loss in 2016
Habitat loss in 2011
Habitat loss in 2014
Habitat loss in 2016 *
Historical habitat amount
Null model
Geographic distance *
Land cover in 1979*
Roughness *
Temperature (Bio06) *
Precipitation (Bio19) *
Precipitation (Bio16) *
Temperature (Bio04) *
Inbreeding (f)
Heterozygosity (H)
Relatedness


Serra Norte
Habitat loss in 2016
Habitat loss in 2014
Habitat loss in 2011
Historical habitat amount
Land cover 1979
Geographic distance
-0.6 -0.4 -0.2 -0.0 0.2 0.4 0.6
-0.03 -0.02 -0.01 -0.00 0.01 0.02 0.03
-0.02 -0.01 0.0 0.01 0.02
Coefficient
Serra Sul
Habitat loss in 2016
Habitat loss in 2014
Historical habitat amount
Geographic distance
Roughness
Temperature (Bio04)
Temperature (Bio06)
Precipitation (Bio16)
Precipitation (Bio19)
-1.0 -0.5 -0.0 0.5 1.0
-0.06 -0.04 -0.02 0.00 0.02 0.04 0.06
-0.04 -0.02 0.00 0.02 0.04
Coefficient
Brasilianthus carajensis
Monogereion carajensis
Inbreeding (f)
Heterozygosity (H)
Relatedness