

1 Kelly Souza^{1*}; Jesús N. Pinto-Ledezma²; Mariana Pires de Campos Telles³; Thannya Nascimento
2 Soares⁴; Lazaro José Chaves⁵ Clarissa Bonafé Gaspar Ruas⁶; Ricardo Dobrovolski⁷; José Alexandre
3 Felizola Diniz-Filho⁸

4

5 **HOW TO MEASURE THE INFLUENCE OF LANDSCAPE ON POPULATION GENETIC**
6 **STRUCTURE: DEVELOPING RESISTANCE SURFACES USING A PATTERN-ORIENTED**
7 **MODELING APPROACH**

8 1. Programa de Pós-Graduação em Genética, PGBM, ICB, Universidade Federal de Goiás. Campus II
9 (Samambaia), Goiânia, GO, Brasil.

10 2. Department of Ecology, Evolution and Behavior, University of Minnesota, 1479 Gortner Ave, Saint
11 Paul, MN 55108, USA.

12 3. Departamento de Genética, Universidade Federal de Goiás. Campus II (Samambaia), Goiânia, GO,
13 Brasil and Escola de Ciências Agrárias e Biologia, PUC-GO, Av. Engler, 286-316 - Jardim Mariliza,
14 Goiânia, GO, Brasil.

15 4. Departamento de Genética, Universidade Federal de Goiás. Campus II (Samambaia), Goiânia, GO,
16 Brasil.

17 5. Escola de Agronomia, Universidade Federal de Goiás. Campus II (Samambaia), Goiânia, GO, Brasil.

18 6. Graduanda em Ciências da Computação, Universidade Estadual Paulista, Rio Claro, SP, Brasil.

19 7. Departamento de Zoologia, Universidade Federal da Bahia, Salvador, BA, Brasil.

20 8. Departamento de Ecologia ICB, Universidade Federal de Goiás, Goiânia, GO, Brasil.

21

22 ***Corresponding Author:** Tel +55 62996000910

23 Email: kellysouzaecol@gmail.com

24

25

26 **Abstract**

27 There are several approaches to understand how a landscape, with its several components, affects the
28 genetic population structure by imposing resistance to gene flow. Here we propose the creation of
29 resistance surfaces using a Pattern-Oriented Modeling approach to explain genetic differentiation,
30 estimated by pairwise FST, among “Baruzeiro” populations (*Dipteryx alata*), a tree species widely
31 distributed in Brazilian Cerrado. To establish the resistance surface, we used land use layers from the area
32 in which the 25 “Baruzeiro” populations were sampled, generating 10000 resistance surfaces. To establish
33 the resistance surface, we used land use layers from the area in which the 25 “Baru” populations were
34 sampled, generating 10000 resistance surfaces. We randomized the cost values for each landscape
35 component between 0 and 100. We use these surfaces to calculate pairwise matrices of the effective
36 resistance among populations. Mantel test revealed a correlation of pairwise FST with a geographical
37 distance equal to $r = 0.48$ ($P < 0.001$), whereas the Mantel correlations between pairwise FST and the
38 generated resistance matrices ranged between $r = -0.2019$ and $r = 0.6736$. Partial regression on distance
39 matrices was used to select the resistance matrix that provided the highest correlation with pairwise FST,
40 based on the AIC criterion. The selected models suggest that the areas with lower resistance are
41 characterized as natural savanna habitats of different forms, mainly arboreal dense savannas. In contrast,
42 roads, big rivers, and agricultural lands cause higher resistance to gene flow.

43 **Keywords:** gene flow, FST, *Dipteryx alata*, circuitscape, mantel

44

45 **Declarations**

46 **Funding:** This manuscript derives from the Geographic Genetics and Natural Resource Conservation
47 Project in Brazilian Cerrado (GENPAC), supported by CNPq. Work by K. S. S. is support by a FAPEG
48 Doctoral Fellowship (proc. 88887.162750/2018-00), work by J.A.F.D.-F. is supported by CNPq
49 productivity fellowships and by the Brazilian Research Network on Climate Change (CNPq No.
50 550022/2014 and FINEP No. 01.13.0353.00). This manuscript is developed in the context of the National
51 Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation,
52 supported by MCTIC/CNPq (proc. 465610/2014-5) and FAPEG (proc. 201810267000023)

53 **Conflicts of interest/Competing interests:** The authors declare that they have no conflict of interest.

54 **Availability of data and material:** The data used in work will be made available after approval of the
55 article by Journal

56 **Code availability:** The custom code created and used in the development of the work will be made
57 available after the Journal approved the manuscript.

58 **Authors' contributions:** All authors contributed to the study conception and design. Material
59 preparation, data collection were performed by Kelly Souza; Mariana Pires de Campos Telles; Thannya
60 Nascimento Soares; Lazaro José Chaves. Analysis was performed by Kelly Souza; Jesús N. Pinto-
61 Ledezma; Clarissa Bonafé Gaspar Ruas; Ricardo Dobrovolski; José Alexandre Felizola Diniz-Filho.
62 Kelly Souza wrote the first draft of the manuscript, and all authors commented on previous versions of
63 the manuscript. All authors read and approved the final manuscript.

64 **Introduction**

65 Landscape structure can influence ecological processes, such as dispersal, migration and flow gene, at
66 different levels of biological organization and spatial scales, including those driving the population
67 genetic structure (Manel et al. 2003; Koffi et al. 2007). For example, a strong genetic structure (i.e., high
68 differentiation among demes or local populations) appears when habitat loss decreased connectivity and,
69 therefore, the dispersal capacity, dividing populations and disrupting gene flow (King and With 2002,
70 Amos et al., 2012; Braga et al. 2019). This decrease in gene flow results in loss of genetic variation and
71 inbreeding depression, most likely increase the probability of local extinctions (Storfer 1999), changing
72 aspects of species' life history (Kramer et al. 2008) and eventually reducing its evolutionary potential
73 (Frankham et al. 2004). The analysis of all these processes is the primary goal of a field that quickly
74 developed and advanced in the last 15 years called "Landscape Genetics" (Manel et al. 2003; Holderegger
75 and Wagner 2006; Storfer et al. 2007; Manel and Holderegger 2013). Landscape genetics seeks to
76 evaluate the interaction between landscape features and microevolutionary processes such as gene flow,
77 selection, and genetic drift, integrating, thus geographical, ecological, and genetic information (Manel et
78 al. 2003; Storfer et al. 2007).

79 In landscape genetics, it is possible to check and interpret the effective distance between
80 individuals or populations, taking into account landscape properties that wild better reflect gene flow
81 (Mateo-Sánchez et al. 2015). This relation can be calculated through the "Isolation by Environment"
82 (IBE) models, which describes a pattern in which genetic differentiation increases with environmental
83 differences, independent of geographical distances (Sexton et al. 2014, Jenkins et al. 2010, Wang and
84 Bradburd 2014). Another way to check the effective distance in landscape genetics is by Isolation by
85 Resistance – IBR (McRae 2006), where the distance calculation incorporates the degree of "permeability"
86 of the different landscape components (e.g., forests, croplands, roads) to the dispersion of individuals
87 throughout the landscape. Therefore, this permeability is related to how the landscape affects the
88 movement of organisms between the areas with resources, in terms of biological, physiological, and
89 behavioral characteristics, thus controlling the natural flows of species (Metzger and Ddcamps 1997;
90 Tischendorf and Fahrig 2000). IBE and IBR researchers tend to concentrate on describing patterns,
91 without necessarily investigating the mechanisms that have generated these patterns (Wang and Bradburd
92 2014). However, all these models are generalized versions of the much older (and simpler) Isolation-by-
93 Distance (IBD) proposed by Sewall Wright in the early 1940's, in which it is possible to predict an

94 exponential decrease of genetic distance as geographic distances increase, by a balance between dispersal
95 and local genetic drift (see Wright 1943). So, both IBR and IBE can be viewed as more complex cases of
96 IBD in terms of dispersal routes and changes in a balance due to landscape features.

97 Resistance surfaces have been used to understand how landscape components influence the
98 connectivity among species populations. (Spear et al. 2010; Koen et al. 2012). These surfaces are
99 representations of the degree of connectivity that is attributed to the original landscape components (i.e.,
100 considering the organisms of interest) that are used to model their movement through the landscape
101 (Spear *et al.* 2010; Taylor et al.,1993; Coulon et al., 2004; Vignieri, 2005). A crucial step in the
102 development of these surfaces is to attribute values, or costs, to each of the landscape components. Such
103 parameterization will determine how users will be this resistance to model species movement throughout
104 the landscape (Spear et al. 2010; Koen et al. 2012). However, this attribution of costs is, in many cases,
105 subjective and is not based on strict knowledge of species' traits.

106 To describe the resistance that the landscape imposes to the gene flow between populations and
107 to reveal information on the processes behind the population's genetic structure observed patterns, it is
108 possible to use a Pattern-Oriented Modeling technique (Grimm 1994; Grimm et al. 1996; Grimm et al.
109 2005; Diniz-Filho et al. 2014). The POM provides a conceptual framework to assess the applicability of
110 models by comparing the patterns generated by the model to observed patterns (Kang e Aldstadt, 2018).
111 One can use computational procedures to create a conceptual framework and find the set of parameters
112 that generates the best models replicating an empirical pattern. The creation of this models resulted in the
113 improvement of the quality of the model and the overall understanding of the system (Kang e Aldstadt,
114 2018; Diniz-Filho et al. 2014), which allows a biological and ecological interpretation of this best set of
115 parameter values (Wiegand et al. 2003).

116 The system we are interested in refers to the landscape influence in the genetic structure of
117 *Dipteryx alata* populations in Central Brazil. Assuming that genetic diversity has a positive relationship
118 with the resistance landscape (McRae 2006), we used here genetic diversity to attribute the values of
119 resistance of the different landscape components to build resistance surfaces that better explain the
120 genetic structure pattern among populations. The main goal is to understand which landscape features (or
121 combination of them) better define the genetic divergence between populations. This statistical definition
122 minimizes the arbitrariness in the parameterization of these surfaces, so we used the POM approach in the

123 search for matches between simulated and observed patterns (Diniz-Filho et al. 2014). Several lines of
124 evidence suggest that anthropogenic features affect the connectivity and the gene flow in natural
125 landscapes (Pérez-Espona 2008, Ayran et al. 2017; Okamiya and Kusano 2019). Thus, we expect that
126 anthropogenic features will be selected for the POM and present a high resistance for the gene flow. We
127 also hope that landscape features, especially those of the species natural habitat and their dispersers, show
128 less resistance to the gene flow.

129

130 **Materials and Methods**

131 For our analyses, we sampled *Dipteryx alata* Vog in 25 localities (local populations hereafter), a tree
132 species widely distributed in the Brazilian Cerrado, popularly known as "Baru" tree or "Baruzeiro." We
133 estimated the genetic variation of 644 individuals collected, with sample sizes ranging from 12 to 37
134 individuals in each local population, covering most of the geographical distribution of the species (Table
135 S1), as detailed described elsewhere (especially Soares et al. 2012; Diniz-Filho et al. 2012; Collevatti et
136 al. 2013). Seven microsatellite loci were used to estimate the genetic divergence between pairs of
137 populations using Wright's F_{ST} , calculated using the *pp.fst* function in the *hierfstat* package (Goudet
138 2005), with modifications on R platform (R Core Team 2014).

139 We define the study area with a polygon covering all populations, marginally buffered by 20
140 kilometers around the local populations, covering an area of approximately 1.25×10^6 km². We used the
141 contemporary habitat configuration to define landscape elements to build the resistance surfaces. The
142 overall reasoning is to determine the influence of a landscape modified by changes in the Cerrado region
143 over the last 50 years due to agricultural expansion and population growth in the region (Klink and
144 Machado 2005; Klink 2013). Initially, we adopted land use layers with land cover classes (hydrography,
145 roads, urban areas, and vegetation types) obtained from the Land Cover Maps Biomes, on a spatial scale
146 of 1:250.000, based on Landsat 7 ETM + with pictures of 2002 and Brazilian roads map (see
147 www.mma.gov.br and www.dnit.gov.br). We reclassified the raster maps with approximately 0.02
148 decimal degrees of resolution into distinct natural (vegetation types of forests, such as savannas,
149 grasslands, and drainage network) or anthropic classes (agricultural areas, urban areas, and roads), (see
150 Table 2 for a description) and calculated the percentage of each class in the landscape. All this
151 information (landscape components) was gathered into a single layer and used to calculate the resistance

152 surfaces. For our analysis, when a cell has two or more classes, the class with higher occurrence was
153 selected to describe the cell, using ArcGIS 9.3 software (ESRI, 2011).

154

155 **Landscape Analysis and Pattern-Oriented Modeling (POM)**

156 The landscape resistance to the gene flow of a species is caused by the interaction between its
157 biological traits for dispersion (or their dispersers) and all land cover classes structured in a resistance
158 surface. The resistance surface is the numerical demonstration of the quantity of resistance imposed on
159 the gene flow among the populations on the study. One of the major challenges in Landscape Genetics is
160 the assignment of resistance values to this landscape components because its interaction (and usually
161 primary biological data) is unknown (Spear et al. 2010; Stevenson-Holt et al. 2014). Each landscape
162 component is a numeric variable, difficult to express in terms of real resistance values, mainly due to lack
163 of knowledge of dispersion forms, and consequently, in the gene flow.

164 To evaluate the potential resistance that the landscape imposes on gene flow among Baru
165 populations, and to reveal information on the values of the parameters and the processes behind the
166 pattern of genetic divergence between populations. We used Pattern-Oriented Modeling (Grimm et al.
167 1996; Grimm et al. 2005; Grimm and Railsback 2012; Topping et al. 2012), an approach based on the
168 genetic divergence between “Baru” populations. Pattern-Oriented Modeling (POM) allows finding which
169 combination of parameters maximize the correlation between genetic and dispersal, and subsequently, the
170 selection of the “best” resistance surface under this relationship.

171 To build the resistance surfaces, resistance values in the interval between 0 and 100, were
172 randomly assigned to each landscape components, generating 10.000 surfaces with different values of
173 resistance for each class. Despite the randomization of resistance values, the landscape configuration
174 (classes) has not changed, remaining constant in all generated surfaces (Figure 2).

175 We started the modeling with a “null” model, where all cells on the resistance surface were
176 zero, equivalent to considering only the Euclidean distance between pairwise populations. Next, we
177 created more complex landscape surfaces incorporating heterogeneity, with randomized values of
178 resistance in each of its components and comparing them with the null model. In a third cycle, we
179 systematically removed the complexity of the surfaces in the selected models. We evaluated the

180 importance of each class and its resistance in the model and validation. Our work emphasizes the second
181 step of the POM, the estimation of parameters in the creation of resistance surfaces. Our most significant
182 interest is to quantify the resistance imposed by each landscape component to the dispersion and,
183 consequently, to the genetic diversity and gene flow between populations.

184 We generate 10.000 resistant surfaces and calculate the effective resistance matrices for pairs
185 of populations according to the Electrical Circuit Theory (McRae 2006; McRae and Beier 2008; McRae et
186 al. 2008). Differently from the more commonly used least-cost path, which indicates a single possible
187 connection path among populations, the effective resistance calculation is based on the Electrical Circuit
188 Theory, allowing calculating the resistance distances considering multiple pathways (McRae et al. 2008).
189 The resistance calculation based on the Electrical Circuit Theory was generated through Circuitscape
190 software (v.4.0 Beta-; McRae, 2006). We used the ResistanceGA package in R (Peterman 2014) to carry
191 out the analysis in Circuitscape.

192

193 **Spatial Analysis**

194 We correlated the spatial arrangements (in terms of resistance distance between populations) with the F_{ST}
195 pairwise using different Mantel tests (see Diniz-Filho et al. 2013, for a recent review). The significance of
196 Pearson matrix correlations was estimated from 999 permutations.

197 To check the influence of landscape resistance on the species' gene flow, maintaining the
198 effect of geographic distance between populations permanent, we also used a partial Mantel test
199 (Legendre and Legendre 2012), calculated using R software (R Core Team 2014) with *mantel.partial*
200 function of the *vegan* package (Oksanen et al. 2012). The Akaike information criterion (AIC) was used to
201 select the best (most satisfactory) models (deLeeuw 1992), which were considered those with ΔAIC
202 values below three. The ΔAIC values were also used to calculate the AIC weight in each model (Akaike
203 weight - W_i) (Burnham and Anderson 2002; Diniz-Filho et al., 2008), revealing the level of certainty in
204 achieving the best explanatory model against all the other tested models.

205 Each landscape component has an individual contribution to the explanatory power of models.
206 To assess this individual contribution of each landscape component in all select models (equation 1),

$$207 \quad \sum_{i=1}^n X_i * W_i \quad [1]$$

208 we multiplied (w_i) by the resistance values of each landscape component (X_i) in the selected
209 models (n). With this approach, we were able to measure the influence of each landscape component on
210 gene flow, translated in genetic distances observed.

211 We then used the mean resistance of selected models to create a final current map to verify
212 possible species dispersion routes and identify the areas that most contribute to connectivity between
213 populations (as seen in Castilho et al., 2011; McRae, 2006; McRae et al., 2008). The value of each cell in
214 the landscape represents an amount of current flowing through it, analogous to the percolation of the
215 species (McRae et al. 2008). Where the landscape is less resistant, the routes of dispersion are more likely
216 to occur.

217

218 **Results**

219 We found a significant amount of genetic divergence between populations, with an average F_{ST} of 0.258
220 to establish a comparison among all the populations. The genetic divergence based on F_{ST} values was
221 positively correlated with the geographic distance (Mantel test, $r = 0.4805$; $P < 0.001$), showing a linear
222 relationship between genetic and geographic distances in which genetic divergence between populations
223 (inverse of F_{ST}) increase with geographic distance increase (Figure 3).

224 For the resistance surfaces generated, Mantel correlation coefficients ranged between -0.2019
225 and 0.6736 (Figure 4). About 56% out of the 10.000 resistance surfaces have a higher correlation with F_{ST}
226 than the one obtained with geographic distance alone. Partial Mantel was used to taking into account the
227 effect of geographical distance, providing thus an estimate of how genetic divergence is explained by
228 landscape alone. The average partial Mantel correlation coefficient was equal to 0.2016, demonstrating
229 explanatory gain for genetic divergence among populations, and increase their respective resistances in
230 the model.

231 Three out of the 10.000 resistance surfaces were selected by ΔAIC under 3. These models are
232 providing the best relationship of landscape with the genetic diversity among the “Baru” populations. The
233 Mantel between landscape resistance and genetic divergence suggest a major adjustment of resistance
234 models, about $r = 0.67$, much larger than a pure geographical model, with $r = 0.48$ (Table 1). All Mantel
235 tests were significant at $P < 0.001$, with 999 permutations. The relative W_i suggests that there is about a

236 30% chance that the surface S-9247 is the more adjusted, a high chance since 10.000 models were
237 generated. The second and third models selected have a reduced chance of 8% and 6.8%, respectively.

238 In partial mantel, the resistance distance explains approximately 45% of the genetic divergence
239 between populations - RR, more than twice the geographical distance explanation provided - RG (23%)
240 (Table 2). By partitioning the effects, about 49.2% out of all the genetic divergence is explained only by
241 resistance (a), about 50.7% of the strength by the overlap with geographical distance (b), and 0.1 % is
242 defined exclusively by the geographic distance (c).

243 The study area is composed of 16 different classes (Table 2). With 28% of the landscape
244 composed by Savannah Agricultural Activities, 21% of Dense Arboreal Savannah, 17.8% of a transition
245 area between different vegetation types (Ecological Tension Areas), 7.7% of Savannah Grassy Woody,
246 6.5% of roads and 4.5% of rivers and lakes. Other classes of landscape account for the remaining 14.5%.
247 The models selected suggest that areas with lower resistance are those with vegetation types Dense
248 Arboreal Savannah and Savanna Grassy Woody (grassland). Savannah Agricultural Activities, Roads,
249 and rivers cause higher resistance.

250 We built a final effective resistance surface from the weighted resistance costs and the weight
251 of the selected models (Figure 5a). The surface is correlated with the genetic divergence of “Baru”
252 populations, with $r = 0.6854$ ($P < 0.001$). The primary gene flow routes of the “Baru” occur where the
253 landscape is less resistant, and dispersal routes are more likely to occur (Figure 5b). The models suggest
254 that areas with lower resistance to gene flow are the Savanna and Savanna Dense Arboreal Grassy
255 Woody. Roads, rivers, and the agricultural regions of Cerrado cause higher resistance to gene flow. We
256 used a current map to preview important connectivity areas between populations; in this case, the minor
257 resistance areas represent areas propitious for species life and to areas suitable to the disperser animals of
258 the species. Warmer colors (purple and red) indicate areas with less current density; areas with higher
259 connectivity are shown in yellow.

260

261 **Discussion**

262 In population genetics, several approaches have been used to investigate patterns and infer
263 microevolutionary processes involved in population differentiation. Landscape genetics is the study of
264 how landscape pattern (the distribution of suitable habitat, barriers, etc.) affects gene flow and genetic

265 differentiation of species (Holderegger and Wagner 2008; Manel and Holderegger 2013). Here we
266 correlated F_{ST} with geographical distances and landscape features using distinct approaches that allow us
267 to decouple their effects. Our results seem to be sufficiently robust to can furnish a description of
268 landscape influence in the genetic variability in a relatively well-known species and corroborate previous
269 analyses of population genetic structure with “Baru” populations (e.g., Soares et al. 2008; Telles et al.
270 2014), adding and complementing information, suggesting influences of anthropic actions among these
271 populations. The Circuitscape has been widely used to check used how landscape features influence
272 genetic connectivity (Adams et al. 2016, Mateo-Sánchez et al. 2015 and Pérez-Espona et al. 2012). For
273 creating the resistance surface is necessary parameterizing cost surfaces by assigning weights to different
274 landscape elements has been challenging, however, because real costs are rarely known (Koen et al.
275 2012). Here, we seek to understand how each landscape components influence genetic divergence
276 between populations of *Dipteryx alata*. Using a Pattern-Oriented Modeling approach, we established
277 which cost configuration best explains the genetic divergence between the studied populations, bringing
278 The approach we used allows getting more insights into the importance of landscape to the genetic flow
279 of populations based on a new way to minimize the arbitrariness in the parameterization of resistance
280 surfaces.

281 The patterns found in this study show the benefits of using an additional set of information to
282 create these surfaces and to interpret the genetic differentiation among populations. Even though another
283 study will have been trying to develop ways to measure the resistance surfaces more clearly by using
284 modeling (e.g., Shirk et al. 2010; Spear et al., 2010) until now, there is no consensus on the most effective
285 approach. Spear et al. (2005) took a big step by using the lowest cost path analysis and discussing the
286 difficulty of developing cost parameters for different habitat types, without having the necessary data (the
287 species biology information and its dispersion). But they did not use cover type because they did not use
288 data that would allow quantifying the specific numerical cost of moving through of the studied
289 salamander species. Koen et al. (2012) carried out a sensitivity analysis of three methods to parameterize
290 a cost surface and two models of landscape permeability. They check that developing a cost surface
291 improves the accuracy of functional connectivity estimates, especially when cost weights are selected
292 through statistical model fitting procedures.

293 The “Baru” has been widely studied in ecological and genetic terms due to its economic,
294 environmental, and cultural importance. Our results seem to be sufficiently robust to can furnish a

295 description of landscape influence in the genetic variability in a relatively well-known species. Our results
296 corroborate and expand our understanding of the factors driving the population genetic structure of this
297 species. Soares et al. (2008) detected that in local populations situated at short geographic distances has
298 been the spatial structure of genetic divergence demonstrated a pattern of genetic discontinuities,
299 suggesting influences of anthropic actions among these populations. Other studies with this species (e.g.,
300 Telles et al. 2014; Diniz-Filho et al. 2015; Soares et al. 2015) have begun to worry about the influence of
301 environmental characteristics on the genetic differentiation of “Baru” populations. Telles et al. (2014)
302 correlate different landscape metrics with the genetic divergence of these same Baru populations finding a
303 strong correlation between the percentage of natural remnant and genetic divergence of populations,
304 demonstrating how human occupation had effects such as habitat loss and fragmentation. Soares et al.
305 (2015) and Smith et al. (2015) discuss and analyze the distribution of these populations from a *center-*
306 *periphery* dynamic, using this environmental suitability and demonstrating a historical influence on the
307 distribution of these populations.

308 The genetic divergence between populations of *D. alata* is better explained by landscape
309 structure than by merely geographic distance. Only about 23% of the genetic divergence is explained by
310 geographic distance, reinforcing that factors other than the geographical distance influence the genetic
311 differentiation among Baru populations. For the resistance surfaces generated, Mantel tests resulted in
312 correlation coefficients varying between -0.2019 and 0.6736, demonstrating the importance of
313 considering landscape components. While the Euclidean distance was a path for this system, this is not
314 usually the best case for studies that have landscape additive information (Coulon et al. 2004, Emel, and
315 Storfer 2015).

316 Regarding landscape components, forest formations did not have a significant influence on the
317 results, mainly due to its small presence in the landscape. The landscape classes that occur in small
318 quantity facility the transposable for “Baru” dispersers, such as birds (macaws) and mammals (monkey,
319 agouti, and livestock), all long-distance dispersers (Ribeiro et al. 2000). Savanna arboreal dense and
320 savanna grassy woody (grassland) behaved as classes with lower resistance to species, observing that low
321 resistance has a direct association with high percolation for animals that disperse their seeds and pollen,
322 as it is a plant with zoochoric dispersion. We expected low resistance in savanna arboreal dense and
323 savanna grassy woody since this species can be found in this type of environment. The classes that
324 showed higher resistance to “Baru” dispersion were the savanna agricultural area, rivers, lakes, and roads,

325 all of them with values close to 35, suggesting high resistance, but also high percolation since it can vary
326 between 0 and 100 (35 is medium resistance). We expected that these landscape classes to have higher
327 resistance considering that they are disturbed areas. We believe that agricultural areas cause increased
328 resistance, mainly due to its instability plant; the Savanna case, they constitute a very high percentage of
329 areas, 28.11%, of the entire study area, with long stretches. The fact that dispersion is mostly associated
330 with terrestrial animals and large and medium-sized flying animals justify the median interference of
331 landscape on species' gene flow. We expected that species with more restrictive dispersal to have more
332 extreme and high values, which reflects mainly on the ability of "Baru" dispersers to overcome barriers
333 and high resistance areas along the dispersion process. The relative importance of landscape components
334 and their spatial patterns can be the key for identifying their influence in microevolutionary processes
335 driving population divergence.

336 From a conservation point of view, the current map (McRae et al. 2008) is a source of
337 information on population connectivity. It has been used to demonstrate critical areas for species'
338 connectivity maintenance (Castilho et al. 2011 and Schwartz et al. 2009). This enhances the
339 understanding of isolated species and facilitates the process of decision-making regarding the main routes
340 of connectivity between populations. Therefore, this work shows an improvement regarding the previous
341 analyses by demonstrating the influence that landscape components have on the processes that generate
342 such genetic variation. Further studies comparing different tree species in this region would allow
343 correlating the weights obtained with their life-history attributes, reinforcing the interpretation of how
344 these differences are captured by IBR and IBE models. Moreover, once these relationships are better
345 established, it would be possible to evaluate how the profound ongoing landscape changes in Brazilian
346 Cerrado (Bonanomi et al. 2019) would disrupt gene flow and, consequently, would lead some
347 economically important species as the "Baru" to local or global extinction.

348

349

350 **References**

- 351 Adams R, Lazerte S and Otter K (2016) Influence of landscape features on the microgeographic genetic
352 structure of a resident songbird. *Heredity* 117: 63–72
- 353 Ayram CC, Mendoza M, Etter A and Pérez-Salicrup D (2017) Anthropogenic impact on habitat
354 connectivity: A multidimensional human footprint index evaluated in a highly biodiverse landscape of
355 Mexico. *Ecological Indicators*. 72:895-909
- 356 Amos JN, Bennett AF, Mac Nally R, Newell G, Pavlova A, Radford JQ, Thomson JR, White M and
357 Sunnucks P (2012) Predicting landscape-genetic consequences of habitat loss, fragmentation and
358 mobility for multiple species of woodland birds. *PLoS ONE* 7(2): e30888
- 359 Bonanomi J, Tortato FR, Santos R, Penha JM, Bueno A S, Peres CA (2019) Protecting forests at the
360 expense of native grasslands: Land-use policy encourages open-habitat loss in the Brazilian Cerrado
361 biome. *Perspectives in Ecology and Conservation*. 17:26-31
- 362 Braga, RT, Rodrigues, JFM, Diniz-Filho JAF, and Rangel TF (2019) Genetic Population Structure and
363 Allele Surfing During Range Expansion in Dynamic Habitats. *Anais da Academia Brasileira de*
364 *Ciências* 91: e20180179
- 365 Burnham KP, Anderson DR (2002) *Model selection and multimodel inference: a practical information-*
366 *theoretic approach*, 2nd edn. Springer, New York
- 367 Castilho CS, Marins-Sá LG, Benedet RC and Freitas TO (2011). Landscape genetics of mountain lions
368 (*Puma concolor*) in southern Brazil. *Mammalian Biology - Zeitschrift Für Säugetierkunde*, 76: 476-
369 483
- 370 Collevatti RG, Telles MP, Nabout JC, Chaves LJ and Soares TN (2013) Demographic history and the low
371 genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. *Hered* 111:97–
372 105
- 373 Coulon A., Cosson JF, Angibault JM, Cargnelutti B, Galan M, Morellet N, Petit E, Aulagnier S and
374 Hewison AJM (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting
375 a fragmented landscape: An individual-based approach. *Mol Ecol* 13:2841–2850
- 376 deLeeuw J (1992). Introduction to Akaike (1973) information theory and an extension of the maximum
377 likelihood principle. Pages 599–609 in S. Kotz, and N.L. Johnson (eds.) *Breakthroughs in statistics*.
378 Vol. 1. Springer-Verlag, London.
- 379 Diniz-Filho JA, Rangel TF and Bini LM (2008) Model selection and information theory in geographical
380 ecology. *Global Ecology and Biogeography*, 17: 479-488
- 381 Diniz-Filho JAF, Rodrigues H, Telles MPDC, Oliveira G, Terribile LC, Soares TN and Nabout JC (2015)
382 Correlation between genetic diversity and environmental suitability: taking uncertainty from
383 ecological niche models into account. *Mol Ecol Resour* n/a-n/a

- 384 Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, Telles MPDC, Rangel TF and Bini
385 LM (2013) Mantel test in population genetics. *Genet Mol Biol* 36:475–485
- 386 Diniz-Filho JAF, Soares TN, Pires M and Telles DEC (2014) Pattern-oriented modelling of population
387 genetic structure. *Biol J Linn Soc* 1152–1161
- 388 Diniz-Filho JAF, Melo DB, de Oliveira G, Collevatti RG, Soares TN, Nabout JC, de Lima JS,
389 Dobrovolski R, Chaves LJ, Naves RV et al. (2012) Planning for optimal conservation of geographical
390 genetic variability within species. *Conserv Genet* 13:1085–1093
- 391 Emel SL, Storfer A (2015) Landscape genetics and genetic structure of the southern torrent salamander,
392 *Rhyacotriton variegatus*. *Conserv Genet* 16: 209-221
- 393 ESRI (2011) ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute
- 394 Frankham R, Ballou JD and Briscoe D a (2004) A Primer of Conservation Genetics. Oryx
- 395 Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol Ecol*
396 *Notes* 5:184–186
- 397 Grimm V (1994) Mathematical models and understanding in ecology. *Ecol Modell* 75–76:641–651
- 398 Grimm V, Frank K, Jeltsch F, Brandl R, Uchmański J and Wissel C (1996) Pattern-oriented modelling in
399 population ecology. *Sci Total Environ* 183:151–166
- 400 Grimm V and Railsback SF (2012) Pattern-oriented modelling: a “multi-scope” for predictive systems
401 ecology. *Philos Trans R Soc B Biol Sci* 367:298–310
- 402 Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke H-H, Weiner J, Wiegand T
403 and DeAngelis DL (2005) Pattern-oriented modeling of agent-based complex systems: lessons from
404 ecology. *Science* 310:987–991
- 405 Holderegger R and Wagner HH (2006) A brief guide to Landscape Genetics. *Landsc Ecol* 21:793–796
- 406 Holderegger R and Wagner HH (2008) Landscape genetics. *Bioscience* 58:199–207
- 407 Jenkins DG, Carey M, Czerniewska J, Fletcher J, Hether T, Jones A, Knight S, Knox J, Long T, Mannino
408 M, McGuire M, Riffle A, Segelsky S, Shappell L, Sterner A, Strickler T and Tursi R (2010) A
409 meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*
410 33(2): 315-320
- 411 Kang, JY and Aldstadt, J (2018) Using Multiple Scale Spatio-Temporal Patterns for Validating Spatially
412 Explicit Agent-Based Models. *International Journal of Geographical Information Science*. 33: 1-21
- 413 King CA and With K (2002) Dispersal success on spatially structure landscapes: when do spatial pattern
414 and dispersal behaviour really matter? *Ecol Modell* 147:23–29
- 415 Klink CA (2013) Policy intervention in the Cerrado Savannas of Brazil: changes in land-use and effects
416 on conservation. *Ecol Conserv Maned Wolf Multidiscip Perspect* 293–307

- 417 Klink CA and Machado RB (2005) Conservation of the Brazilian Cerrado. *Conserv Biol* 19:707–713
- 418 Koen EL, Bowman J and Walpole A A. (2012) The effect of cost surface parameterization on landscape
419 resistance estimates. *Mol Ecol Resour* 12:686–696
- 420 Koffi KJ, Deblauwe V, Sibomana S, Neuba DFT, Champluvier CDC, Barbier N, Traore D, Habonimana
421 B, Robbrecht E, Lejoly J and Bogaert J (2007) Spatial Pattern Analysis as a Focus of Landscape
422 Ecology to Support Evaluation of Human Impact on Landscapes and Diversity. In: Hong SK.,
423 Nakagoshi N., Fu B., Morimoto Y. (eds) *Landscape Ecological Applications in Man-Influenced*
424 *Areas*. Springer, Dordrecht
- 425 Kramer AT, Ison JL, Ashley M V. and Howe HF (2008) The paradox of forest fragmentation genetics.
426 *Conserv Biol* 22:878–885
- 427 Legendre P and Legendre L (2012) *Numerical Ecology*, 3rd edition. Elsevier, Amsterdam
- 428 Manel S and Holderegger R (2013) Ten years of landscape genetics. *Trends Ecol Evol* 28:614–621
- 429 Manel S, Schwartz MK, Luikart G and Taberlet P (2003) Landscape genetics: Combining landscape
430 ecology and population genetics. *Trends Ecol Evol* 18:189–197
- 431 Mateo-Sánchez MC, Balkenhol N, Cushman S, Pérez T, Domínguez A and Saura S (2015) Estimating
432 effective landscape distances and movement corridors: comparison of habitat and genetic data.
433 *Ecosphere* 6(4):59
- 434 McRae BH (2006) Isolation by resistance. *Evolution* 60:1551–1561
- 435 McRae BH and Beier P (2008) Circuit theory predicts gene flow in plant and animal populations.
436 *Academy of Sciences of the United States of America*. 104: 19885-90
- 437 McRae BH, Schumaker N, McKane R, Busing R, Solomon A and Burdick C (2008) A Multi-Model
438 Framework for Simulating Wildlife Population Response to Land-Use and Climate Change.
439 *Ecological Modelling* 219: 77-91
- 440 Metzger J and Ddcamps H (1997) The structural connectivity threshold: an hypothesis in conservation
441 biology at the landscape scale. 18:1–12
- 442 Okamiya H, Kusano T (2019) Effects of landscape features on gene flow among urban frog populations.
443 *Ecol. Res.* 34: 497– 508
- 444 Oksanen J, Blanchet JG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P,
445 Stevens MHH and Wagner H (2012) *Vegan: Community Ecology Package*. R package version 2.0-5.
- 446 Pérez-espona S, Pérez-barbería FJ, Mcleod JE, Jiggins CD, Gordon IJ and Pemberton JM (2008)
447 Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). *Molecular*
448 *Ecology*, 17: 981-996
- 449 Perez-Espona, S. et al. (2012) Landscape genetics of a top neotropical predator. *Mol. Ecol.* 21:5969-5985

- 450 Peterman W (2014) ResistanceGA: An R package for the optimization of resistance surfaces using
451 genetic algorithms. *Methods in Ecology and Evolution*. 9
- 452 R Core Team (2014) R: A language and environment for statistical computing. *R A Lang. Environ. Stat.*
453 *Comput.* 1
- 454 Ribeiro JF, Sano SM, Brito MA, et al (2000) Baru (*Dipteryx alata* Vog.). Funep, Jaboticabal
- 455 Schwartz M K, Copeland JP, Anderson NJ, Squires JR, Inman RM, McKelvey KS, Pilgrim KL, Waits LP
456 and Cushman SA (2009) Wolverine gene flow across a narrow climatic niche. *Ecology* 90: 3222-3232
- 457 Sexton JP, Hangartner SB and Hoffmann AA (2014) Genetic isolation by environment or distance: which
458 pattern of gene flow is most common? *Evolution*, 68: 1-15
- 459 Shirk A, Wallin D, Cushman SA, Rice C, Warheit K (2010) Inferring landscape effects on gene flow: a
460 new model selection framework. *Mol Ecol* 19:3603–3619
- 461 Soares TN, Chaves LJ, Telles MPC, Resende LV, Diniz-Filho JAF (2008) Landscape conservation
462 genetics of *Dipteryx alata* (Baru tree: Fabaceae) from Cerrado region of Central Brazil. *Genetica*,
463 132:9-19
- 464 Soares TN, Diniz-Filho JAF, Nabout JC, de Campos Telles MP, Terribile LC and Chaves LJ (2015)
465 Patterns of genetic variability in central and peripheral populations of *Dipteryx alata* (Fabaceae) in the
466 Brazilian Cerrado. *Plant Syst Evol* 301:1315–1324
- 467 Soares TN, Melo DB, Resende LV, Vianello RP, Chaves LJ, Collevatti RG and de Campos Telles MP
468 (2012) Development of microsatellite markers for the neotropical tree species *Dipteryx alata*
469 (Fabaceae). *Am J Bot* 99:5–6
- 470 Spear SF, Balkenhol N, Fortin MJ, McRae BH and Scribner K (2010) Use of resistance surfaces for
471 landscape genetic studies: Considerations for parameterization and analysis. *Mol Ecol* 19:3576–3591
- 472 Spear SF, Peterson CR, Matacq M, Storfer A (2005) Landscape genetics of the blotched tiger salamander
473 (*Ambystoma tigrinum melanostictum*). *Mol Ecol* 14: 2553–2564
- 474 Stevenson-Holt CD, Watts K, Bellamy CC, Nevin OT and Ramsey AD (2014) Defining Landscape
475 Resistance Values in Least-Cost Connectivity Models for the Invasive Grey Squirrel: A Comparison
476 of Approaches Using Expert-Opinion and Habitat Suitability Modelling. *PLoS One* 9:e112119
- 477 Storfer A (1999) Gene flow and endangered species translocarions: a topic revisited. *Biol Conserv*
478 87:173–180.
- 479 Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling
480 L and Waits LP (2007) Putting the “landscape” in landscape genetics. *Heredity* (Edinb) 98:128–142
- 481 Taylor PD, Fahrig L, Henein K, Merriam G (1993) Connectivity is a vital element of landscape structure.
482 *Oikos* 68:571–573

- 483 Telles MPDC, Dobrovolski R, Souza KS, Lima JDS, Collevatti RG, Soares TN, Chaves LJ and Diniz-
484 Filho JAF (2014) Disentangling landscape effects on population genetic structure of a Neotropical
485 savanna tree. *Nat Conserv* 12:65–70. doi: 10.4322/natcon.2014.012
- 486 Tischendorf L and Fahrig L (2000) On the usage and measurement of landscape connectivity. *Oikos*
487 90:7–19
- 488 Topping CJ, Dalkvist T and Grimm V (2012) Post-Hoc Pattern-Oriented Testing and Tuning of an
489 Existing Large Model: Lessons from the Field Vole. *PLoS One*. doi: 10.1371/journal.pone.0045872
- 490 Vignieri SN (2005) Streams over mountains: Influence of riparian connectivity on gene flow in the
491 Pacific jumping mouse (*Zapus trinotatus*). *Mol Ecol* 14:1925–1937
- 492 Wang I and Bradburd G (2014) Isolation by Environment. *Ecol Lett* 5649–5662
- 493 Wiegand T, Jeltsch F, Hanski I and Grimm V (2003) Using pattern-oriented modeling for revealing
494 hidden information: a key for reconciling ecological theory and application. *Oikos* 65:209–222
- 495 Wright S (1943) Isolation by distance. *Genetics* 28:114-138.

Surface	Mantel	Partial Mantel	RT	RR	RG	a	b	c	ΔAIC	Wi
S-9247	0.6736	0.5404	0.456	0.454	0.23	0.225	0.228	0.001	0	0.29
S-3939	0.6702	0.5364	0.452	0.449	0.23	0.221	0.227	0.002	2.528	0.082
S-4515	0.6697	0.5319	0.449	0.449	0.23	0.218	0.23	-0.001	2.895	0.068

496 **Table 1. Resistance surface selected according to ΔAIC**

497

498

499 **Table 2. Influence of landscape features on the gene of flow *Dipteryx alata*. Costs on the resistance**
 500 **surfaces selected and the sum of the pondered cost per wi.**

PHYTOECOLOGICAL REGIONS	PLAND	S-9247	S-3939	S-4515	S-mean	Σ(Xi*Wii)
Savannah Agricultural Activities	28.11	79	91	83	84	36.05
Roads	6.52	86	90	77	84	37.60
Rivers Lakes	4.53	78	66	97	80	34.66
Deciduous Forest Secondary Vegetation and Agricultural Activities	2.24	80	72	71	74	33.96
Savannah Park	4.79	45	93	70	69	25.46
Ecological Tension Areas of Savannah Savannah Estépica	0.24	7	95	93	65	16.17
Savannah arborous	2.99	81	56	56	64	31.92
Deciduous Forest Submontane	0.04	36	51	100	62	21.45
Areas of Pioneer Formations Influences Fluvial	0.10	92	3	92	62	33.22
Ecological Tension Areas Savanna Forest Estacional	17.50	50	70	65	62	24.68
Semideciduous Forest Secondary Vegetation and Agricultural Activities	3.00	49	61	57	56	23.11
Semideciduous Forest Alluvial	0.22	20	38	51	36	6.55
Semideciduous Forest Submontane	0.79	52	38	16	35	17.08
Ecological Tension Areas Forest Rain Forest Estacional	0.10	47	19	15	27	16.22
Savannah Grassy Woody	7.70	22	32	17	24	12.21
Dense Arboreal Savannah	21.13	5	4	9	6	2.39

501

pland = percentage area in the landscape. Σ (xi * wii) = sum of each weighted resistance by wi your model.

502

503 **Figure captions**

504 **Fig. 1** Distribution of *Dipteryx alata* population studied in the Brazilian Cerrado, with the landscape
505 components (classes) analyzed in resistance surface. Areas in green tones are natural areas, and brown
506 tones are predominantly disturbed areas. Red circles indicate the number of 25 populations genotyped,
507 ranging between 12 and 37 individuals

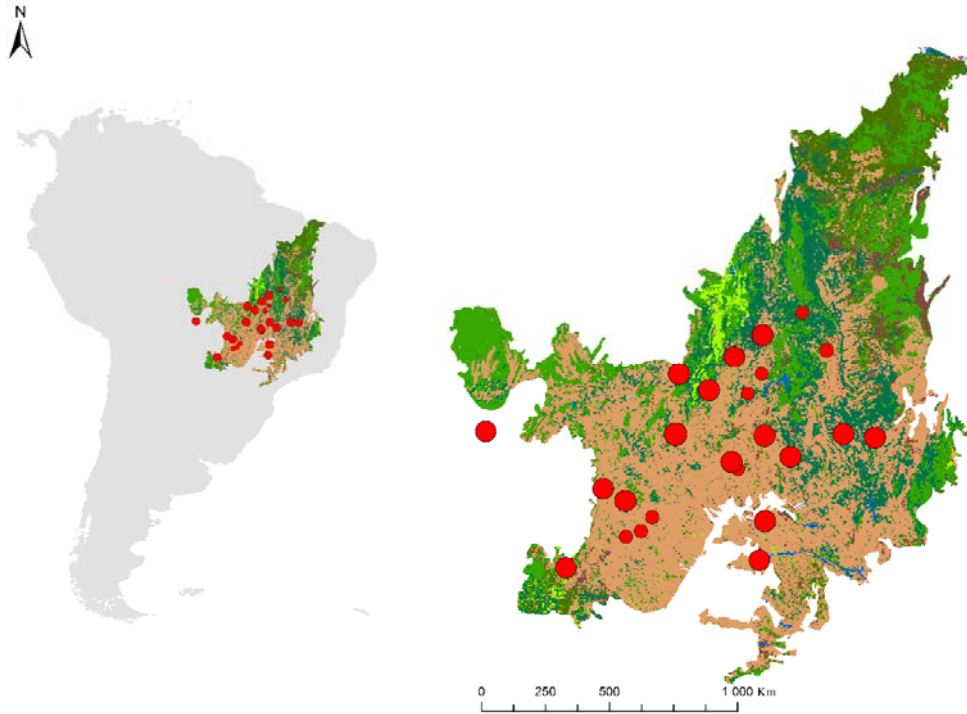
508 **Fig. 2 a)** and **b)** are homogeneous surfaces with respectively resistance values of 0 and 100 in all areas.
509 The surfaces **c)** and **d)** are heterogeneous and contain the parameterized classes with values of 0 to 100 at
510 intervals of 1

511 **Fig. 3** Relationship between pairwise F_{ST} and geographical distance

512 **Fig. 4** Histogram of Pearson's correlation coefficient (r) values obtained using the Mantel test between
513 pairwise F_{ST} and resistance matrices, highlighting the correlation coefficient between the F_{ST} and
514 geographical distances, $r = 0.4805$ (red line)

515 **Fig. 5 a)** landscape resistance map for Baru (*Dipteryx alata*) in the Brazilian Cerrado with average
516 resistance costs ranging between 2.3928 and 37.587. The resistance gene flow was parameterized with
517 average model parameter estimates for these variables. The lighter areas have lower resistance, and the
518 darker has a higher resistance. **b)** current map with the main routes of lower resistance to the dispersion of
519 Baru

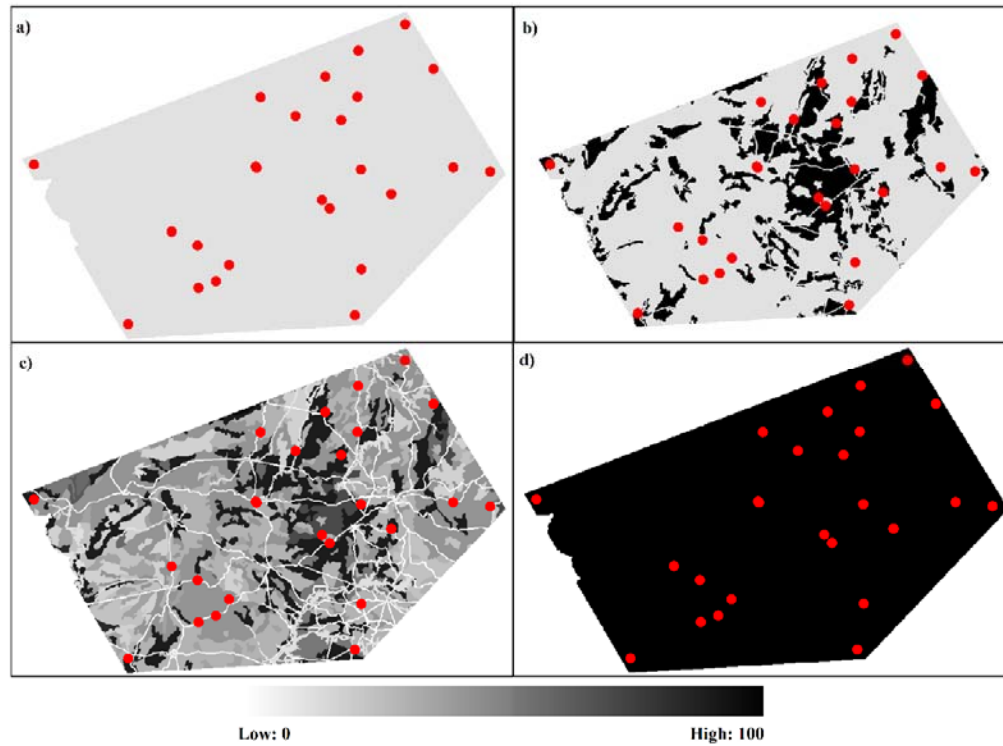
520



521

522 **Fig 1**

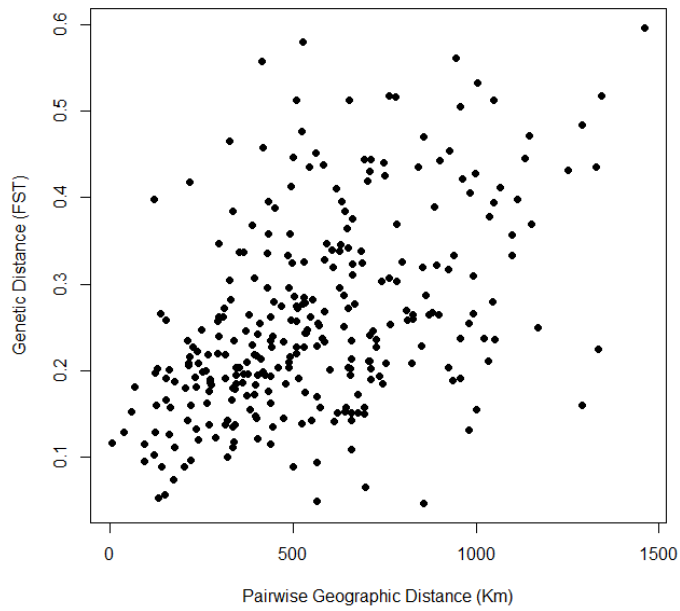
523



524

525 **Fig 2**

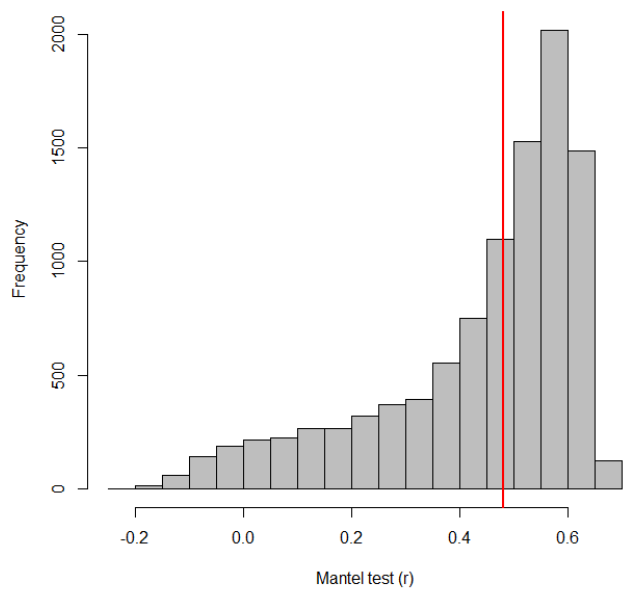
526



527

528 **Fig 3**

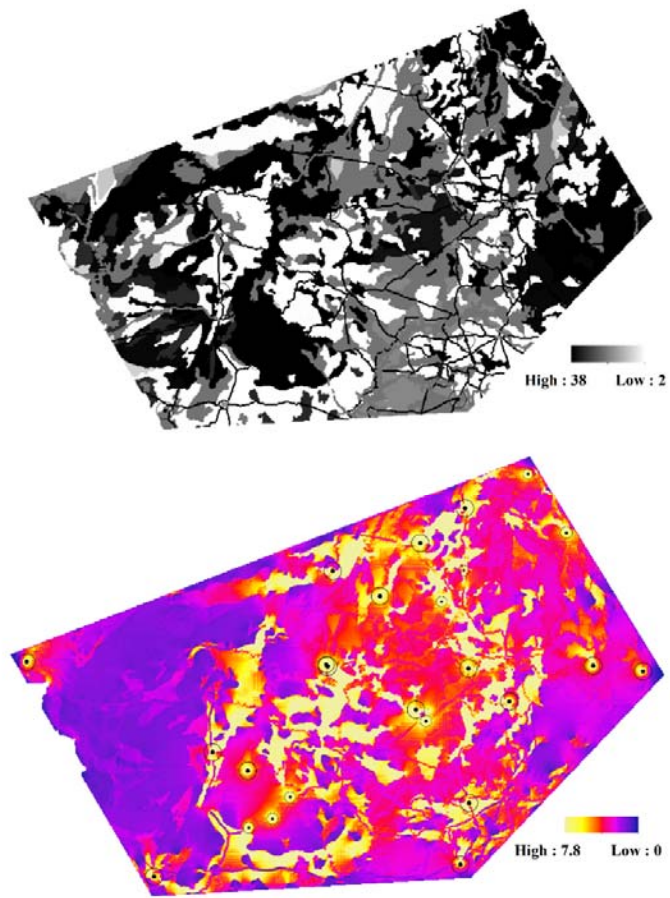
529



530

531 **Fig 4**

532



533

534 **Fig 5**

535

536 **Electronic Supplementary Material**

537 **Table S1:** Geographical coordinates of Baru Populations

Population	City / State	Longitude	Latitude	Number of Samples
1	Cocalinho/MT	-14.397	-50.996	32
2	Nova Nazaré/MT	-13.839	-52.042	32
3	Pirenópolis/GO	-15.997	-49.034	32
4	Sonora/MS	-17.853	-54.704	31
5	Alcinópolis/MS	-18.268	-53.926	32
6	Alvorada/TO	-12.449	-49.115	32
7	São Miguel do Araguaia/GO	-13.225	-50.103	32
8	Luziânia/GO	-16.732	-48.13	32
9	Icém/SP	-20.347	-49.216	31
10	Monte Alegre de Minas/MG	-18.978	-49.024	32
11	Estrela do Norte/GO	-13.828	-49.14	12
12	Santa Terezinha de Goiás/GO	-14.52	-49.628	12
13	Arinos/MG	-15.934	-46.273	32
14	Pintópolis/MG	-16.061	-45.166	32
15	Chapadão do Sul/MS	-18.846	-52.982	13
16	Água Clara/MS	-19.33	-53.377	13
17	Camapuã/MS	-19.528	-53.9	13
18	Indiara/GO	-17.162	-49.973	13
19	Aragarças/GO	-15.9112	-52.187	27
20	Aragarças/GO	-15.9482	-52.158	37
21	Palminópolis/GO	-15.9121	-50.201	32
22	Chapada da Natividade/TO	-11.6614	-47.714	12
23	Arraias/TO	-12.9904	-46.863	15
24	Anastácio/MS	-20.611	-56.004	31
25	Porto Esperidião/MT	-15.853	-56.8222	30

538