DR. LUCIANA CUNHA RESENDE-MOREIRA (Orcid ID : 0000-0003-2977-8433)

MS. JOYCE RODRIGUES PRADO (Orcid ID : 0000-0002-2025-5479)

Author Manuscript

Article type : Research Paper

RESEARCH PAPER

EVOLVING IN ISOLATION: GENETIC TESTS REJECT RECENT CONNECTIONS OF AMAZONIAN SAVANNAS WITH THE CENTRAL CERRADO

Running title: Amazonian savannas: history of isolation

Luciana C. Resende-Moreira[ab1], L. Lacey Knowles[b1], Andréa T. Thomaz[b], Joyce R. Prado[c], Andrea P. Souto[a], José P. Lemos-Filho[d], Maria Bernadete Lovato[a2]

[a] Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil

[b] Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Michigan 48109

[c] Departamento de Ciências Biológicas, Escola Superior de Agricultura 'Luiz de Queiroz', Universidade de São Paulo, Piracicaba, SP 13418-900, Brazil

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jbi.13468

This article is protected by copyright. All rights reserved

[d] Departamento de Botânica, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil

[1] Luciana C. Resende-Moreira and L. Lacey Knowles contributed equally to this work

[2] Corresponding author: Maria Bernadete Lovato

E-mail address: lovatomb@icb.ufmg.br

Luciana C. Resende-Moreira e-mail address: resendelc@gmail.com

Departamento de Biologia Geral, Universidade Federal de Minas Gerais, CP 486, Belo Horizonte, MG 31270-901, Brazil

Word count: 6,825 words

Author Manuscript

This article is protected by copyright. All rights reserved

ACKNOWLEDGMENTS

The authors thank the Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), for providing sampling permits, and the Instituto Nacional de Ciência e Tecnologia Herbário Virtual da Flora e dos Fungos (INCT-HVFF, funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq) and Ary Oliveira-Filho for compiling information regarding species occurrence. We are especially thankful to taxonomists Maria Candida Mamede and Augusto Francener from Instituto de Botânica for help with species identification. We also thank for all assistance for sampling AS populations: Hugo Alencar from Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) in Vilhena/RO, Cláudia Lima from ICMBio in Porto Velho/RO, Marcelo Carvalho and Bruno Souza from ICMBio in Boa Vista/RR, Adilson Lopes and Rogério Alves from Embrapa Macapá/AP. We are grateful to Anna Papadopoulou for helpful discussion of analyses and processing of ddRADseq data. Funding was provided for field work and genetic analysis by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and CNPq. This research was also partially supported by funds to LLK from the University of Michigan, Ann Arbor. LCRM received a PhD scholarship from the Brazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil (CAPES PDSE 99999.003882/2015-02), and MBL and JPLF received research fellowships from CNPq.

Author Manuscript

This article is protected by copyright. All rights reserved

Author Manuscript

1 **ABSTRACT**

2 **Aim:** The effects of past climatic shifts remain enigmatic for the Amazon region, especially for
3 islands of savanna within the tropical forest known as "Amazonian savannas" (AS). These
4 disjunct savanna areas share many plant and animal species with the Cerrado biome in central
5 Brazil (the CC), fuelling debate over historical connections. We evaluate hypothesized corridors
6 between the CC and the AS, and specifically investigate whether a history of isolation versus
7 recent connections are supported by genetic tests.

8 **Location:** Cerrado and Amazon biomes

9 **Taxon:** Two woody plant species: *Byrsonima coccolobifolia* and *B. crassifolia* (Malpighiaceae).

10 **Methods:** Analyses of genomic data (SNPs from more than 4,500 loci) in 28 populations, as
11 well as chloroplast DNA (cpDNA), were used to test for parallel geographic structuring between
12 the CC and AS – an expected structure if putative corridors provided regional connections
13 between different areas of the CC and AS, and divergence times between the CC and AS were
14 estimated using a composite-likelihood method based on the site frequency spectrum.

15 **Results:** Genomic data, in contrast with cpDNA, generally show strong, concordant genetic
16 structure between the CC and AS in both species, rather than regional grouping of CC with AS
17 populations. In addition, divergence between the CC and AS predates the last glacial maximum.

18 **Main conclusions:** Our results suggest the AS have remained relatively isolated from the CC
19 even though the strong structure of genomic variation is not shared by cpDNA. We note that past
20 evidence of putative corridors between the CC and AS based solely on cpDNA should be
21 interpreted cautiously since the lack of structure may reflect limited genetic resolution rather
22 than gene flow. As such, the uniqueness of AS may be more pronounced than previously
23 thought, highlighting the importance of protecting these highly threatened areas.

24 **Keywords:** Amazon, *Byrsonima*, Cerrado, corridor, Malpighiaceae, phylogeography, RAD-seq,
25 relict, savanna

This article is protected by copyright. All rights reserved

26 **INTRODUCTION**

27 Climate change has induced historical shifts in landscapes, including the fragmentation of
28 once widespread biomes into relatively isolated patches. The persistence of such populations and
29 the evolutionary dynamics shaping their current genetic structure are commonly considered in
30 studies of the northern hemisphere following the glacial retreat of the Pleistocene (Hewitt, 2004;
31 Knowles & Massatti, 2017; Pielou, 1992). However, the impact of past climatic shifts is not
32 unique to these areas. The effects of climatic extremes are worldwide, with documented shifts of
33 biomes leaving behind relict populations (e.g., Bonatelli et al., 2014; Migliore et al., 2013;
34 Ornelas, Ruiz-Sanchez, & Sosa, 2010). However, tropical regions remain critically understudied
35 relative to their northern counterparts. The evolutionary history of many tropical biomes is also
36 enigmatic because of particularly sparse palynological or fossil evidence (e.g., Jaramillo et al.,
37 2010) and limited or inconsistent support for a range of different hypotheses regarding the
38 magnitude of climate-induced distributional shifts.

39 Such uncertainty is exemplified by debates over the evolutionary history of the central
40 Cerrado (CC) and Amazonian savannas (AS) of Brazil (Fig. 1). The CC is a hyper-diverse, yet
41 relatively understudied savanna biome that covers over 2 million km^2. Many plant and animal
42 taxa (including over 70 woody species) are present in the CC and AS, with some AS displaying
43 higher floristic similarity to locations within the CC than to geographically proximate AS
44 (Ratter, Bridgewater, & Ribeiro, 2003), suggesting past connections between the CC and AS
45 (Prance, 1996; Silva, 1995; Silva & Bates, 2002), rather than independent long-distance dispersal
46 events (see Pennington, Lewis, & Ratter, 2006). However, different hypotheses narrate how the
47 retraction of the Cerrado from its former maximum extent might have occurred, which include
48 past connections – that is, corridors – between the Cerrado and areas where AS persist today.
49 Where such corridors might have existed, and which geographic areas they might have
50 connected are still debated. For example, three different corridors between the CC and AS have
51 been proposed: a coastal corridor, a central Amazonian corridor and an Andes corridor (Haffer,
52 1967; 1974; Webb, 1991). Depending on the study, support for hypothesized corridors differ, as
53 does the purported timing of past connections between the CC and AS (e.g., Bueno et al., 2017;
54 Quijada-Mascareñas et al., 2007; Savit & Bates, 2015; Vargas-Ramírez, Maran, & Fritz, 2010;

Author Manuscript

This article is protected by copyright. All rights reserved

55 Werneck, Nogueira, Colli, Sites, & Costa, 2012). That is, the uncertainty over the geographic
56 location of corridors is paralleled by debate over when such connections might have occurred
57 (e.g., during the Miocene and Pliocene, Pascual & Jaureguizar, 1990, versus the Pleistocene,
58 Haffer, 1969; Prance, 1982; van der Hammen, 1991), including whether such connections might
59 have been forged during the drier climate of the last glacial maximum, LGM, especially given
60 the lack of support for such late Pleistocene expansion based on palynological evidence
61 (Colinvaux, Irion, Rasanen, Bush, & de Mello, 2001; Kastner & Goni, 2003; Mayle, Burn,
62 Power, & Urrego, 2009).
63 Here we address the extent to which the AS has evolved in isolation from the CC by
64 quantifying population genetic structure of two widely distributed tree species that are common
65 in both the CC and AS – *Byrsonima coccolobifolia* Kunth and *Byrsonima crassifolia* (L.) Kunth
66 (Ratter et al., 2003). Specifically, we test the degree to which Cerrado populations are genetically
67 distinct from the AS, as opposed to exhibiting parallel geographic structuring of genetic variation
68 within the CC and among AS, as expected if multiple corridors provided regional connections
69 between different areas of the CC and different subsets of AS. We conducted this tests using
70 genomic data (i.e., more than 7,000 and 4,500 loci sequenced in 86 and 68 individuals of *B.*
71 *coccolobifolia* and *B. crassifolia*, respectively), as well as assays of the geographic structure of
72 chloroplast DNA (cpDNA) across an even broader sampling of populations. In addition to the
73 individual histories, we consider the degree to which the taxa show concordant patterns of
74 genetic variation. As ecologically similar, dominant and co-distributed taxa, concordance would
75 lend support to common factors structuring the history of constituent taxa in this diverse biome
76 (Avise, 2004), thereby overriding stochastic processes associated with the biomes dynamic
77 history (Behling & Hooghiemstra, 2001; Ledru, 2002; but see Massatti & Knowles, 2014; 2016).
78 Lastly, we estimate divergence times between the CC and AS to determine how long the AS may
79 have been evolving independently of the CC.

80 **MATERIAL AND METHODS**

81 **Study species**

82 *Byrsonima* Rich. Ex Kunth is a common genus, with most of its diversity represented by
83 South American savanna taxa, many of which co-occur (Anderson, Anderson & Davis, 2006;

Author Manuscript

This article is protected by copyright. All rights reserved

84 Ratter et al., 2003). *Byrsonima coccolobifolia* and *B. crassifolia* are the most common species
85 from the genus in the Cerrado and in the Amazonian savannas (Ratter et al., 2003), with the
86 range of *B. crassifolia* extending into the savanna woodlands of Central America and Mexico
87 (Anderson, 1981). Its fleshy fruits are bird-dispersed (Anderson, 1983) and flowers are
88 pollinated by oil-collecting bees, especially *Centris* species (Vinson, Williams, Frankie &
89 Shrum, 1997; Benezar & Pessoni, 2006).

90 **Sampling and DNA extraction**

91 Population sampling of *B. coccolobifolia* and *B. crassifolia* covered both species' ranges
92 across the CC and AS (Fig. 1; for details see Table S1.1, Appendix S1 in Supporting
93 Information) and was informed by occurrence data from NeoTropTree (Oliveira-Filho, 2017)
94 and the INCT - Virtual Herbarium of Flora and Fungi (http://inct.splink.org.br/index). A total of
95 158 individuals across 16 populations of *B. coccolobifolia* and 15 populations of *B. crassifolia*
96 were sequenced using RADseq (described below; see also Table 1 and Fig. 1). In addition,
97 cpDNA was sequenced in a larger set of populations (i.e., 46 populations and 218 individuals;
98 for detailed sampling information see Appendix S2). Voucher specimens of all sampled
99 populations were deposited in the Herbarium of Departamento de Botânica, Universidade
100 Federal de Minas Gerais (BHCB) and the Herbarium of Universidade Estadual do Oeste do
101 Paraná (UNOP).

102 DNA was extracted using a CTAB protocol (Novaes, Rodrigues, & Lovato, 2009) from
103 silica-gel dried leaves that were stored at -20ºC until DNA extraction. DNA quality was
104 evaluated with Nanodrop® (Thermo Scientific, Walthham, USA) and quantified with Qubit®
105 (Thermo Scientific).

106 **Genomic dataset**

107 Two genomic libraries were prepared (one for each species) following the double-digest
108 restriction site-associated DNA sequencing (ddRADseq) protocol of Peterson, Weber, Kay,
109 Fisher, & Hoekstra (2012). Briefly, genomic DNA was digested with the restriction enzymes
110 *Eco*RI and *Mse*I, ligated to adaptors with unique barcodes, pooled and size-selected using Pippin
111 Prep (Sage Science, Beverly, USA), and sequenced on an Illumina HiSeq2500 to generate

This article is protected by copyright. All rights reserved
Author Manuscript

112 single-end 50bp reads at The Centre for Applied Genomics, Toronto, Canada (protocol details
113 given in Appendix S3).

114 Genomic data for each species was processed separately using the pipeline STACKS 1.35
115 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Reads were demultiplexed and
116 filtered using the program PROCESS_RADTAGS, with sequences from each individual assembled
117 *de novo* in USTACKS to identify putative loci, and a catalog of consensus loci built in CSTACKS.
118 Individual genotypes were identified with SSTACKS (for details of raw data processing see
119 Appendix S3). Individuals were grouped according to their sampling localities in POPULATIONS
120 (STACKS pipeline), and biallelic loci from a minimum of two populations were used in
121 population genetic analysis (described below). We chose this parameter to maximize the number
122 of loci retained (i.e., for any given minimum of missing data, there is a drop out of loci as the
123 number of individuals increases; Huang & Knowles, 2016). A custom script (available on
124 https://github.com/KnowlesLab; Thomaz, Malabarba, & Knowles, 2017) was used in R 3.2.2 (R
125 Core Team, 2017) to exclude loci with high theta values (located within the upper 95% quantile)
126 and SNPs from the two last nucleotides (Fig. S1.1 in Appendix S1) to guard against sequencing
127 and assembly errors. Following this step, the software PLINK 1.07 (Purcell et al., 2007) was used
128 to identify SNPs with a maximum of 20% of missing data and with a minimum stack depth per
129 individual (m) of five for inclusion in the final dataset.

130 Processed genomic data resulted in 28,487 SNPs for *B. coccolobifolia* and 14,855 SNPs
131 for *B. crassifolia*, and a total of 7,115 and 4,543 loci with one biallelic SNP per locus in each
132 species, respectively; hereafter we refer to this genetic variation sampled across the genome as
133 "genomic" variation or structure. An average of 81% of reads per individual were retained, with
134 a mean coverage depth per locus of 23.6 ± 8.8x after processing and assembly, which is
135 considered adequate for population genomics inference (see Buerkle & Gompert 2013 for
136 details). These loci were identified from the 226 million reads for the 183 individuals sequenced
137 on two Illumina lanes (average of 1,236,269.5 ± 684,965.7 reads per individual; 29 individuals
138 were excluded due to large amounts of missing data – for details see Table S1.2 in Appendix
139 S1).

140 **Characterizations of genomic variation and structure**

This article is protected by copyright. All rights reserved

Author Manuscript

141 Genetic structure was investigated using two different strategies: principal components
142 analysis (PCA), which does not require any assumptions about the underlying genetic model
143 (Jombart, Pontier, & Dufour, 2009), and Bayesian clustering, which applies a coalescent model
144 for inferences about genetic structure. The packages "adegenet" v2.0 (Jombart, 2008; Jombart &
145 Ahmed, 2011) and "ade4" v.1.7-2 (Dray & Dufour, 2007) were used to perform a PCA in R;
146 missing data were replaced by the mean frequency of the most frequent allele. The robustness of
147 PCA results was evaluated using datasets with different levels of missing data (5 and 20%; see
148 Huang & Knowles, 2016) and with an additional minimum stack depth per individual of 10.
149 Because these results were qualitatively similar (Fig. S3.1 in Appendix S3), the results are not
150 discussed further. Bayesian clustering was performed with the software STRUCTURE 2.3.4
151 (Pritchard, Stephens, & Donnelly, 2000), with only one SNP per locus. These analyses included
152 admixture among populations and a correlation among allele frequencies with 1 to 10 genetic
153 clusters (*K*) tested. Ten independent runs were performed for each *K*-value, with 100,000 burn-in
154 and 300,000 MCMC iterations (the number of burn-in and MCMC iteration were increased when
155 necessary to reach convergence). The most probable number of cluster was identified with
156 STRUCTURE HARVESTER (Earl & Vonholdt, 2012), and the posterior probability of individual
157 assignment to each cluster was permuted across different runs and visually displayed with
158 CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). A hierarchical
159 analysis with subsets of populations from each inferred genetic cluster was used to test for
160 additional structure within the initial clusters identified by STRUCTURE (e.g. Massatti & Knowles,
161 2014; Papadopoulou & Knowles, 2015). Hierarchical analyses were performed with the same
162 parameter settings described above, with *K*-values ranging from 1 to the maximum number of
163 populations in each sequential analysis. Note that analyses of genetic structure in *B.*
164 *coccolobifolia* suggested the presence of a cryptic taxon (i.e., PCA analysis revealed that the
165 individuals were quite divergent, and distinct, from all the other populations; see Fig. 2). Because
166 inclusion of these populations (specifically, coAGN, coNAT and coFOR populations) would
167 confound comparisons of CC to AS (e.g., compare PCA with and without these individuals; Fig.
168 2), the populations were removed and are not included in the geographic structure results.
169 Tests of the association between geography and genetic structure were performed using
170 two approaches in each species. Isolation-by-distance (IBD) was tested by evaluating whether

This article is protected by copyright. All rights reserved

Author Manuscript

171 there was a significant correlation between geographic distance and genetic distance (F_{ST} / 1-
172 F_{ST}, Slatkin, 1995) based on 100,000 permutations with the package "vegan" 2.3-1 in R
173 (Oksanen et al., 2017). Additionally, a Procrustes analysis, which retains the relative longitudinal
174 and latitudinal position of populations to test for an association between genetic variation and
175 geography was used (for additional details see Appendix S3), with the significance of the
176 association, t_0, (Wang, Zöllner, & Rosenberg, 2012) evaluated by 10,000 permutations (package
177 "vegan"). The robustness of the association between genes and geography was assessed using a
178 sequential population drop out procedure (see Knowles and Massatti 2017). Geographic
179 structuring of genetic variation was also assessed with additional Procrustes analyses conducted
180 on the CC and AS separately.

181 Lastly, levels of genetic diversity were characterized for each population using the
182 dataset with all SNPs (i.e., not the dataset with only a single SNP per locus). These include
183 estimation of standard population genetics statistics such as nucleotide diversity (π), expected
184 heterozygosity (H_{EXP}), observed heterozygosity (H_{OBS}), and Wright's F-statistics (F_{IS} and
185 pairwise F_{ST}), which were calculated using the POPULATIONS module from the STACKS pipeline
186 (Catchen et al., 2013).

187 **Estimates of divergence times**

188 Divergence times were estimated between the CC and AS using a composite-likelihood
189 method based on the site frequency spectrum (SFS) and implemented in FASTSIMCOAL2
190 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Excoffier & Foll, 2011). To
191 improve the accuracy of parameter estimates from the SFS (and following the recommendations
192 of the program; see Excoffier & Foll, 2011), we fixed the effective population size of the CC,
193 which was calculated directly from the empirical data, whereas the other parameters were
194 estimated (i.e., the population size of AS, N_{AS}, the ancestral population size, N_{ANC}, and the
195 divergence time, T_{DIV}). Specifically, the population size of CC was calculated from the
196 nucleotide diversity, π, of fixed and variable sites using a nuclear genomic mutation rate of $7x10^{-}$
197 ^{9} subs/site/generation (Ossowski et al., 2010). This mutation rate was estimated based on
198 spontaneous mutations of *Arabidopsis thaliana,* a herbaceous annual plant, and therefore
199 divergence times estimated here will tend to be relatively more recent than expected if mutation

This article is protected by copyright. All rights reserved

Author Manuscript

200 rates in *B. coccolobifolia* and *B. crassifolia* are lower, as suggested for other woody plants
201 (Smith & Donoghue, 2008; Yang et al., 2015).
202 Point estimates for each parameter were obtained from the run with the highest maximum
203 likelihood from 40 FASTSIMCOAL2 runs with 100,000 to 250,000 simulations per run, and 10 to
204 40 expectation-conditional maximization (ECM) cycles based upon a stopping criterion of 0.001
205 as a minimum relative difference between two iterations. Confidence intervals were calculated
206 for each parameter from 100 parametric bootstrap replicates of simulated SFS under a model
207 based on the point estimates. Since there is no literature about generation time for those species,
208 divergence time estimates were converted from generations to years assuming a generation time
209 of three years, which was observed by a domestication program to be the age of first fruiting in
210 natural populations (Nascimento W. M. O & Carvalho J. E. U, Embrapa Amazônia Oriental,
211 personal comm.), although we recognize time estimates may be considerably older if generation
212 times of 10-15 years for Cerrado trees were applied (de Lima, Lima-Ribeiro, Tinoco, Terribile,
213 & Collevatti, 2014; Collevatti, Terribile, Rabelo, & Lima-Ribeiro, 2015).
214 Analyses were run with 15 individuals selected from each of the sampled populations that
215 had the smallest amount of missing data (see Table S3.1 in Appendix S3). Since *B.*
216 *coccolobifolia* displayed some admixture between the CC and AS, we estimated divergence
217 times with and without the populations that displayed admixture (i.e., populations A-coHTA,
218 coCHG and coVHA). We used a python script to calculate the folded joint SFS based on the vcf
219 file from POPULATIONS (script is available on https://github.com/KnowlesLab; Papadopoulou &
220 Knowles, 2015). Only loci with a minimum coverage of 10 that were present in all selected
221 individuals were used to calculate the SFS. Divergence times were estimated excluding
222 monomorphic sites (i.e., using the *"removeZeroSFS"* option in FASTSIMCOAL2) and assuming no
223 migration between the CC and AS (this assumption is corroborated by other analyses - see
224 below); note that any violation of this assumption would result in underestimated divergence
225 times (i.e., this assumption is conservative with respect to evaluating whether the AS has had a
226 relatively short history of isolation from the CC).

227 **Chloroplast DNA data and analysis**

This article is protected by copyright. All rights reserved
Author Manuscript

228 The *trnS-trnG* (Hamilton, 1999) and *trnH-trnK* (Demesure, Sodzi, & Petit, 1995) regions
229 of chloroplast DNA were sequenced following protocols described in Resende-Moreira et al.
230 (2017) on the ABI 3730XL DNA Analyzer (ThermoFisher Scientific, Walthan, USA). A total of
231 126 sequences of *B. coccolobifolia* (including 49 sequences from Resende-Moreira et al. 2017)
232 and 116 sequences of *B. crassifolia*, with 1 to 6 individuals per population (Table S2.1 in
233 Appendix S2), were analyzed. Sequences were aligned using the software MUSCLE implemented
234 in MEGA 5.2 (Tamura et al., 2011) and all polymorphisms confirmed by visual inspection. We
235 excluded polymorphisms in microsatellites, which are prone to homoplasy, and indels and
236 inversions were recoded as one mutational step. Haplotypes were identified in DNASP 5.10
237 (Librado & Rozas, 2009) and their distribution plotted geographically to highlight haplotype
238 diversity across the range of each species, as well as the distribution of widespread versus
239 localized haplotypes. Additional analyses were performed with cpDNA data to calculate
240 diversity indices and to evaluate population structure (see Appendix S2 for details).

241 **RESULTS**

242 Measures of genomic diversity were generally similar across populations in both species
243 (Table 1), whereas cpDNA diversity varied somewhat between taxa and among populations
244 (Table S2.1 in Appendix S2), including the fixation of a single cpDNA haplotype in some
245 populations (Fig. 3), which contrasts with genomic diversity estimates (see Table 1). Populations
246 with fixed cpDNA were not disproportionately represented by AS populations (i.e., most AS
247 populations were polymorphic in cpDNA), despite their relatively small size and geographic
248 fragmentation.
249 In both species the CC populations were genetically differentiated from the AS
250 populations. STRUCTURE analyses (Fig. 4) identified separate ancestries for the CC and the AS
251 (with the exception of coHTA in *B. coccolobifolia*), which was corroborated by pairwise F_{ST} -
252 values, which were generally higher between populations from the CC and AS than among
253 populations within the respective regions (Tables S3.2 and S3.3 in Appendix S3). Likewise,
254 Procrustes analyses showed genetically distinct clusters separating the CC and AS regions,
255 except for the coHTA in *B. coccolobifolia*, which clustered with individuals from the CC (Fig.
256 4a), with a significant association between genes and geography ($t_0 = 0.770$, $P < 0.0001$ for *B.*

This article is protected by copyright. All rights reserved

Author Manuscript

257 *coccolobifolia* and $t_0 = 0.795$, $P < 0.0001$ for *B. crassifolia*). Sequential population drop-out
258 analysis showed the results from the Procrustes analyses are robust (i.e., no single population had
259 a disproportionate effect on the strength of the association between genes and geography, Table
260 S3.4 in Appendix S3). Little admixture between the CC and AS regions were detected in
261 STRUCTURE analyses, with only one AS population of each species (A-coHTA in *B.*
262 *coccolobifolia* and A-crSVA in *B. crassifolia*) showing any appreciable sign of admixture when
263 all individuals were analysed (Fig. 4). Significant IBD was also detected in a correlation analysis
264 between genetic differentiation (measured by F_{ST}) and geographic distances among populations
265 in both species ($r = 0.64$ and $r = 0.61$, $P < 0.001$, in *B. coccolobifolia* and *B. crassifolia*,
266 respectively).
267 Within the CC and AS regions, both species showed significant geographic structure.
268 Within the CC, this local substructure was evident in both the hierarchical STRUCTURE analysis
269 (Fig. 5), and the separate Procrustes analyses (Fig. 6c and 6d), which showed three genetic
270 clusters in both species. Note that for *B. coccolobifolia* one of the groups is based on only one
271 population because this is the area where the other three sampled populations appear to belong to
272 a previously unrecognized species (see Fig. 2). Separate analyses of the AS populations also
273 detected substructure, with three genetic clusters in STRUCTURE analysis (Fig. 5) and the
274 Procrustes analyses (Fig. 6a and 6b). With the primary axes of genetic variation from the
275 Principle Components Analyses separating the CC and AS regions in both species (Fig. 4; as
276 well as hierarchical structure detected in the sequential STRUCTURE analyses; Fig. 5), the
277 substructure observed within the AS and within the CC clearly accumulated after the separation
278 of AS populations from the CC.
279 Comparing geographic structuring of genomic data with cpDNA, the genomic variation
280 in *B. coccolobifolia* was generally congruent with the cpDNA (Fig. S2.1 and Tables S2.2, S2.4 in
281 Appendix S2), even with broader sampling of the cpDNA dataset (Fig. 1). In contrast, cpDNA
282 results for *B. crassifolia* differed from the genomic results and showed a lack of regional or local
283 geographic structure (Figs. 3 and S2.1, and Tables S2.3 and S2.5 in Appendix S2).
284 Divergence time estimates between the CC and AS were 119,379 and 290,541 years for
285 *B. coccolobifolia* and *B. crassifolia,* respectively. These results clearly do not support a LGM
286 divergence even considering that the shortest possible generation time of three years was used.

This article is protected by copyright. All rights reserved

Author Manuscript

287 We also note the confidence intervals surrounding the parameter estimate for *t* do not overlap
288 with the LGM. This conclusion is also robust to inclusion of admixed populations of *B.*
289 *coccolobifolia* (Table 2). Even with considering potential errors in the mutation rate, the
290 mutation rate would have to be six to twelve-times faster than the one applied here to
291 accommodate a divergence time consistent with the LGM. However, as noted in the methods,
292 mutation rates in woody plants are thought to be slower – not faster – than the one applied here,
293 so a LGM divergence is extremely unlikely.

294 **DISCUSSION**

295 The genomic distinctiveness of the CC populations from the disjunct AS and lack of any
296 regional structure that group populations from the two regions indicate the Amazonian
297 populations have evolved independently (for the most part) from the Cerrado. As such, it is
298 isolation and not recent connections (either via long-distance dispersal or expansion/retraction
299 via corridors) that dominates in this tropical biome; substantial admixture was limited to a single
300 population that borders the Cerrado (see also Buzatti, Lemos, Bueno & Lovato, 2017 for
301 localized study of another plant species from this focal area). The extent to which these results
302 are generalizable to other taxa from the Cerrado and Amazonian savannas are discussed below,
303 as is what our results suggest about the evolutionary dynamics of relictual populations in tropical
304 systems. In both discussions we advocate for a more nuanced approach to tests of the relative
305 isolation versus connections of AS populations. In addition, with reference to our own results,
306 we highlight how some refined hypotheses might provide more insight about why one process
307 might predominate over the other in particular taxa or geographic regions. Lastly, given the
308 general lack of phylogeographic studies of the AS, we reflect on the relevance of our results on
309 the processes contributing to savanna species diversity, as well as to future conservation efforts.
310

311 **Past connections versus isolation of Amazonian savannas**

312 To explain the similarity between CC and AS, three regional connections or corridors
313 were proposed: the coastal, the central Amazonian, and the Andean corridor (Haffer, 1967, 1974;
314 Silva & Bates, 2002; Webb, 1991). These corridors are hypothesized to have connected the CC

This article is protected by copyright. All rights reserved

Author Manuscript

315 and AS during waves of Pleistocene savanna expansions (Haffer, 1969; Silva & Bates, 2002),
316 possibly as recent as the Holocene (see de Freitas et al., 2001). However, our genomic data did
317 not provide strong support for the existence of such corridors in either species. Instead, analyses
318 suggest a history of restricted gene flow between the CC and AS (Fig. 4), with the AS evolving
319 in relative isolation from the CC over a history of divergence that predates the LGM (Table 2).
320 This is corroborated by palynological evidence that draws into question any recent large savanna
321 expansions that might have served as connections between the CC and AS (Colinvaux, Oliveira,
322 Moreno, Miller, & Bush, 1996; Colinvaux et al., 2001; Kastner & Goni, 2003; Mayle et al.,
323 2009). The only exception is the admixture detected in one southwestern Amazon population
324 (see Fig. 4), which is a region where joint analyses of ENMs and cpDNA data in an unrelated
325 plant also found evidence of a connection to the CC (see Buzatti et al., 2017; note this study
326 focused only on this single site so it is not possible to determine if other AS populations in the
327 species remained isolated from the CC).
328 The independent evolutionary history of CC and AS has important implications for
329 questions beyond those focused on genetic structure per se. Although *Byrsonima* species are
330 suggested to display effective long-distance dispersal (Willis et al. 2014), there is no clear
331 evidence of recent gene flow in *B. coccolobifolia* and *B. crassifolia,* indicating that long distance
332 dispersal is not common. The genetic isolation of CC from AS (Fig. 5), in addition to the
333 compositional similarities in their constituent plant communities (with over 70 woody species in
334 common; Ratter et al., 2003) suggest a more ancient common history, rather than the
335 maintenance by corridors per se. Moreover, it implies that the differences in species composition
336 between the AS and CC might reflect the cumulative loss of species in the AS (community
337 relaxation – Connor & McCoy, 1979), rather than differences in the maintenance of diversity
338 through successful/unsuccessful utilization of corridors. Additional circumstantial evidence of
339 localized extinctions rests in the observation that few Cerrado taxa are found across all AS
340 populations (Ratter et al., 2003). Alternatively, with many taxa restricted in distribution to the
341 CC, there might have been historical restrictions to expansion for many taxa such that they were
342 never part of the AS, even when Cerrado reached its broadest historical distribution. Additional
343 tests will be needed to evaluate this hypothesis. These might include testing for evidence of
344 environmental filtering or differences in the dispersal capabilities of exclusively CC taxa

This article is protected by copyright. All rights reserved
Author Manuscript

345 compared with those distributed across the AS, although no significant difference in seed
346 dispersal syndromes for species present in CC and AS has been suggested in past studies (see
347 Vieira, Aquino, Brito, Fernandes-Bulhão, & Henriques, 2002).

348 **Conflicting support for connections of the CC and AS**

349 When comparing our results to past studies purported to support hypothesized
350 connections between the CC and AS, several non-mutually exclusive explanations might account
351 for such contrasting support of the corridor hypothesis. These include: (i) differences in the
352 resolution of genetic markers, (ii) relying solely upon applications of distributional or ecological-
353 niche models, and (iii) differences among taxa in access to the corridors due to historical
354 contingencies or differences in the taxa themselves (i.e., species-specific traits). Below we
355 consider each of these explanations in turn with reference to results from our analyses of *B.*
356 *coccolobifolia* and *B. crassifolia.*

357 The genetic marker applied to test a phylogeographic hypothesis can impact the
358 likelihood that a study might find support for or refute a particular hypothesis (Knowles, 2009).
359 In particular, tests that rely upon genetic structure as evidence of isolation (e.g., when support for
360 putative corridors is based on the lack of genetic differentiation between CC and AS regions;
361 Savit & Bates, 2015) may be obscured by limited genetic resolution, including an insufficient
362 time for the sorting of ancestral polymorphism or the lack of mutation variation for detecting
363 historical divisions (Ball, Neigel, & Avise, 1990; Papadopoulou & Knowles, 2015; Thomaz et
364 al., 2017). For example, sharing of chloroplast haplotypes between CC and AS populations in *B.*
365 *coccolobifolia* and *B. crassifolia* (Fig. 3) could suggest recent gene flow or connections, as might
366 the lack of regional cpDNA clades separating the CC and AS regions (Fig. S2.1 in Appendix
367 S2). However, analysis of genomic data clearly shows that CC and AS regions are genetically
368 distinct in both species (Fig. 4). In other words, if chloroplast data by itself is going to be used to
369 refute a hypothesis of isolation, it is important to test whether the data may be consistent with a
370 history of isolation, which can be evaluated using computer simulations (see Knowles &
371 Maddison, 2002). Alternatively, and as we apply here, additional markers can be used to model
372 the divergence history of the species (as opposed to the history of a single locus; see Knowles,

Author Manuscript

This article is protected by copyright. All rights reserved

373 2009). Here the parameterized divergence models support a long history of isolation between CC
374 and AS regions that predate the LGM (Table 2).

375 For the Cerrado, evidence for the existence of corridors connecting areas north and south
376 of Amazon comes primarily from distributional data (e.g., Ávila-Pires, 1995; Nogueira &
377 Rodrigues, 2006; Silva & Bates, 2002), with some support for distinct routes of movement
378 suggested by a few phylogeographic studies (e,g., Buzatti et al., 2017; Quijada-Mascareñas et al.,
379 2007; Savit & Bates, 2015). As with concerns regarding inferences based on a single locus,
380 inferences based on ENMs alone might also be misleading (as opposed to considering ENMs
381 jointly with molecular data; reviewed in Alvarado-Serrano & Knowles, 2014). Specifically,
382 although ENMs might be used to identify possible connections, without genetic data, it is not
383 possible to test whether species actually utilized purported corridors (i.e., gene flow might not
384 have been associated with the corridors inferred from ENMs for many of the different reasons
385 discussed above). Here we advocate that, as with the interpretation of single locus data, extreme
386 caution is needed. More specifically, studies based solely on ENMs should be used to generate
387 hypotheses, but do not (by themselves) constitute evidence for supporting the corridors
388 hypothesis.

389 Assuming that such corridors existed, it is possible that some species just by chance,
390 found themselves in the right place at the right time to have access to a corridor, whereas others
391 did not. On the other hand, the lack of consistent support for corridors could also reflect
392 deterministic processes related to species-specific differences (Massatti & Knowles, 2014, 2016;
393 Papadopoulou & Knowles, 2016). Indeed, the organisms investigated in the phylogeographic
394 studies that sampled broadly the AS include different taxonomic groups (e.g., birds, snakes and
395 this study with plants) with distinct dispersal abilities and the distribution of some investigated
396 taxa are not restricted to the savannas (e.g., Savit & Bates, 2015). Although it is possible that *B.*
397 *coccolobifolia* and *B. crassifolia* differ from taxa for which future genomic analyses might show
398 corridors between the CC and AS regions, it is not obvious why *B. coccolobifolia* and *B.*
399 *crassifolia* would not have utilized corridors (if they existed). First, they are very common
400 species and widely distributed (Fig. 1), so unlike rare or patchily distributed species, they most
401 probably would have had access to any putative corridor. Second, these attributes also make it

Author Manuscript

This article is protected by copyright. All rights reserved

402 less likely that any species-specific traits would have restricted their movement (i.e., they
403 obviously can readily disperse to occupy vast areas of the Cerrado biome).

404 **Scale-specific effects of climate-induced distributional shifts?**

405 As possible remnants of a dynamic historical past, the tropical Amazonian savannas are
406 similar to relict populations in northern latitudes (Pielou, 1991). However, this dynamic history,
407 with cycles of climate-induced distributional shifts, contributes to the enigmatic nature of
408 tropical relicts and debate over their role as drivers of divergence (e.g., Capurucho et al., 2013).
409 By rejecting hypothesized periods of connectivity between CC and AS through putative
410 expansions during glacial-interglacial periods (Prance, 1996; Silva & Bates, 2002), our study
411 raises some intriguing questions about divergence of Cerrado species. Here we make the
412 argument that connections forged during cycles of expansion, while not extensive enough to
413 support corridors between the CC and AS, may have played an important role in divergence
414 within the CC and within the AS. In other words, geographic scale determines whether climatic
415 oscillations promote connections. Likewise, we note that the existence of regional structure itself
416 within both the CC and AS, suggests a limit on the level of connectiveness across populations in
417 the past (otherwise, the regional structure would have been lost, and the only structure would be
418 the population level structure that was also observed; see Fig. 5).

419 What might limit the role of climate-induced distributional shifts at the larger scale – that
420 is, why were distributional shifts not associated with connections between the CC and the AS?
421 The most obvious answer is that the extent of savanna expansion (or conversely forest
422 contraction) may have been more limited than previous proposals. For example, suggestions of a
423 fairly stable forest during the LGM, especially for the western part of Amazon (e.g., Bush,
424 Silman, & Urrego, 2004; Cheng et al., 2013; Colinvaux, Oliveira, & Bush, 2000), offer an
425 alternative to Haffer's (1969) scenario of forest fragmentation during glacial periods. In addition,
426 recent isotopic data sampled from the Amazon dry corridor (i.e., an area of current lower
427 precipitation within the Amazon, Haffer, 1969) suggests forest physiognomies during the LGM
428 consistent with the maintenance of rainforest (Wang et al., 2017), and/or its replacement by dry-
429 forest habitats, instead of savanna (Bush, 2017; Pennington, Prado, & Pendry, 2000).

Author Manuscript

This article is protected by copyright. All rights reserved

430 Within the CC, past phylogeographic studies of plants have documented an east-west
431 split that is generally concordant with *Byrsonima* (Fig. 6c and 6d) (reviewed in Leal, Palma-
432 Silva, & Pinheiro, 2016). Likewise, a regional genetic structuring of individuals sampled in
433 southern, central-northern and north-western portions of Cerrado have been observed in other
434 organisms, including frogs and lizards (Prado, Haddad, & Zamudio, 2012; Santos, Nogueira,
435 Giugliano, & Colli, 2014). This spatial concordance across studies highlights how these CC
436 communities may be shaped by similar historical processes. Similarly, among the AS
437 populations, regional divergence is clear, as is differentiation among individual populations
438 (Figs. 5 and 6). However, it is not clear what accounts for the observed regional structure of AS
439 populations of both *Byrsonima* species because unfortunately, unlike the CC, there is extremely
440 limited data in terms of genetic analyses of broadly sampled AS populations (in fact, we are not
441 aware of any other studies besides ours). We note that in other types of open habitat (e.g., birds
442 inhabiting white sand vegetation), genetic data provides evidence of recent population expansion
443 during the late Pleistocene (Capurucho et al., 2013; Matos et al., 2016), suggesting that the
444 connections we propose among AS populations based on regional structuring of genetic variation
445 may not be an anomaly. It is clear that future research will be charting new directions about the
446 drivers of divergence within the CC and AS as the focus shifts from one built on a history of
447 corridors connections, to the independent evolutionary trajectories of the CC and AS.

448 **Conservation of Cerrado and Amazonian savannas**

449 Despite the high endemism and species diversity, the Cerrado is rapidly being loss (less
450 than 20% remains undisturbed; Strassburg et al., 2017), especially with the expansion of
451 agriculture, cattle ranching, and charcoal production, and conservation of the Cerrado biome has
452 received little attention. Although rates of loss have decreased over the last several years (i.e.,
453 since 2010), we are nevertheless loosing Cerrado faster than Amazon Forest (Françoso et al.,
454 2015).

455 Given the extent of the biome, covering 2 million km^2, assessments of genetic diversity
456 and population structure arguably could provide important guidance in conservation efforts. Yet,
457 with relatively sparse geographic sampling, and limited genomic study, such information is
458 rarely considered in conserving this highly threatened biome. Analyses of broadly distributed

Author Manuscript

This article is protected by copyright. All rights reserved

459 taxa in particular, like *B. coccolobifolia* and *B. crassifolia*, could be used to devise conservation
460 strategies that protect not only the constituent taxa, but also preserve diversity generating
461 processes (see Moritz, 2002). For example, our study revealed an unexpected cryptic species in
462 *B. coccolobifolia* from the central and northern areas of the CC, an area reportedly of high
463 species richness (Ratter et al., 2003). Other phylogeographic studies on Cerrado trees suggest the
464 highest genetic diversity occurs in central areas of the Cerrado as well (e.g., Collevatti, Castro,
465 Lima, & Telles, 2012; Novaes, Ribeiro, Lemos-Filho, & Lovato, 2010; Souza, Collevatti, Lima-
466 Ribeiro, Lemos-Filho, & Lovato, 2017); however, many of these have limited sampling of
467 northern areas. Our results, as those with more extensive sampling (Collevatti, Terribile, Diniz,
468 & Lima-Ribeiro, 2015; Ribeiro, Lemos, Buzatti, Lovato, & Heuertz, 2016) have revealed high
469 genetic diversity in north-eastern plant populations, highlighting the importance of these areas to
470 conservation efforts of the Cerrado, especially since these areas are part of an expanding
471 agricultural frontier.

472 Perhaps most importantly, our broad sampling of plant species from the AS identifies a
473 number of factors relevant to developing conservation priorities for the AS. First, we show that
474 these populations display levels of genetic diversity similar to CC populations, which is
475 somewhat reassuring about their general health from a genetic prospective (i.e., they do not show
476 disproportionately depressed levels of diversity; see Fig. 3 and Table 1). However, their apparent
477 genetic isolation does place them at substantial risk (Fig. 4). Moreover, these populations
478 arguably should be considered as unique Cerrado environments in conservation efforts of the
479 biome, given their relatively long isolated history from the CC (see Table 2). Even though most
480 AS display much less species diversity than the CC (but see Ratter et al., 2003 for exceptions),
481 some AS contain more than 250 plant taxa (Miranda, Absy, & Rebelo, 2003; Sanaiotti, 1997), in
482 addition to vulnerable and endemic species of birds, reptiles, amphibians and plants (Barbosa,
483 Campos, Pinto, & Fearnside, 2007; Carvalho, 1997; França, Mesquita, & Colli, 2006; Rocha &
484 Miranda, 2014). It is also important to note that the number of species in the AS most likely is
485 larger considering that these areas are highly understudied (Carvalho & Mustin, 2017). Lastly,
486 AS are under particularly high anthropogenic disturbance because they are misleadingly
487 considered as natural pastures in an environment largely dominated by forest (Miranda et al.,

Author Manuscript

This article is protected by copyright. All rights reserved

488 2003), making immediate attention as conservation units an imperative (Carvalho & Mustin,
489 2017).

490 **Conclusions**

491 Our results show independent evolution of the CC and AS populations of both broadly
492 distributed tree species studied here (*B. coccolobifolia* and *B. crassifolia*), casting doubt on the
493 importance of corridors in structuring Cerrado plant communities. In the context of
494 understanding the evolutionary history of AS populations in particular, it is possible that climatic
495 change in the tropics, and/or differences in the traits of the species themselves, might make
496 certain corridors more or less accessible during different geologic periods (Silva & Bates, 2002;
497 Wüster et al., 2005), but careful consideration of this hypothesis will require expanding the
498 dataset to other broadly distributed taxa. Specifically, our genomic data suggests a relatively long
499 history of isolation between the CC and AS regions that predates the LGM, as well as population
500 structuring of genetic variation within regions in both species. The contrast between genetic
501 structure of genomic versus chloroplast datasets also highlights the need for cautious
502 interpretation of what constitutes evidence for the corridor hypothesis. Our findings suggest that
503 methodology, not biology, may contribute to some of the differences in support for the corridor
504 hypothesis reported across studies. Lastly, as a biodiversity hotspot, these results have direct
505 implications for diversification in the Cerrado, as well as its conservation, especially given
506 extensive and ongoing habitat destruction (Carvalho & Mustin, 2017; Mittermeier et al., 2004).

Author Manuscript

This article is protected by copyright. All rights reserved

TABLES

Table 1. Number of individuals sampled, *N*, and estimates of genetic diversity per population of *Byrsonima coccolobifolia* and *B. crassifolia* (see Figure 1 for distributional map of sampled populations). Amazonian savanna populations are identified with an A preceding the population label. Estimates of genetic diversity per population are based on all polymorphic nucleotide positions of filtered genomic data; H_{OBS}, observed heterozygosity; H_{EXP}, expected heterozygosity; π, nucleotide diversity.

Pop	*N*	H_{OBS}	H_{EXP}	π
Byrsonima coccolobifolia				
A-coBON	4	0.053	0.049	0.058
A-coBVT	5	0.056	0.052	0.059
A-coCNE	6	0.059	0.054	0.059
A-coMCP	6	0.059	0.056	0.063
A-coSAN	5	0.057	0.053	0.060
A-coHTA	5	0.063	0.060	0.067
coCAI	5	0.065	0.064	0.071
coVHA	6	0.064	0.065	0.072
coCHG	5	0.064	0.060	0.068
coCGD	6	0.062	0.062	0.068
coJPO	6	0.067	0.068	0.075
coUDA	6	0.063	0.064	0.070
coPRA	5	0.068	0.062	0.070
Byrsonima crassifolia				
A-crATA	5	0.060	0.055	0.063
A-crPAC	1	0.057	0.029	0.057
A-crCNE	5	0.065	0.067	0.076
A-crMCP	6	0.067	0.070	0.078
A-crSAN	6	0.065	0.068	0.075

Author Manuscript

This article is protected by copyright. All rights reserved

A-crSVA	6	0.067	0.071	0.079
crPRI	2	0.065	0.051	0.070
crAGN	5	0.055	0.055	0.062
crSLA	6	0.056	0.055	0.061
crMAG	6	0.059	0.061	0.068
crFOR	5	0.064	0.064	0.072
crCHG	5	0.058	0.059	0.067
crCVE	2	0.059	0.044	0.062
crPPB	2	0.056	0.040	0.060
crITI	6	0.057	0.059	0.066

Author Manuscript

This article is protected by copyright. All rights reserved

Table 2. Divergence time estimates (assuming a minimum of a three year generation time; see methods for details) and other demographic parameters for each *Byrsonima* species based on the model of divergence between the Amazonian savannas (AS) and central Cerrado (CC) regions using FASTSIMCOAL2. Specifically, we show results for divergence time, T_{DIV}, ancestral effective population size, N_{ANC}, effective population size for AS, N_{AS}, and number of loci used to calculate the folded joint site frequency spectrum (SFS). Confidence intervals based on 100 parametric boostraps are shown in parentheses. Note that effective population size of the CC (N_{CC}) was calculated directly from the empirical data (i.e., was a fixed parameter in the model) to improve the accuracy of the other parameters estimated from the SFS (following the recommendations for the program; see Excoffier & Foll, 2011).

Species	Loci	T_{DIV} (years)	N_{ANC}	N_{CC}	N_{AS}
Byrsonima coccolobifolia (all populations)	2285	109,611 (87,432-143,886)	100,400 (60,069-176,369)	978,571	343,875 (263,662-502,760)
Byrsonima coccolobifolia (excluding admixed populations A-coHTA, coCHG and coVHA)	1945	119,379 (96,195-169,311)	56,282 (31,697-102,681)	992,857	210,635 (162,674-322,062)
Byrsonima crassifolia	1032	290,541 (240,696-355,920)	117,638 (80,184-166,741)	1,000,000	399,628 (331,352-498,123)

Author Manuscript

This article is protected by copyright. All rights reserved

FIGURES

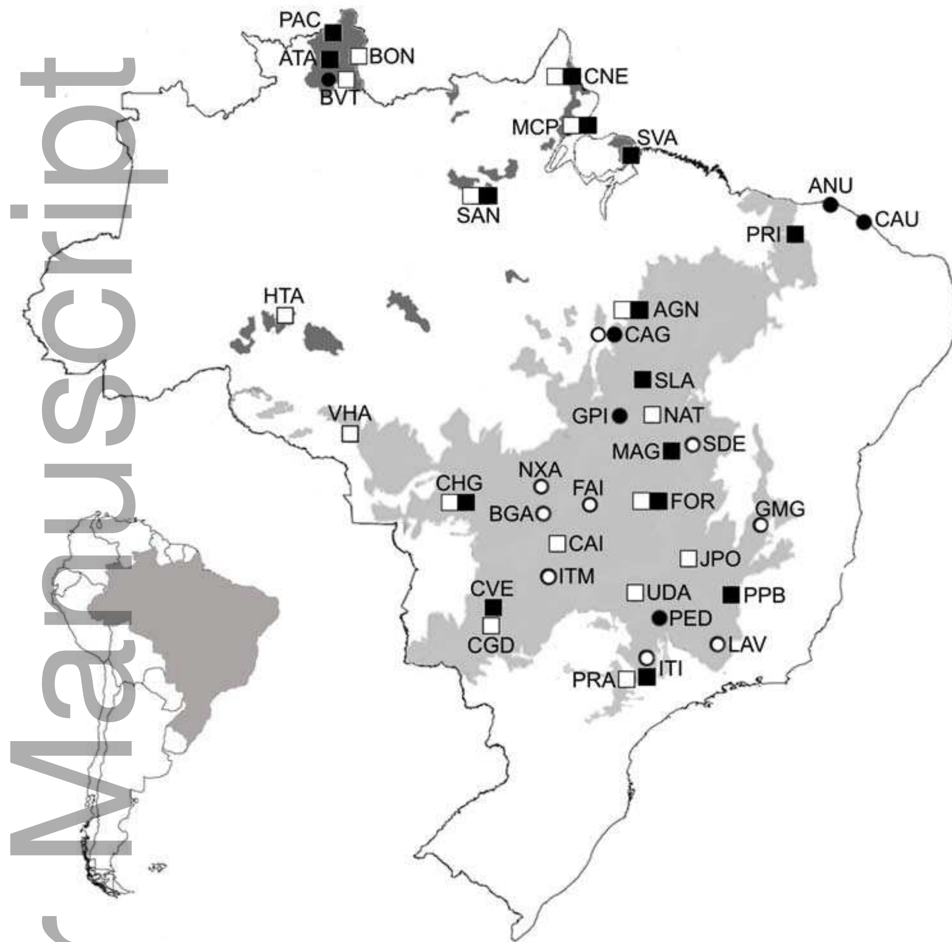
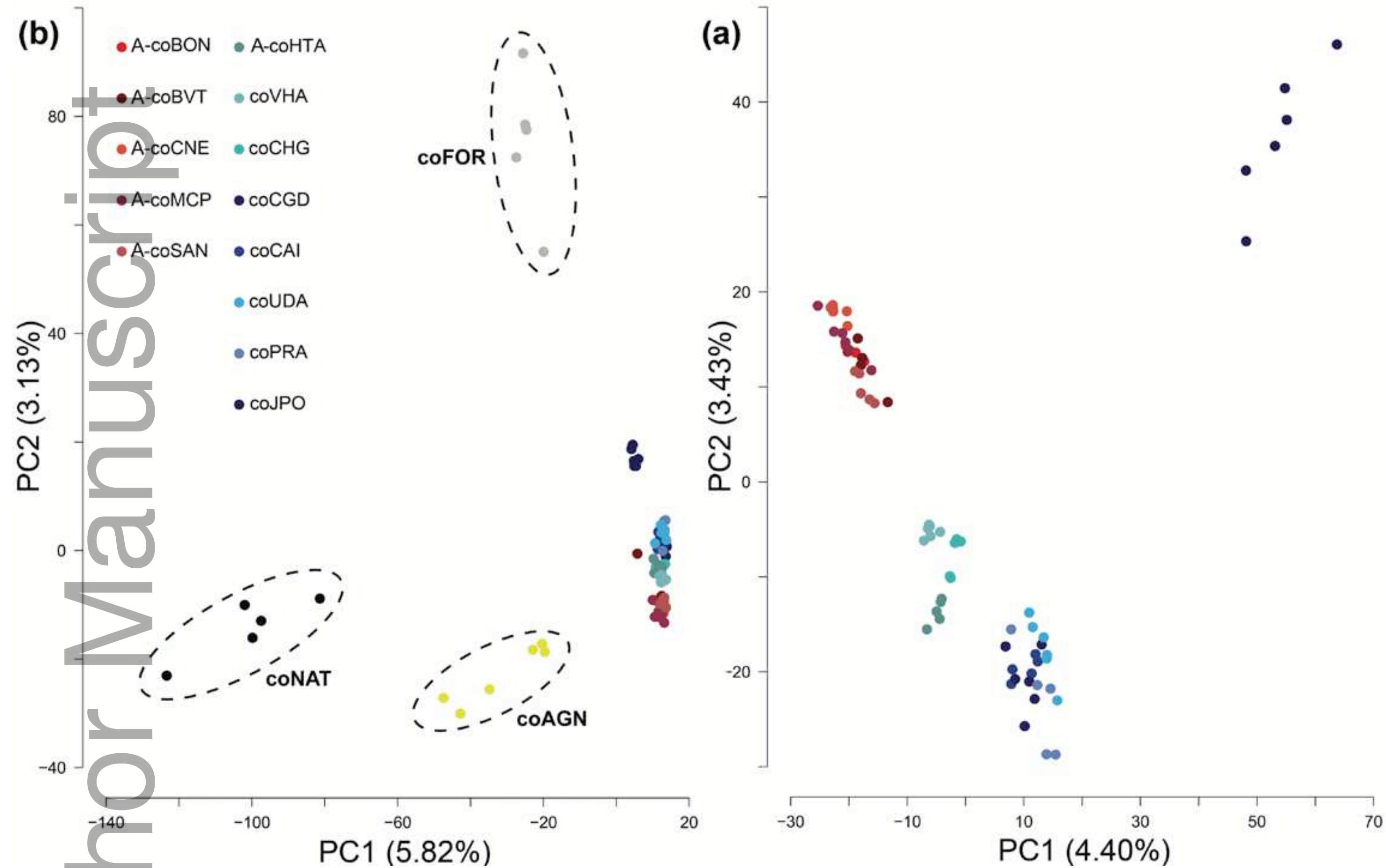


Figure 1. Geographic location of sampled *B. coccolobifolia* and *B. crassifolia* populations (white and black symbols, respectively) across the Cerrado (light grey) and Amazonian savannas (dark grey). Populations with both genomic and cpDNA sequences are marked by squares, whereas those with only cpDNA are marked by circles.

Author Manuscript

This article is protected by copyright. All rights reserved



Author Manuscript

Figure 2. Principle Components Analysis (PCA) of *Byrsonima coccolobifolia* including (a) and excluding (b) populations that revealed cryptic genetic diversity indicative of potentially different species (i.e., the three divergent sampled populations: coAGN, coNAT and coFOR). The amount of variation explained by each axis is given in parentheses and colors indicate population identity.

This article is protected by copyright. All rights reserved

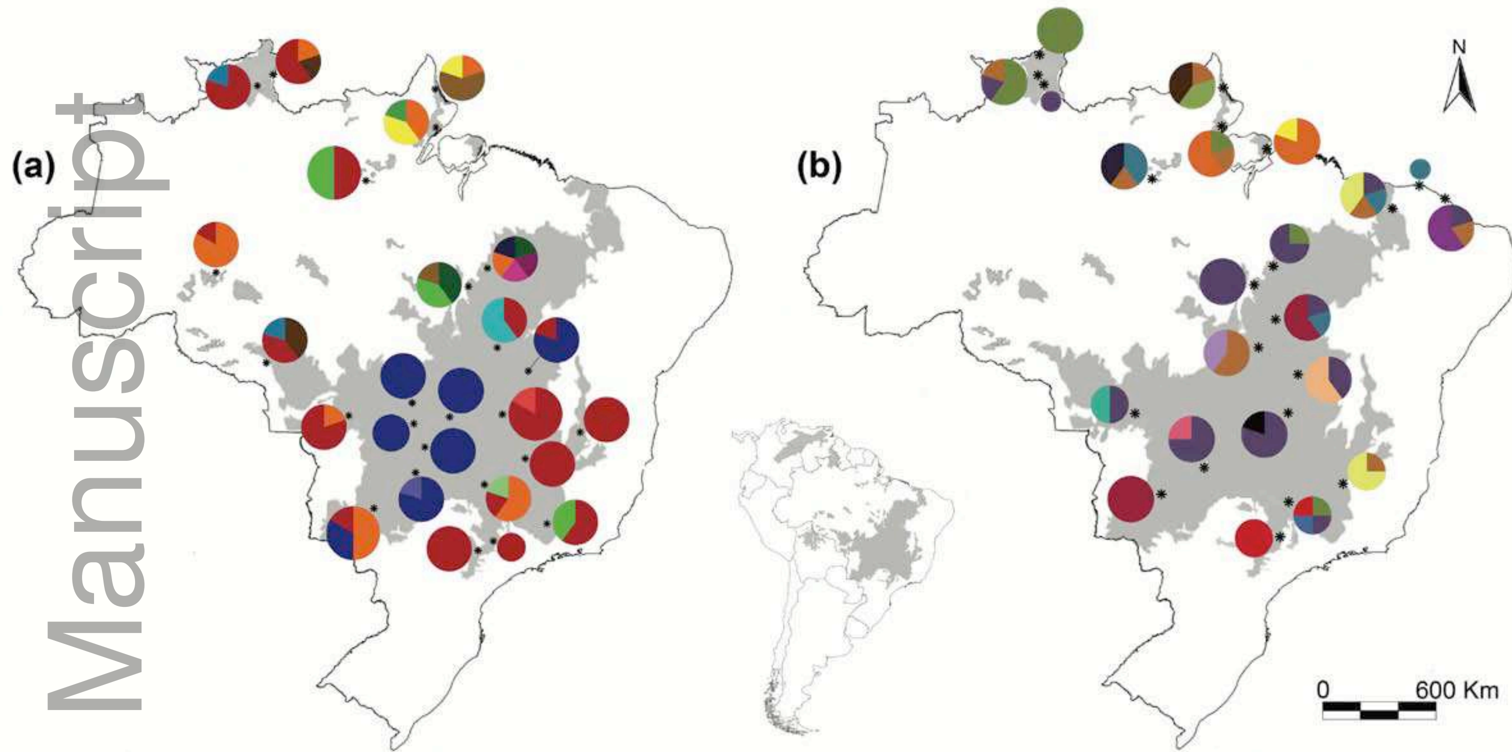
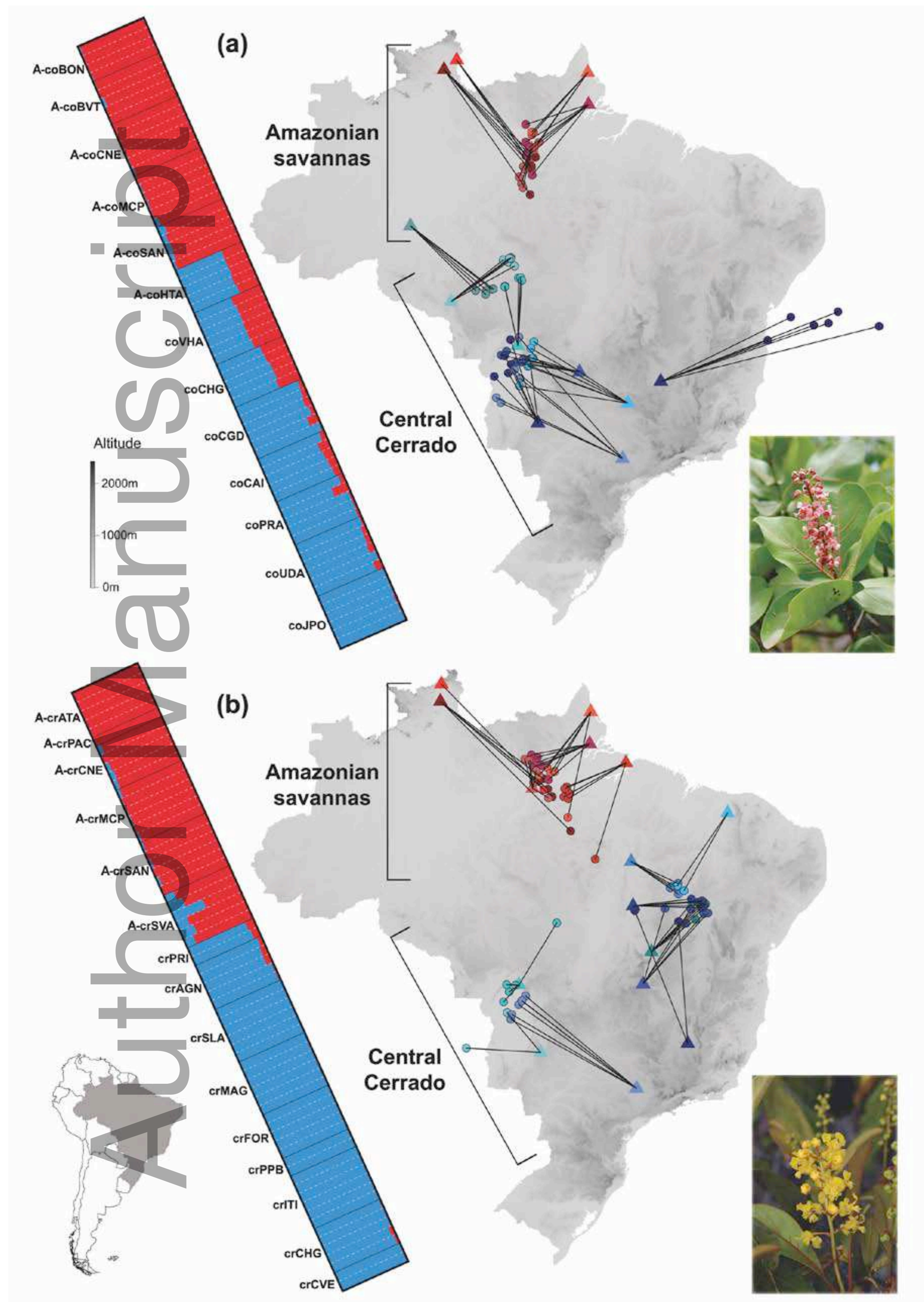


Figure 3. Geographic distribution of cpDNA haplotypes (sampling location is marked by small black dot) of *Byrsonima coccolobifolia* (a) and *B. crassifolia* (b), with each distinct haplotype represented by a different color and the number of individuals sampled in each population indicated by the size of the circles. Shaded areas approximate the distribution of the Cerrado (both the central Cerrado and Amazonian savannas).

Author Manuscript

This article is protected by copyright. All rights reserved

Author Manuscript



This article is protected by copyright. All rights reserved

Figure 4. Population structure of *Byrsonima coccolobifolia* (a) and *B. crassifolia* (b). Barcharts show the most probable number of groups (*K*) according to STRUCTURE results for each species as different colors along with Procrustes analysis of genetic variation. Each individual in the barchart is demarcated by a white dashed line, and the posterior probability of each individual's ancestry is depicted as the proportion of each color per individual, whereas populations are labelled and separated by black lines. In each map, the position of an individual in genomic space (shown as circles) relative to sampled locality (show as triangles) is indicated (with individuals color coded by population). The lines connecting individuals (circles) to localities (triangles) indicate the deviation of an individual genetically from expectations based on their geographic location (i.e., departures from isolation by distance). Amazonian savanna populations are identified with an A. Photographs of *B. coccolobifolia* and *B. crassifolia* were provided by Maurício Mercadante and Daniel Nickrent (source: http://www.phytoimages.siu.edu), respectively.

Author Manuscript

This article is protected by copyright. All rights reserved

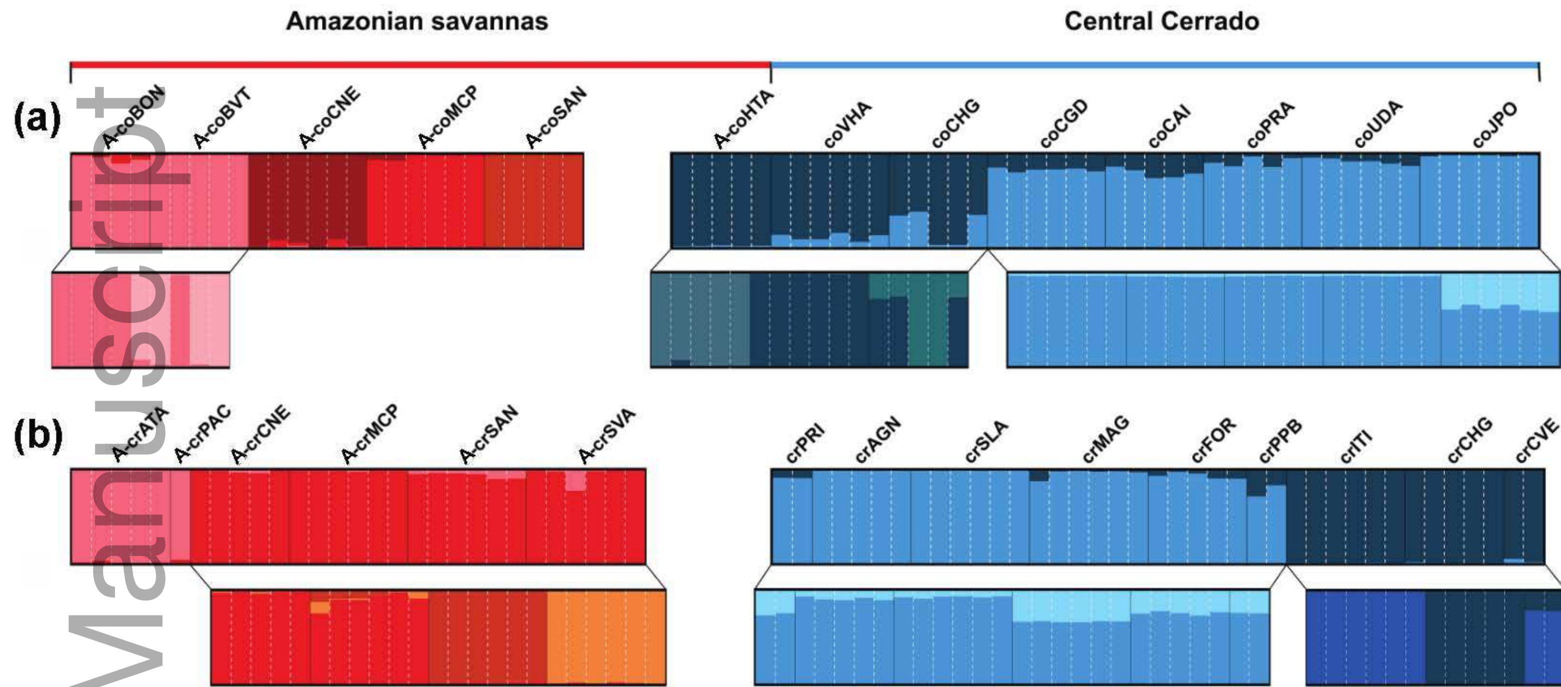


Figure 5. Hierarchical population structure of *Byrsonima coccolobifolia* (a) and *B. crassifolia* (b) based on sequential, and separate, STRUCTURE analyses of Amazonian savanna and central Cerrado datasets (i.e., data subsets identified from global analyses; see Fig. 4). The most probable number of groups (*K*) is displayed as different colours in each plot, with populations marked by thin black lines, and white lines demarcating sampled individuals with the posterior probability of belonging to each cluster depicted as the proportion of each colour in the bar. Amazonian savanna populations are identified with an A.

Author Manuscript

This article is protected by copyright. All rights reserved

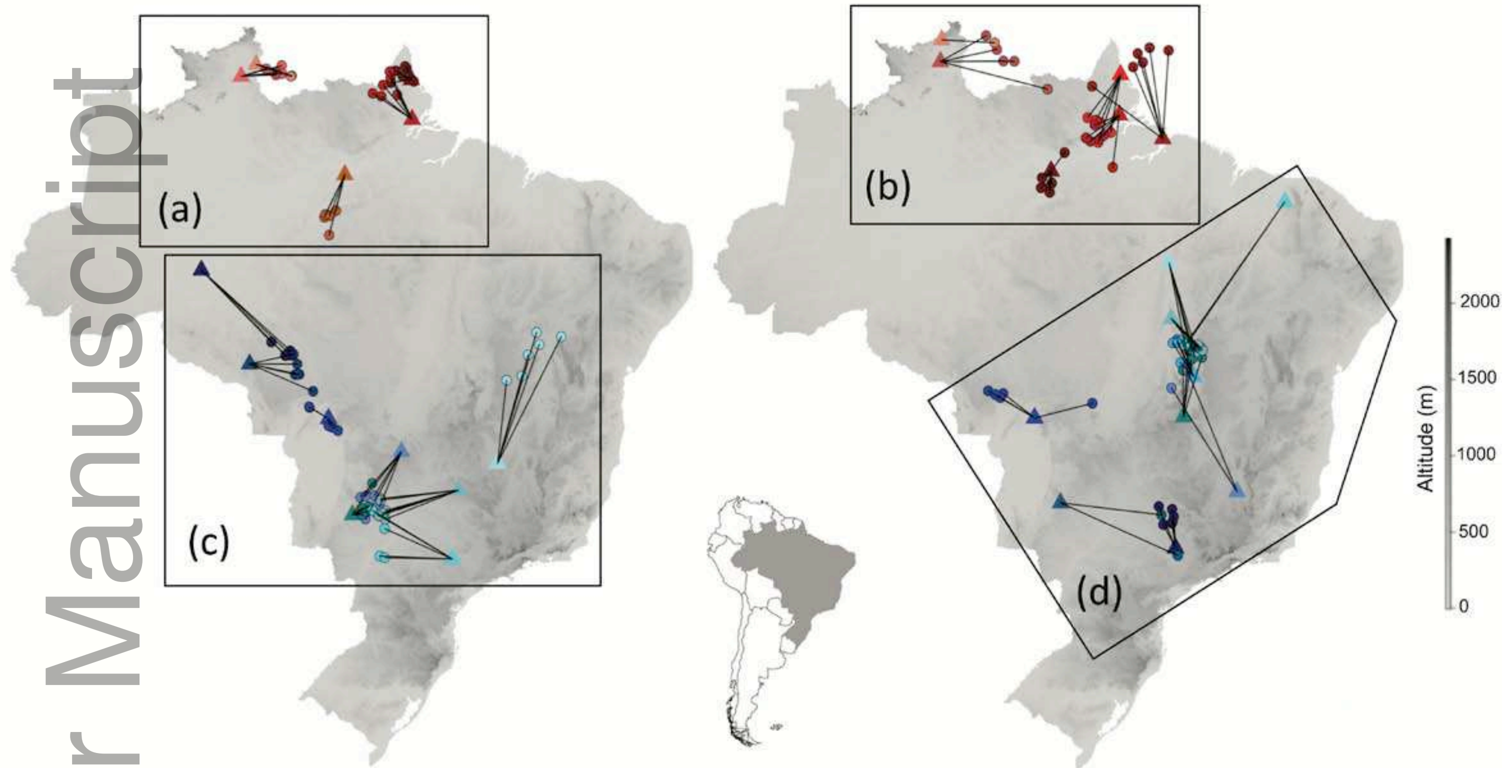


Figure 6. Plots of Procrustes analyses carried out separately on regional datasets of the Amazonian savanna (a and b) and central Cerrado populations (c and d) of *Byrsonima coccolobifolia* and *B. crassifolia*, respectively. The lines connect individuals (shown as circles) to sampling location (shown as triangles) indicate deviations from the expected pattern of genetic variation based on isolation by distance, where longer lines indicate greater departures from expectations based on where an individual was sampled geographically. Colors indicate population identity.

Author Manuscript

This article is protected by copyright. All rights reserved

REFERENCES

Alvarado-Serrano, D. F., & Knowles, L. L. (2014). Ecological niche models in phylogeographic studies: applications, advances and precautions. *Molecular Ecology Resources, 14*, 233-248.

Anderson, W. R. (1981). Malpighiaceae, The Botany of the Guayana Highland, Part XI, vol. 32, Memoirs of the New York Botanical Garden 32, 21-305, New York, NY.

Anderson, W. R. (1983) *Byrsonima crassifolia* (Nance, Nancite, Shoemaker's Tree). In: Janzen, D. H. (Ed.) Costa Rican Natural History, pp. 202-204. Chicago, IL: University Chicago Press.

Anderson, W. R.; Anderson, C., & Davis, C. C. (2006) Malpighiaceae. http://herbarium.lsa.umich.edu/malpigh/index.html Accessed on: 01 May 2017.

Ávila-Pires, T. C. S. (1995). Lizards of Brazilian Amazonia (Reptilia: Squamata). *Zoologische Verhandelingen, 299*, 1-706.

Avise, J.C. (2004). *Molecular markers, natural history and evolution* (2nd ed.). Sunderland, MA: Sinauer Associates.

Ball, R. M., Neigel, J. E., & Avise, J. C. (1990). Gene genealogies within the organismal pedigrees of random-mating populations. *Evolution, 44*, 360-370.

Barbosa, R. I., Campos, C., Pinto, F., & Fearnside, P. M. (2007). The "Lavrados" of Roraima: biodiversity and conservation of Brazil's Amazonian savannas. *Functional Ecosystems & Communities, 1*, 29-41.

Behling, H., & Hooghiemstra, H. (2001). Neotropical savanna environments in space and time: late Quaternary inter-hemispheric comparisons. In V. Markgraf (Ed.), *Interhemispheric Climate Linkages* (pp. 307-323). San Diego, CA: Academic Press.

Benezar, R. M. & Pessoni, L.A. (2006) Floral biology and the reproductive system of *Byrsonima coccolobifolia* (Kunth) in an Amazonian savanna. *Acta Amazonica*, *36*,159-68.

Bonatelli, I. A. S., Perez, M. F., Peterson, A. T., Taylor, N. P., Zappi, D. C., Machado, M. C., . . . Moraes, E. M. (2014). Interglacial microrefugia and diversification of a cactus species complex: phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. *Molecular Ecology, 23*, 3044-3063.

Bueno, M. L., Pennington, R. T., Dexter, K. G., Kamino, L. H. Y., Pontara, V., Neves, D. M., . . . Oliveira-Filho, A. T. (2017). Effects of Quaternary climatic fluctuations on the distribution of Neotropical savanna tree species. *Ecography, 40*, 403-414.

Buerkle, C. A. & Gompert, Z. (2013) Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology, 22,* 3028-3035.

Author Manuscript

Bush, M. B. (2017). The resilience of Amazonian forests. *Nature, 541*, 167-168.

Bush, M. B., Silman, M. R., & Urrego, D. H. (2004). 48,000 years of climate and forest change in a biodiversity hotspot. *Science, 303*, 827-829.

Buzatti, R. S. D., Lemos, J. P., Bueno, M. L., & Lovato, M. B. (2017). Multiple Pleistocene refugia in the Brazilian cerrado: evidence from phylogeography and climatic nichemodelling of two *Qualea* species (Vochysiaceae). *Botanical Journal of the Linnean Society, 185*, 307-320.

Capurucho, J. M. G., Cornelius, C., Borges, S. H., Cohn-Haft, M., Aleixo, A., Metzger, J. P., & Ribas, C. C. (2013). Combining phylogeography and landscape genetics of *Xenopipo atronitens* (Aves: Pipridae), a white sand campina specialist, to understand Pleistocene landscape evolution in Amazonia. *Biological Journal of the Linnean Society, 110*, 60-76.

Carvalho, A. M. (1997). A synopsis of the genus *Dalbergia* (Fabaceae: Dalbergieae) in Brazil. *Brittonia, 49*, 87-109.

Carvalho, W. D., & Mustin, K. (2017). The highly threatened and little known Amazonian savannahs. *Nature Ecology & Evolution, 1*,100.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology, 22*, 3124-3140.

Cheng, H., Sinha, A., Cruz, F. W., Wang, X. F., Edwards, R. L. A., d'Horta, F. M., . . . Auler, A. S. (2013). Climate change patterns in Amazonia and biodiversity. *Nature Communications, 4*, 1411.

Colinvaux, P. A., Oliveira, P. E., & Bush, M. B. (2000). Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses. *Quaternary Science Reviews, 19*, 141-169.

Colinvaux, P. A., Oliveira, P. E., Moreno, J. E., Miller, M. C., & Bush, M. B. (1996). A long pollen record from lowland Amazonia: Forest and cooling in glacial times. *Science, 274*, 85-88.

Colinvaux, P. A., Irion, G., Rasanen, M. E., Bush, M. B., & de Mello, J. (2001). A paradigm to be discarded: Geological and paleoecological data falsify the Haffer & Prance refuge hypothesis of Amazonian speciation. *Amazoniana, 16*, 609-646.

Collevatti, R. G., Castro, T. G., Lima, J. S., & Telles, M. C. (2012) Phylogeography of *Tibouchina papyrus* (Pohl) Toledo (Melastomataceae), an endangered tree species from rocky savannas, suggests bidirectional expansion due to climate cooling in the Pleistocene. *Ecology and Evolution, 2*, 1024-1035.

Author Manuscript

This article is protected by copyright. All rights reserved

Collevatti, R. G., Terribile, L. C., Rabelo, S. G., & Lima-Ribeiro, M. S. (2015). Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a Neotropical savanna tree species in the deeper Quaternary. *Frontiers in Plant Sciences, 6,* 653.

Collevatti, R. G., Terribile, L. C., Diniz-Filho, J. A. F., & Lima-Ribeiro, M. S. (2015). Multi-model inference in comparative phylogeography: an integrative approach based on multiple lines of evidence. *Frontiers in Genetics, 6*, 31.

Connor, E. F., & McCoy, E. D. (1979). Statistics and biology of the species-area relationship. *American Naturalist, 113*, 791-833.

de Freitas, H. A., Pessenda, L. C. R., Aravena, R., Gouveia, S. E. M., Ribeiro, A. D., & Boulet, R. (2001). Late Quaternary vegetation dynamics in the southern Amazon basin inferred from carbon isotopes in soil organic matter. *Quaternary Research, 55*(1), 39-46.

de Lima, N. E., Lima-Ribeiro, M. S., Tinoco, C. F., Terribile, L. C., & Collevatti, R. G. (2014). Phylogeography and ecological niche modelling, coupled with the fossil pollen record, unravel the demographic history of a Neotropical swamp palm through the Quaternary. *Journal of Biogeography, 41*, 673-686.

Demesure, B., Sodzi, N., & Petit, R. J. (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology, 4*, 129-131.

Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software, 22*, 1-20.

Earl, D. A., & Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources, 4*, 359-361.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *Plos Genetics, 9*, e1003905.

Excoffier, L., & Foll, M. (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics, 27*, 1332-1334.

França, F. G. R., Mesquita, D. O., & Colli, G. R. (2006). A checklist of snakes from Amazonian savannas in Brazil, housed in the Coleção Herpetológica da Universidade de Brasília, with new distribution records. *Occasional Papers Oklahoma Museum of Natural History, 17*, 1-13.

Author Manuscript

This article is protected by copyright. All rights reserved

Françoso, R. D., Brandão, R., Nogueira, C. C., Salmona, Y. B., Machado, R. B., & Colli, G. R. (2015). Habitat loss and the effectiveness of protected areas in the Cerrado biodiversity hotspot. *Natureza & Conservação, 13*, 35-40.

Haffer, J. (1967). Zoogeographical notes on the "nonforest" lowland bird faunas of northwestern South America. *Hornero, 10*, 315-333.

Haffer, J. (1969). Speciation in Amazonian Forest Birds. *Science, 165*, 131-137.

Haffer, J. (1974). Avian speciation in tropical South America, with a systematic survey of the toucans (Ramphastidae) and jacamars (Galbulidae). *Nuttall Orthological Club, 14*, 1-390.

Hamilton, M. B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecololgy, 8*, 521-523.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 359*, 183-195.

Huang, H. T., & Knowles, L. L. (2016). Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology, 65*, 357-365.

Jaramillo, C., Hoorn, C., Silva, S. A. F., Leite, F., Herrera, F., Quiroz, L., . . . Antonioli, L. (2010). The origin of the modern Amazon rainforest: implications of the palynological and palaeobotanical record. In C. Hoorn & F. P. Wesselingh (Eds.), *Amazonia: Landscape and Species Evolution: a look into the past*, (pp. 317-334). Hoboken, NJ: Wiley-Blackwell.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*, 1403-1405.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics, 27*, 3070-3071.

Jombart, T., Pontier, D., & Dufour, A. B. (2009). Genetic markers in the playground of multivariate analysis. *Heredity, 102*, 330-341.

Kastner, T. P., & Goni, M. A. (2003). Constancy in the vegetation of the Amazon basin during the late Pleistocene: Evidence from the organic matter composition of Amazon deep sea fan sediments. *Geology, 31*, 291-294.

Knowles, L. L. (2009). Statistical phylogeography. *Annual Review of Ecology Evolution and Systematics, 40*, 593-612.

Knowles, L. L., & Maddison, W. P. (2002). Statistical phylogeography. *Molecular Ecology, 11*, 2623-2635. doi:10.1046/j.1365-294X.2002.01637.x

Author Manuscript

This article is protected by copyright. All rights reserved

Knowles, L. L., & Massatti, R. (2017). Distributional shifts - not geographic isolation - as a probable driver of montane species divergence. *Ecography, 40*, 1475-1485.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources, 15*, 1179-1191.

Leal, B. S. S., Palma-Silva, C., & Pinheiro, F. (2016). Phylogeographic Studies Depict the Role of Space and Time Scales of Plant Speciation in a Highly Diverse Neotropical Region. *Critical Reviews in Plant Sciences, 35*, 215-230.

Ledru, M. P. (2002). Late Quaternary history and evolution of the cerrados as revealed by palynological records. In P. S. Oliveira & R. J. Marquis (Eds.), *The Cerrados of Brazil: ecology and natural history of a Neotropical savanna* (pp. 33-50). New York, NY: Columbia University Press.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics, 25*, 1451-1452.

Massatti, R., & Knowles, L. L. (2014). Microhabitat differences impact phytogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution, 68*, 2833-2846.

Massatti, R., & Knowles, L. L. (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Molecular Ecology, 25*, 3974-3986.

Matos, M. V., Borges, S. H., d'Horta, F. M., Cornelius, C., Latrubesse, E., Cohn-Haft, M., & Ribas, C. C. (2016). Comparative phylogeography of two bird species, *Tachyphonus phoenicius* (Thraupidae) and *Polytmus theresiae* (Trochilidae), specialized in Amazonian white-sand vegetation. *Biotropica, 48*, 1

Mayle, F. E., Burn, M. J., Power, M., & Urrego, D. H. (2009). Vegetation and fire at the Last Glacial Maximum in Tropical South America. In F. Vimeux, F. Sylvestre & M. Khodri (Eds.), *Past Climate Variability in South America and Surrounding Regions* (pp. 89-112). Dordrecht, Netherlands: Springer.

Migliore, J., Baumel, A., Juin, M., Fady, B., Roig, A., Duong, N., & Medail, F. (2013). Surviving in mountain climate refugia: new insights from the genetic diversity and structure of the relict shrub *Myrtus nivellei* (Myrtaceae) in the Sahara desert. *Plos One, 8*, e73795.

Miranda, I. S., Absy, M. L., & Rebelo, G. H. (2003). Community structure of woody plants of Roraima Savannahs, Brazil. *Plant Ecology, 164*, 109-123.

Author Manuscript

This article is protected by copyright. All rights reserved

Mittermeier, R. A., Gil, P. R., Hoffman, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., . . . Fonseca, G. A. B. (2004). *Hotspots revisited: Earth's biologically richest and most threatened terrestrial ecoregions*. Mexico City: CEMEX.

Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology, 51*, 238-254.

Nogueira, C., & Rodrigues, M. T. (2006). The genus *Stenocercus* (Squamata: Tropiduridae) in extra-Amazonian Brazil, with the description of two new species. *South American Journal of Herpetology, 12*, 149-165.

Novaes, R. M. L., Lemos-Filho, J. P., Ribeiro, R. A., & Lovato, M. B. (2010). Phylogeography of *Plathymenia reticulata* (Leguminosae) reveals patterns of recent range expansion towards northeastern Brazil and southern Cerrados in Eastern Tropical South America. *Molecular Ecology, 19*, 985-998.

Novaes, R. M. L., Rodrigues, J. G., & Lovato, M. B. (2009). An efficient protocol for tissue sampling and DNA isolation from the stem bark of Leguminosae trees. *Genetics and Molecular Research, 8*, 86-96.

Oksanen, J., Blanchet, F. G., Friendly, M., Kinda, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2017). vegan: Community Ecology Package (Version 2.4-4). R package. Retrieved from https://CRAN.R-project.org/package=vegan. Accessed 2017-07-06.

Oliveira-Filho, A. T. (2017). NeoTropTree, flora arbórea da região Neotropical: um banco de dados envolvendo biogeografia, diversidade e conservação. Retrieved from http://www.icb.ufmg.br/treeatlan. Accessed 2017-01-06.

Ornelas, J. F., Ruiz-Sanchez, E., & Sosa, V. (2010). Phylogeography of *Podocarpus matudae* (Podocarpaceae): pre-Quaternary relicts in northern Mesoamerican cloud forests. *Journal of Biogeography, 37*, 2384-2396.

Ossowski, S., Schneeberger, K., Lucas-Lledo, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., . . . Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science, 327*, 92-94.

Papadopoulou, A., & Knowles, L. L. (2015). Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution, 69*, 1501-1517.

Author Manuscript

This article is protected by copyright. All rights reserved

Papadopoulou, A., & Knowles, L. L. (2016). Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 8018-8024.

Pascual, R., & Jaureguizar, E. O. (1990). Evolving climates and mammal faunas in Cenozoic South-America. *Journal of Human Evolution, 19*, 23-60.

Pennington, R. T., Prado, D. E., & Pendry, C. A. (2000). Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography, 27*, 261-273.

Pennington, R. T., Lewis, G. P., & Ratter, J. A. (2006) An overview of the plant diversity, biogeography and conservation of Neotropical savannas and seasonally dry forests. In R. T. Pennington, G. P. Lewis & J. A. Ratter (Eds.), *Neotropical savannas and seasonally dry forests: plant diversity, biogeography, and conservation,* (pp. 1-30). Boca Raton, FL: Taylor & Francis Group.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *Plos One, 7,* e37135.

Pielou, E. C. (1991). *After the Ice Age: the return of life to glaciated North America* (E. C. Pielou Ed.). Chicago, IL: University of Chicago Press.

Pielou, E. C. (1992). *Biogeography* (E. C. Pielou Ed.). Malabar, FL: Krieger Publishing Company.

Prado, C. P. A., Haddad, C. F. B., & Zamudio, K. R. (2012). Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. *Molecular Ecology, 21*, 921-941.

Prance, G. T. (1982). A review of the phytogeographic evidences for Pleistocene climate changes in the Neotropics. *Annals of the Missouri Botanical Garden, 69*, 594-624.

Prance, G. T. (1996). Islands in Amazonia. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 351*, 823-833.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*, 945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics, 81*, 559-575.

Quijada-Mascareñas, J. A., Ferguson, J. E., Pook, C. E., Salomão, M. D. G., Thorpe, R. S., & Wüster, W. (2007). Phylogeographic patterns of trans-Amazonian vicariants and Amazonian biogeography: the Neotropical rattlesnake (*Crotalus durissus* complex) as an example. *Journal of Biogeography, 34*, 1296-1312.

Author Manuscript

This article is protected by copyright. All rights reserved

R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ratter, J. A., Bridgewater, S., & Ribeiro, J. F. (2003). Analysis of the floristic composition of the Brazilian Cerrado vegetation III: comparison of the woody vegetation of 376 areas. *Edinburgh Journal of Botany, 60*, 57-109.

Resende-Moreira, L. C., de Vasconcelos, P. N., Andrea, P. S. A., Menezes, A. P. A., Lemos-Filho, J. P., & Lovato, M. B. (2017). East-west divergence in central Brazilian Cerrado revealed by cpDNA sequences of a bird-dispersed tree species. *Biochemical Systematics and Ecology, 70*, 247-253.

Ribeiro, P. C., Lemos, J. P., Buzatti, R. S. D., Lovato, M. B., & Heuertz, M. (2016). Species-specific phylogeographical patterns and Pleistocene east-west divergence in *Annona* (Annonaceae) in the Brazilian Cerrado. *Botanical Journal of the Linnean Society, 181*, 21-36.

Rocha, A. E. S., & Miranda, I. S. (2014). Poaceae communities in the savannas of the Amazon estuary in Brazil. *Acta Botanica Brasilica, 28*, 559-568.

Sanaiotti, T. M. (1997). Comparação fitossociológica de quatro savanas de Roraima. In R. I. Barbosa, E. J. G. Ferreira, & E. G. Castellon (Eds.), *Ambiente e Ecologia em Roraima* (pp. 481-488). Manaus, Brazil: INPA.

Santos, M. G., Nogueira, C., Giugliano, L. G., & Colli, G. R. (2014). Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography, 41*, 1506-1519.

Savit, A. Z., & Bates, J. M. (2015). Right around the Amazon: the origin of the circum-Amazonian distribution in *Tangara cayana*. *Folia Zoologica, 64*, 273-283.

Silva, J. M. C. (1995). Biogeographical analysis of the South American avifauna. *Steenstrupia, 21*, 49-67.

Silva, J. M. C., & Bates, J. M. (2002). Biogeographic patterns and conservation in the South American Cerrado: A tropical savanna hotspot. *Bioscience, 52*, 225-233.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics, 139*, 457-462.

Smith, S. A., & Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science, 322*, 86-89.

Souza, H. A., Collevatti, R. G., Lima-Ribeiro, M. S., Lemos-Filho, J. P., & Lovato, M. B. (2017). A large historical refugium explains spatial patterns of genetic diversity in a Neotropical savanna tree species. *Annals of Botany, 119*, 239-252.

Author Manuscript

This article is protected by copyright. All rights reserved

Strassburg, B. B. N., Brooks, T., Feltran-Barbieri, R., Iribarrem, A., Crouzeilles, R., Loyola, R., . . . Balmford, A. (2017). Moment of truth for the Cerrado hotspot. *Nature Ecology & Evolution, 1*, 0099.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution, 28*, 2731-2739.

Thomaz, A. T., Malabarba, L. R., & Knowles, L. L. (2017). Genomic signatures of paleodrainages in a freshwater fish along the southeastern coast of Brazil: genetic structure reflects past riverine properties. *Heredity, 119*, 287-294.

van der Hammen, T. (1991). Paleoecological background: Neotropics. *Climatic Change, 19*, 37-47.

Vargas-Ramírez, M., Maran, J., & Fritz, U. (2010). Red- and yellow-footed tortoises, *Chelonoidis carbonaria* and *C. denticulata* (Reptilia: Testudines: Testudinidae), in South American savannahs and forests: do their phylogeographies reflect distinct habitats? *Organisms Diversity & Evolution, 10*, 161-172.

Vieira, D. L. M., Aquino, F. G., Brito, M. A., Fernandes-Bulhão, C., & Henriques, R. P. B. (2002) Síndromes de dispersão de espécies arbustivo-arbóreas em cerrado *sensu strictu* do Brasil Central e savannas amazônicas. *Brazilian Journal of Botany, 25,* 215-220.

Vinson, S. B.; Williams, H. J.;, Frankie, G. W; & Shrum, G. (1997) Floral lipid chemistry of *Byrsonima crassifolia* (Malpigheaceae) and a use of floral lipids by *Centris* bees (Hymenoptera: Apidae). *Biotropica*, *1*, 76-83.

Wang, C. L., Zöllner, S., & Rosenberg, N. A. (2012). A quantitative comparison of the similarity between genes and geography in worldwide human populations. *Plos Genetics, 8*, e1002886.

Wang, X. F., Edwards, R. L., Auler, A. S., Cheng, H., Kong, X. G., Wang, Y. J., . . . Chiang, H. W. (2017). Hydroclimate changes across the Amazon lowlands over the past 45,000 years. *Nature, 541*, 204-207.

Webb, S. D. (1991). Ecogeography and the great American interchange. *Paleobiology, 17*, 266-280.

Werneck, F. P., Nogueira, C., Colli, G. R., Sites, J. W., & Costa, G. C. (2012). Climatic stability in the Brazilian Cerrado: implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. *Journal of Biogeography, 39*, 1695-1706.

Author Manuscript

This article is protected by copyright. All rights reserved

Willis, C. G., Franzone, B. F., Xi, Zhenxiang, & Davis, C. C. (2014) The establishment of Central American migratory corridors and the biogengraphic origins of seasonally dry tropical forests in Mexico. *Frontier in Genetics, 5,* 433.

Wüster, W., Ferguson, J. E., Quijada-Mascareñas, J. A., Pook, C. E., Salomão, M. D., & Thorpe, R. S. (2005). Tracing an invasion: landbridges, refugia, and the phylogeography of the Neotropical rattlesnake (Serpentes: Viperidae: *Crotalus durissus*). *Molecular Ecology, 14*, 1095-1108.

Yang, Y., Moore, M. J., Brockington, S. F., Soltis, D. E., Wong, G. K. S., Carpenter, E. J., . . . Smith, S. A. (2015). Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution, 32,* 2001-2014.

Author Manuscript

This article is protected by copyright. All rights reserved

Biosketch

The authors are broadly interested in biogeographical and evolutionary history of the Neotropics. Together the authors bring different expertise, forming a complementary team that spans from sampling and data collection to hypothesis generation and testing about the Cerrado evolutionary history. This work was part of a collaboration between UFMG and the University of Michigan.

Author contributions: LCRM, MBL, JPLF and LLK designed the study; LCRM, MBL and JPLF collected samples; LCRM, ATT, JRP and APS generated data and performed data analysis; LCRM, LLK and MBL wrote the paper and all authors contributed with comments.

Data Accessibility

Chloroplast DNA sequences will be submitted to GenBank and Illumina raw data will be submitted to NCBI Sequence Read Archive (SRP158434).

Supporting Information

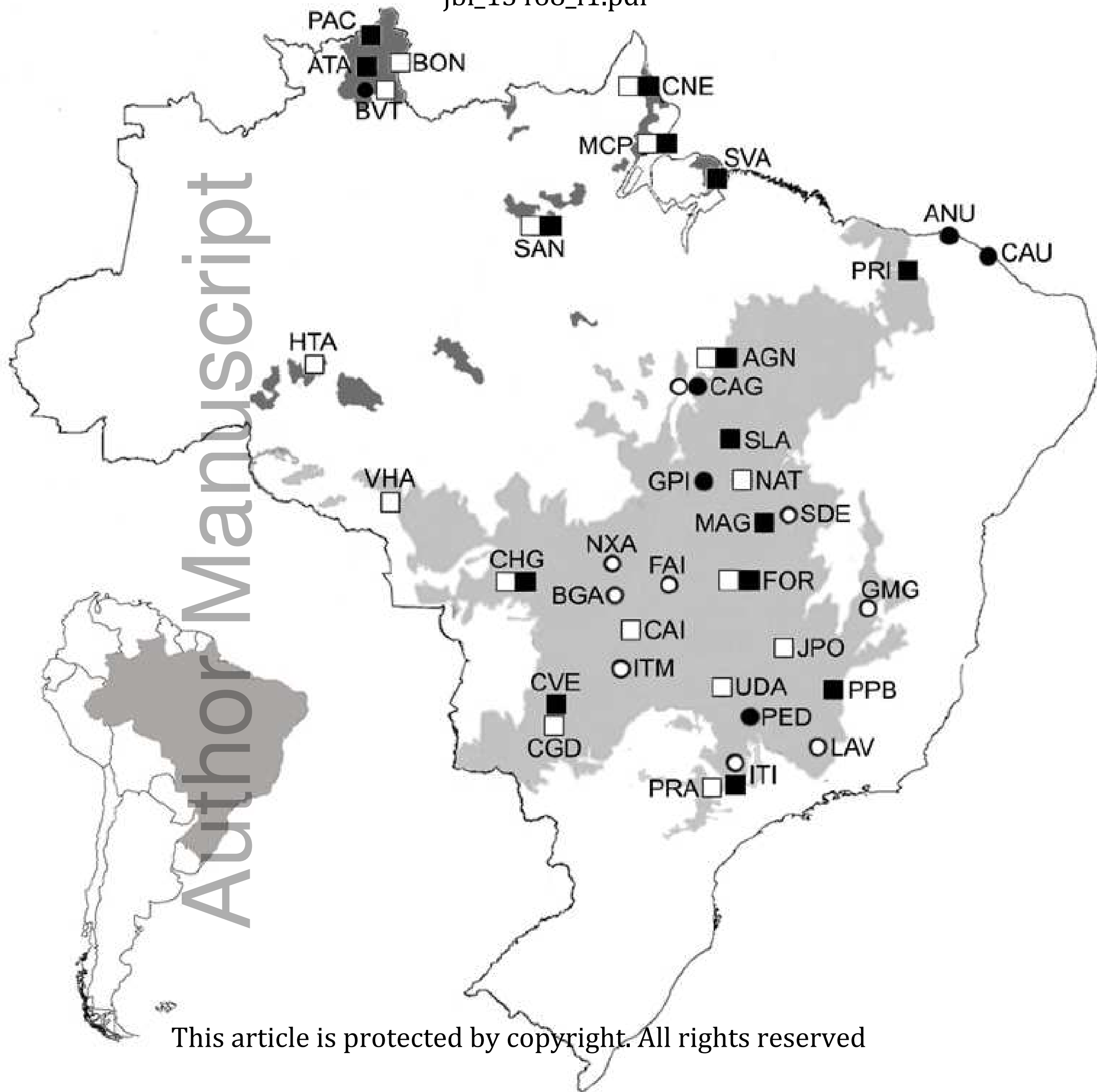
Appendix S1 - Locality information and collected genomic data

Appendix S2 - Analysis of chloroplast DNA and related results

Appendix S3 - Additional details about methods and analyses of genomic data

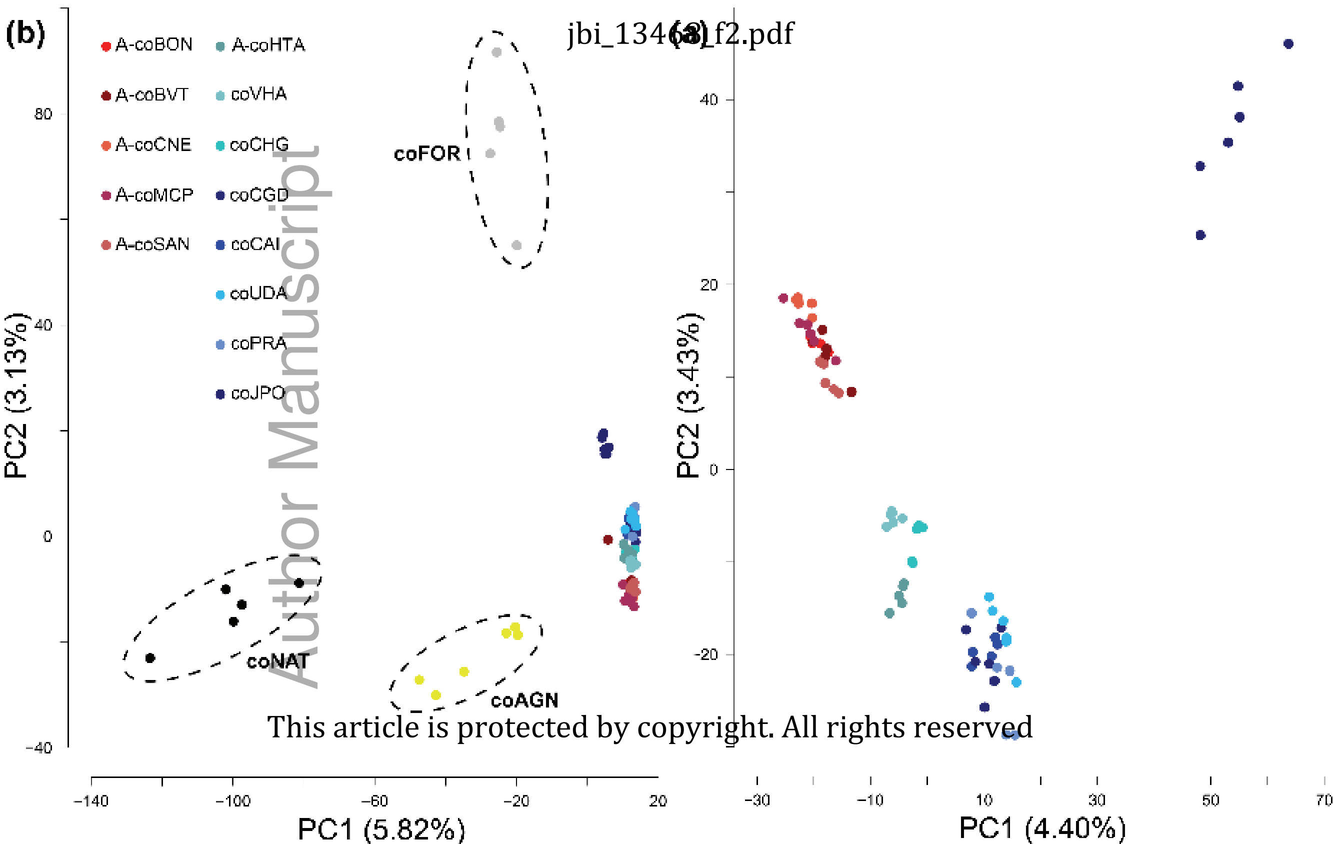
Author Manuscript

This article is protected by copyright. All rights reserved



Author Manuscript

This article is protected by copyright. All rights reserved



jbi_13463_f2.pdf

Author Manuscript

This article is protected by copyright. All rights reserved

jbi_13468_f3.pdf

(a)

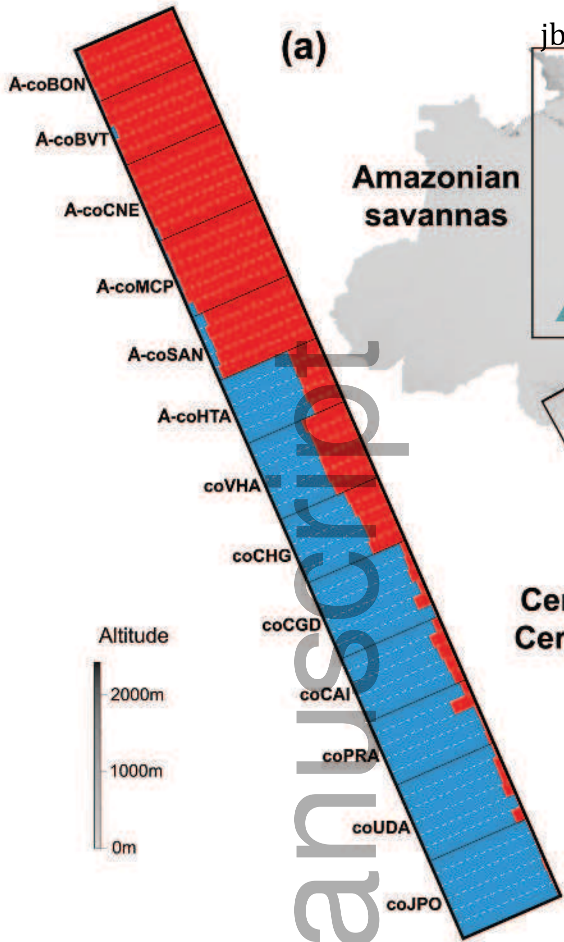
(b)

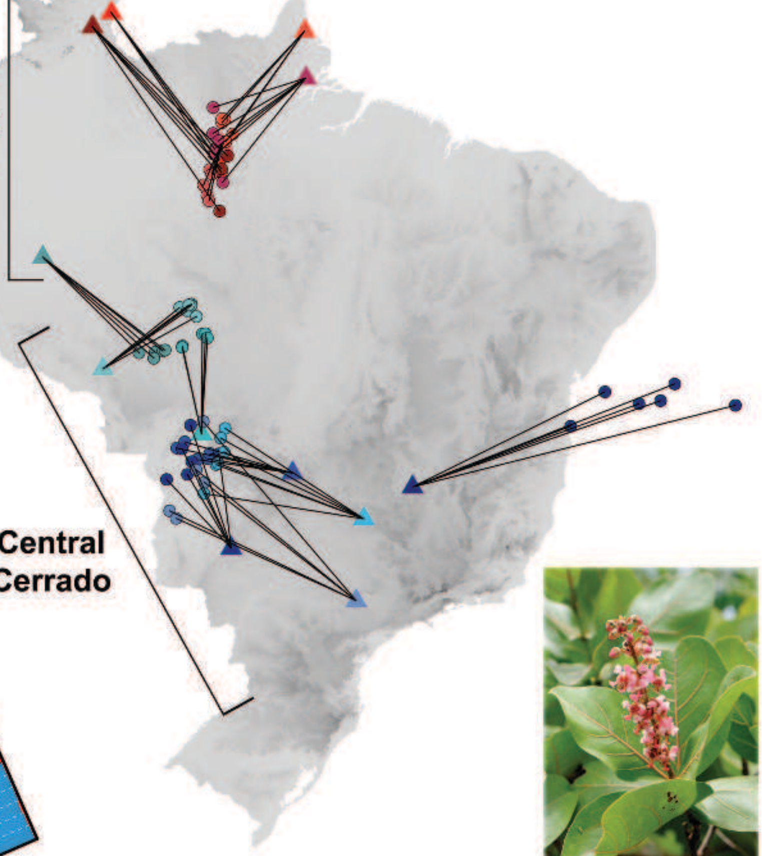
N

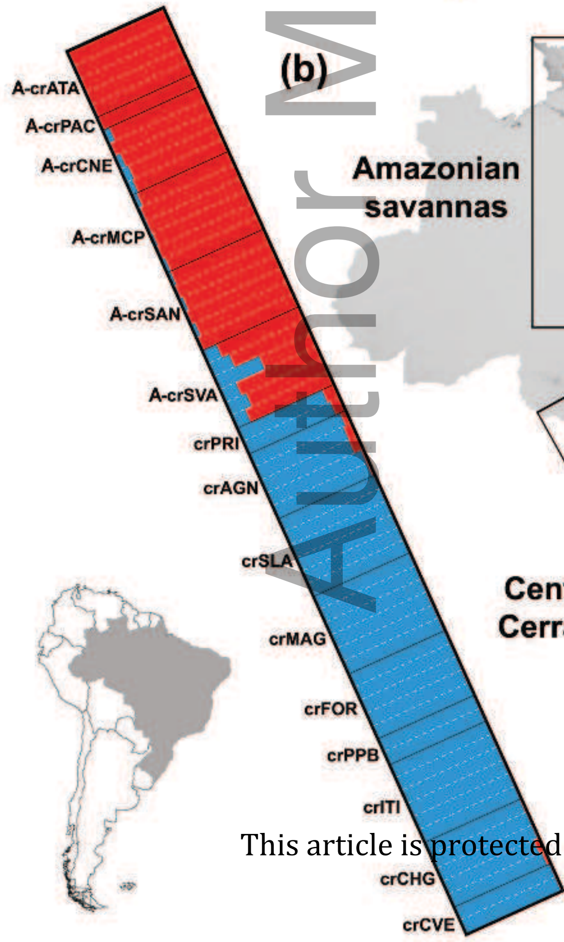
0 600 Km

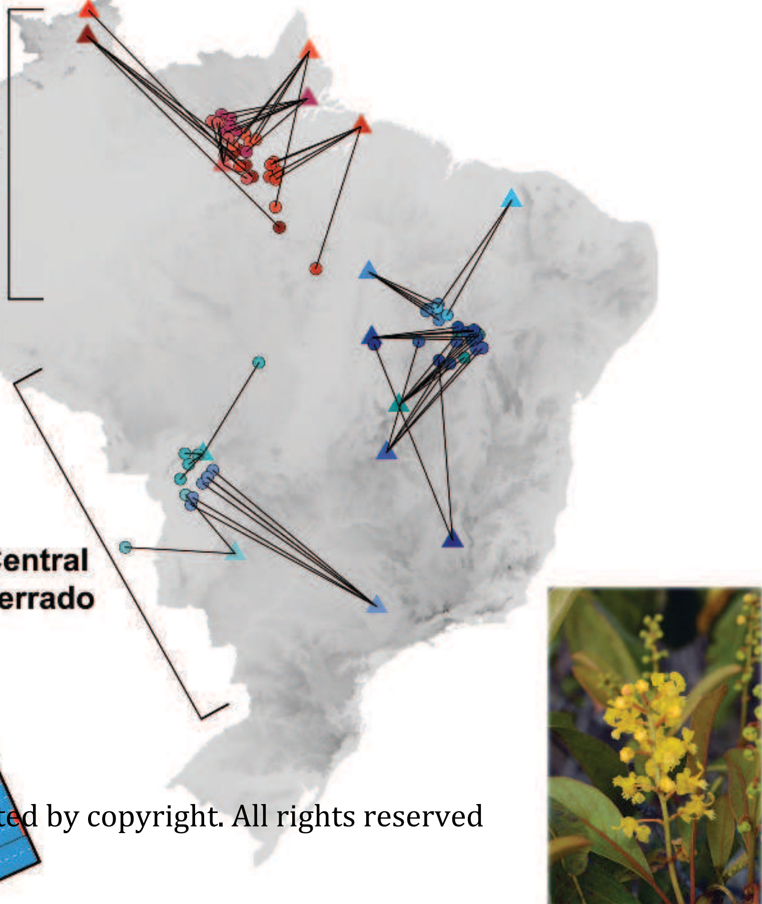
Author Manuscript

This article is protected by copyright. All rights reserved





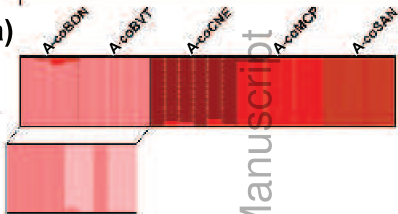


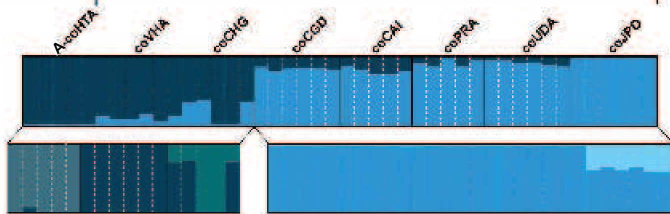


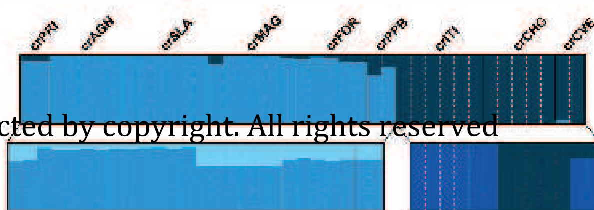
Author Manuscript

This article is protected by copyright. All rights reserved

jbi_13468_f5.pdf



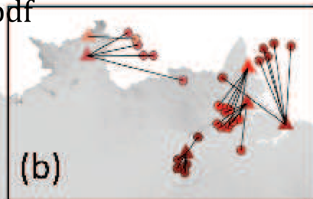


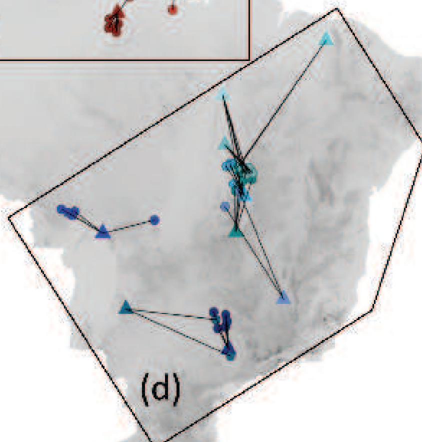


Author Manuscript

This article is protected by copyright. All rights reserved

jbi_13468_f6.pdf





Author Manuscript

This article is protected by copyright. All rights reserved