Botanical Journal of the Linnean Society, 2019, **190**, 67–82. With 3 figures.

Elucidating the *Clusia criuva* species 'complex': cryptic taxa can exhibit great genetic and geographical variation

MARIA BEATRIZ DE S. CORTEZ[1,†], DANILO A. SFORÇA[2], FÁBIO DE M. ALVES[1], JOÃO DE D. VIDAL[3,‡], ALESSANDRO ALVES-PEREIRA[1,2], GUSTAVO M. MORI[2,4,5,§], ISABELA A. ANDREOTTI[1,‖], JOSÉ E. DO NASCIMENTO JR[7], VOLKER BITTRICH[6], MARIA I. ZUCCHI[4], MARIA DO CARMO E. AMARAL[1] and ANETE P. DE SOUZA[1,2,*]

[1]*Departamento de Biologia Vegetal, Instituto de Biologia (IB), Universidade Estadual de Campinas, Campinas, São Paulo, Brazil*
[2]*Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, CP 6010, Campinas, São Paulo, Brazil*
[3]*Instituto de Biociências de Botucatu, Departamento de Botânica, Botucatu, Universidade Estadual Paulista 'Júlio de Mesquita Filho' (Unesp), São Paulo, Brazil*
[4]*Agência Paulista de Tecnologia dos Agronegócios, Pólo Centro Sul, Piracicaba, São Paulo, Brazil*
[5]*Instituto de Biociências, Campus do Litoral Paulista, Universidade Estadual Paulista 'Júlio de Mesquita Filho' (Unesp), São Vicente, São Paulo, Brazil*
[6]*Rua Dr. Mario de Nucci 500, 13083–290, Campinas, São Paulo, Brazil*
[7]*curso de Ciências Biológicas, Universidade Federal do Triângulo Mineiro, Iturarama, Minas Gerais, Brazil*
[†]*Current Address: Department of Biology, University of Florida, Gainesville, FL, USA*
[‡]*Current Address: Afromontane Research Unit & Department of Geography, University of the Free State, QwaQwa campus, Private Bag X13, Phuthaditjhaba, 9866, Republic of South Africa*
[§]*Current Address: Instituto de Biociências, Campus do Litoral Paulista, Universidade Estadual Paulista 'Júlio de Mesquita Filho' (Unesp), São Vicente, São Paulo, Brazil*
[‖]*Current Address: Department of Nature Sciences, Mathematics and Education, Federal University of São Carlos, 13600-970, Araras, São Paulo, Brazil*

Received 11 April 2018; revised 11 January 2019; accepted for publication 30 January 2019

In the *Clusia criuva* Cambess. species complex, the two subspecies *C. criuva* subsp. *parviflora* Vesque and *C. criuva* subsp. *criuva* can only be distinguished on the basis of stamen/staminode morphology and geographical occurrence. Despite being recently restructured, taxonomic relationships in this complex remain unclear. Therefore, to illuminate the evolutionary mechanisms involved in the diversification of these two lineages we investigated their population structure, phylogeographical and niche distribution patterns using plastid and nuclear microsatellites (plastid SSRs and nuSSRs, respectively). We obtained ten polymorphic nuSSRs from a microsatellite-enriched library and used six previously described plastid SSRs to genotype *c.* 300 samples. We conducted F-statistics, genetic distance and population structure analyses to test whether the subspecies presented distinct genotypic clusters. Putative phylogeographic breaks were also identified and tested. Finally, we developed distribution models to contrast genetic and environmental information. We found extensive genetic differentiation between the subspecies. Three significant breaks were identified, two of which coincide with geographical barriers. Niche modelling predictions indicated that *C. criuva* subsp. *criuva* potentially occupied a much wider area during the Last Glacial Maximum than it does today. These results indicate that both lineages are evolving independently because of limited gene flow and restriction to different environments, suggesting that they should again be elevated to species status. To clarify this issue, we recommend further phylogenetic and morphological studies.

KEYWORDS: microsatellite markers – niche modelling – phylogeography – population genetics – speciation.

*Corresponding author. E-mail: anete@unicamp.br

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

INTRODUCTION

Species concepts and delimitation have always been highly controversial and complicated, especially when the focal organisms are considered cryptic or hyper-cryptic. Recently, De Queiroz (2007) reviewed six major classes of species concepts with contrasting properties, identified an 'underlying conceptual unity' and proposed a unified concept of species. According to this concept, species are metapopulation lineages that are evolving separately and for which separation and divergence (speciation process) provide evidence for species delimitation (De Queiroz, 2007). Thus, different phases along the isolation of populations relate to different effects on ecological, genotypic or morphological differentiation. Identifying these effects can be informative concerning the actual relationship among lineages, especially if one line of evidence alone does not offer conclusive support in determining kinship among closely related organisms. To assist in the delimitation of a species complex, a recent study (Pessoa *et al.*, 2012) used four distinct criteria following the strong recommendation for the use of different frameworks for species identification (Padial *et al.*, 2010). Nonetheless, other studies have succeeded in defining new species based on fewer criteria (Edwards *et al.*, 2009; Caddah *et al.*, 2013), demonstrating that it is not imperative to consider many criteria if a clear and satisfactory separation can be achieved through the use of fewer parameters.

The application of genetic markers to characterize the relationships between lineages is useful in detecting signals of isolation, especially in cryptic species (Adams *et al.*, 2014). Complementarily, Van Valen (1976) argued that the niche or adaptive zone is relevant for establishing boundaries between morphologically similar species. This consideration is particularly important for widespread species occurring in more than one environment.

There are 261 genera in the Brazilian cerrado, among which 205 species also inhabit the Atlantic Forest (Heringer *et al.*, 1977). Considering these closely related taxa that are not restricted to one domain or even to one vegetation type, delimitation based on environmental differences can detect ongoing speciation events. In this context, the *Clusia criuva* Cambess. *s.s.* 'complex' can serve as a model for analysing the influence of ecological boundaries on the process of speciation, given its occurrence in various physiognomies presenting distinct ecological characteristics. This species recently underwent a taxonomic reorganization (Bittrich, 2003), causing it to be divided into two currently recognized subspecies (*C. criuva* subsp. *criuva* and *C. criuva* subsp. *parviflora* Vesque) based on a subtle but unambiguous difference in the length of the connective prolongation of the anthers and antherodes. In *C. criuva* subsp. *criuva*, the anthers are three to five times larger than the apical connective extension and in *C. criuva* subsp. *parviflora* the anthers are two to six times smaller than the apical connective extension, which in this case is acute and easily observable. Furthermore, the geographical distributions of the two taxa do not overlap, with the former being distributed in the Brazilian cerrado, in the states of Minas Gerais, Bahia and Goiás and the latter occuring in the eastern Atlantic Forest between the northern part of Rio Grande do Sul and the state of Rio de Janeiro. Another important consideration is that both subspecies occur in different physiognomies: *C. criuva* subsp. *criuva* inhabits both rocky fields and gallery forests, whereas *C. criuva* subsp. *parviflora* occurs across different elevational ranges along the coastal region of the Atlantic Forest, being subject to higher levels of precipitation and inhabiting locations where the soil is predominantly sandy and deep. Although the domains inhabited by *C. criuva* represent distinct habitats, they share one relevant and concerning aspect: both regions have been subjected to intense deforestation through anthropogenic action. By 2008, the cerrado had lost *c.* 50% of its original vegetation (MMA/IBAMA 2011), and the Atlantic Forest had lost *c.* 76% by 2009 (MMA/IBAMA, 2012).

Due to its confusing taxonomic history, it is possible to consider *C. criuva* as a potentially cryptic species, as, according to Adams *et al.* (2014:13), 'cryptic biodiversity is theoretically possible for any species that has been defined solely on morphological criteria'. Therefore, we employed methods based on the species delimitation criteria described above to generate novel data for the taxonomic circumscription of this complex, allowing for possible modifications to be endorsed, if necessary. Population genetic studies were applied to understand how these subspecies are genetically structured and differentiated, and phylogeographic and ecological niche modelling analyses were employed to understand the relationships of the taxa with geography and time. Vaasen, Scarano & Hampp (2007) previously assessed the genetic structure of *C. criuva* subsp. *parviflora* in the state of Rio de Janeiro, when the taxon was still considered a distinct species ('*C. parviflora* Engl.' nom. illeg.) to evaluate whether distinct environments could have accounted for potential genetic differences using a limited set of nuclear microsatellites (nuSSRs). We followed a similar approach but focused on comparing the two subspecies of *C. criuva*, which had not previously been attempted. To this end, we developed a new set of nuSSRs to obtain a higher number of polymorphic markers that could provide a broader sampling of the genome and more robust results to solve taxonomic problems at the species or infraspecific level. Additionally, we included

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

plastid microsatellite (plastid SSR) information in our study since the combination of these markers enables a better understanding of the relationship between taxa, especially in a phylogeographical framework (Turchetto-Zolet *et al.*, 2012). Furthermore, analysis of plastid SSR is useful to understanding the connection between genetic processes and the geographical distributions of distinct lineages at the intraspecific level, providing insights into seed dispersal because such markers are, for most angiosperms, maternally inherited (Powell *et al.*, 1995). We also employed ecological niche modelling to provide a thorough understanding of how ecological factors influence speciation (Hickerson *et al.*, 2010). This information is particularly important for comparisons of two similar taxa that occur in contrasting environments that could exert considerable influences on their distributions, as is the case for *C. criuva*. When comparing genetic variation in combination with models of niche distribution, it is possible to find support for demographic changes and/or stability revealed by molecular markers (Alvarado-Serrano & Knowles, 2013). Therefore, we combined distribution models, niche identity tests and nuclear and plastid molecular markers to elucidate the evolutionary relationships of subspecies of *C. criuva* and to possibly gather insight into the historical relationship between the different environments occupied by this species complex.

MATERIAL AND METHODS

Species description

Clusia criuva is a dioecious tree or treelet, occasionally shrub, with white exudate. The leaf blade is coriaceous and distinctly discolorous *in vivo*, with a whitish-green coloration on the abaxial side. Both male and female flowers have four to ten greenish-red sepals and five free white petals. In the staminate flowers the pistillode is absent and there are numerous stamens, whereas in the pistillate flowers there are usually five to ten staminodes and a five-locular ovary with a sessile stigma. The fruit is a fleshy septifragal capsule with a subapical stigma and persistent sepals and staminodes. Pollination is carried out by beetles that feed on pollen in the staminate flowers and afterwards visit the pistillate flowers by mistake (Correia *et al.*, 1993). Seeds are dispersed by birds that feed from the aril (V. Bittrich, pers. com.; J. Nascimento-Jr, pers. com.).

Sample collection and preparation

We sampled at ten locations in Brazil (Table 1, Figs 1, 2), distributed between latitudes 12° and 29°S and longitudes 41° and 49°W, thus including almost the entire species distribution. Sampling localities comprised areas in the Brazilian cerrado and Atlantic Forest, and GPS coordinates (datum WG84) were recorded for each location. Distances between pairs of populations ranged from 89 km to 1599 km. From each *C. criuva* sampling site, 30 individuals were collected, except from Parque Estadual do Cerrado (PRC) and Riviera de São Lourenço (SPR), from which 21 and 28 individuals were collected, respectively (Table 1). To prevent the collection of parents and their offspring, we attempted to sample individuals that were considerably separated from one another (usually at least 5 m apart). A hand lens was used to inspect the male flowers in each population to determine identification of the subspecies based on stamen differentiation. The leaves were lyophilized and stored in –20 °C freezers prior to DNA extraction using a CTAB protocol (Doyle & Doyle, 1987). See the Supplementary Material (Table S8) for a list of the vouchers.

Development of microsatellite library

We developed a microsatellite-enriched library following the protocol proposed by Billotte *et al.* (1999). The DNA for the library was obtained from an individual of *C. criuva* subsp. *parviflora* located at the Central Experimental Center 'Fazenda Santa Elisa' of the Agronomical Institute of Campinas, Campinas, São Paulo, Brazil (22°52'15" S, 47°4'31" W – accession number: UEC122246). We obtained 40 pairs of primers complementary to the flanking sequences of dinucleotide microsatellites with more than six motif repetitions through the software Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) following the standard parameters: 18–22 base pairs, T_m between 45 and 65 °C with a salt concentration of 50 mM, to ensure comprehensive and specific amplification. Additionally, differences of 3 °C in T_m between paired oligonucleotides, a GC content of 40–60%, absence of complementarity between paired oligonucleotides and polymerase chain reaction (PCR) products between 100 and 360 base pairs were enforced to ensure high resolution during genotyping. Characterization of polymorphisms was performed with two individuals of each sampled population, resulting in a successful amplification of 28 out of the 40 pairs of primers developed. Subsequently, the products were inspected in a 3%(w/v) agarose gel, stained with ethidium bromide and genotyped in a 6%(w/v) silver-stained polyacrylamide gel, following Creste, Neto & Figueira (2001). Sixteen loci were found to be polymorphic. We deposited the annotated sequences in GenBank (CCSSR01-CCSSR10).

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

Table 1. Sampling locations, their codes used in figures and tables, GPS coordinates, domains and corresponding vegetation type where each subspecies of *Clusia criuva* occurs

Sampling location	Code	Latitude	Longitude	Biome	Vegetation	Subspecies
Floresta Estadual do Palmito	PR	25° 37' 8" S	48° 38' 16" W	Atlantic Forest	Lowland dense ombrophilous forest (restinga)	*C. criuva* subsp. *parviflora*
Riviera de São Lourenço	SPR	23° 48' 43" S	46° 2' 18" W	Atlantic Forest	Lowland dense ombrophilous forest (restinga)	*C. criuva* subsp. *parviflora*
P. E. S. M. Santa Virgínia	SPV	23° 20' 6" S	45° 8' 43" W	Atlantic Forest	Montane dense ombrophilous forest	*C. criuva* subsp. *parviflora*
Parque Nacional de Itatiaia	RJ	22° 41' 24" S	44° 45' 23" W	Atlantic Forest	Montane dense ombrophilous forest	*C. criuva* subsp. *parviflora*
Parque Estadual de Itapeva	RS	24° 23' 7" S	49° 42' 36" W	Atlantic Forest	Lowland dense ombrophilous forest (restinga)	*C. criuva* subsp. *parviflora*
Porto Belo	SC	27° 21' 33" S	48° 38' 9.9" W	Atlantic Forest	Submontane dense ombrophilous forest	*C. criuva* subsp. *parviflora*
P. N. Chapada Diamantina	BA	12° 40' 35" S	41° 35' 18 " W	Cerrado	Rocky fields	C. *criuva* subsp. *criuva*
RECOR	DF	16° 0' 12" S	47° 55' 43" W	Cerrado	Gallery forest	C. *criuva* subsp. *criuva*
Parque Estadual do Ibitipoca	MG	21° 54' 21" S	44° 7' 3" W	Atlantic Forest	Rocky fields	C. *criuva* subsp. *criuva*
Jaguariaíva	PRC	24° 23' 7 " S	49° 42' 36" W	Atlantic Forest	Disturbed fragment	*C. criuva* subsp. *parviflora*

Nuclear and plastid microsatellite amplification and genotyping

Of the 16 polymorphic loci, six showed unclear peak patterns and were not included in analyses (Supplementary Material, Table S1). The forward primers used in the PCR were fluorescently labelled with 6-FAM (Sigma-Aldrich, Saint Louis, MO), NED, PET or VIC (Applied Biosystems, Foster City, CA). The following PCR protocol (15 µL) was used: 1× buffer, 1.25 mM $MgCl_2$, 0.125 mM each dNTP, 0.3 µM fluorescence-labelled forward primer, 0.6 µM reverse primer, bovine serum albumin (BSA) 5 µg $µL^{-1}$, 1 U Taq DNA polymerase and 10 ng template DNA. PCR conditions were as follows: initial denaturing at 94 °C for 5 min followed by 30 cycles of 94 °C (1 min), T_m (56 °C, 60 °C and 62 °C, depending on the pair of primers), 72 °C (2 min) and a final elongation at 72 °C for 5 min. We inspected the quality of the amplification in 3% (w/v) agarose gels stained with ethidium bromide before genotyping on an ABI 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA). The peaks were called using the software Geneious v.8.1.7 (http://www.geneious.com/, last accessed December 2017), and the size standard GeneScan 600 LIZ (Applied Biosystems, Foster City, CA) was used to calibrate amplicon sizes.

We tested 20 plastid SSRs and obtained six polymorphic markers for *C. criuva*: ccmp 1, ccmp 2, ccmp 10 (Weising & Gardner, 1999), cc 4, cc 7 and cc 9 (Chung & Staub, 2003). We used IRDye-700- or IRDye-800-labeled primers for posterior genotyping. The following PCR protocol (10 µL) was used for amplifications: 1× buffer, 2 mM MgCl, 2 mM each dNTP, 0.3 µM forward primer, 0.8 µM reverse primer, BSA 5 µg $µL^{-1}$, 1 U Taq DNA polymerase and 10 ng template DNA. PCR conditions were as follows: initial denaturing at 94 °C for 4 min; 30 cycles of 94 °C (1 min), 58 °C (1 min) and 72 °C (1 min); followed by ten cycles of 94 °C (40 s), 53 °C (40 s) and 72 °C (40 s) and final elongation at 72 °C for 10 min. We inspected the quality of the amplifications using 3% (w/v) agarose gels stained with ethidium bromide. Genotyping was performed using a 4300 DNA Analyzer (LI-COR), 50–350 bp size standards (IRDye 700 IRDye 800) and the software GIMP (http://www.gimp.org/, last accessed February 2018).

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

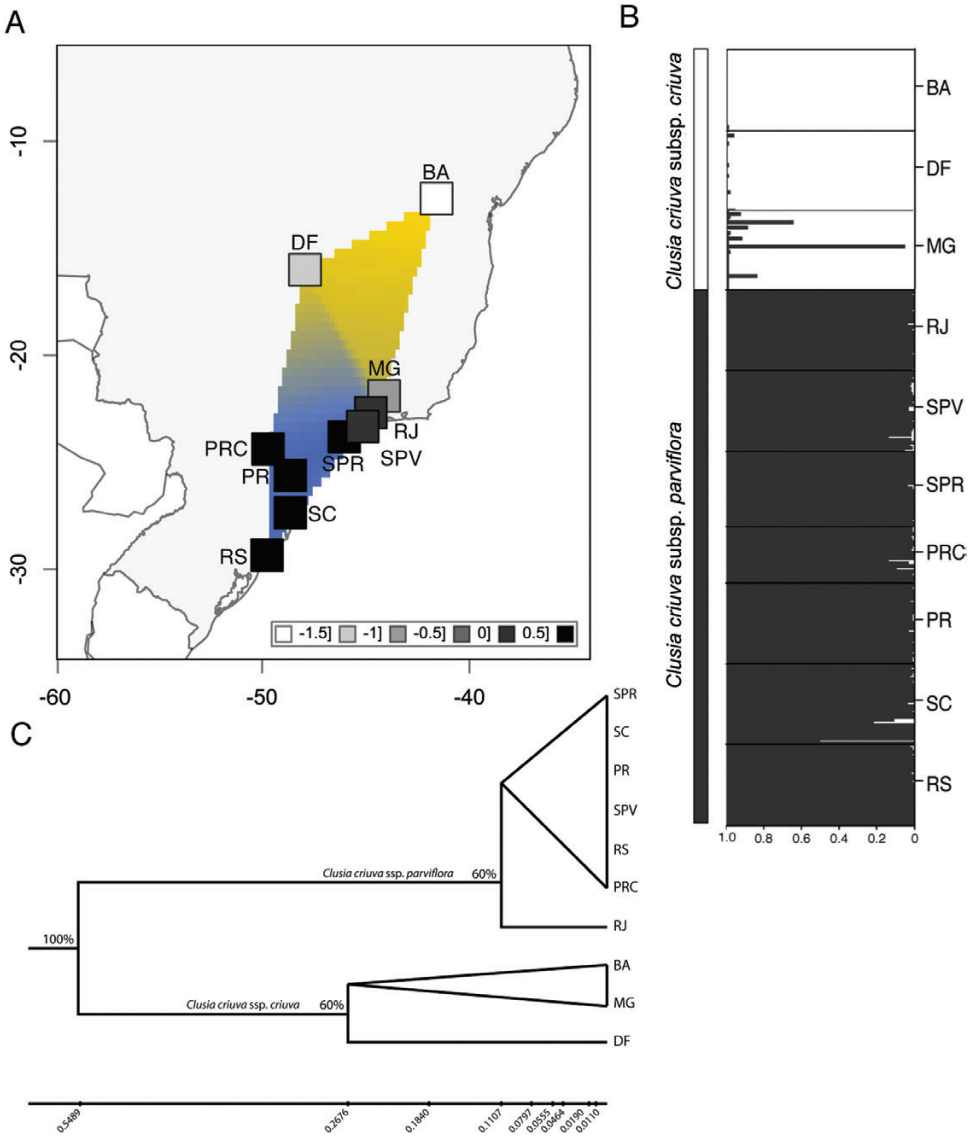

Figure 1. A, Interpolation of the first Eigenvalue using lagged scores obtained through sPCA in the statistical software R. The map plots show loading scores for each sampling site, represented as squares, with colors representing their values: dark squares designate positive values, light squares indicate negative values. B, Structure bar plot of the ten sampling locations of C. criuva, considering $K = 2$, based on the variation of nuSSRs. C, Dendrogram based on Nei's (1972) genetic distances using nuclear microsatellites computed using the UPGMA clustering method with 10 000 iterations. Branching support values (> 60%) are shown for each node as percentages. Sampling locality codes follow Table 1.

GENETIC DIVERSITY AND POPULATION STRUCTURE

We identified the presence of private alleles, expected (H_E) and observed heterozygosity (H_O) and allele frequencies of nuSSR loci (by sampling location and locus) using the software GenAlEx v.6.5 (Peakall & Smouse, 2012). The software FSTAT v.2.9.3.2 (Goudet, 2001) was used to estimate allelic richness based on a minimal sample size of 13 individuals and genotypic disequilibrium based on 900 permutations with a significance level of 5%. For each location, we used the software GenePop v.4.0.7 (Rousset, 2008) to test for deviations from the Hardy–Weinberg equilibrium, with exact tests based on 100 Markov chain Monte Carlo (MCMC) batches, 10 000 iterations per batch and a dememorization number of 10 000. To estimate the frequency of null alleles and its influence on F_{ST}, the algorithm proposed by Dempster, Laird & Rubin (1977) was applied in the software FreeNA (Chapuis & Estoup, 2007), considering $N > 0.20$ as indicative of the presence of null alleles.

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

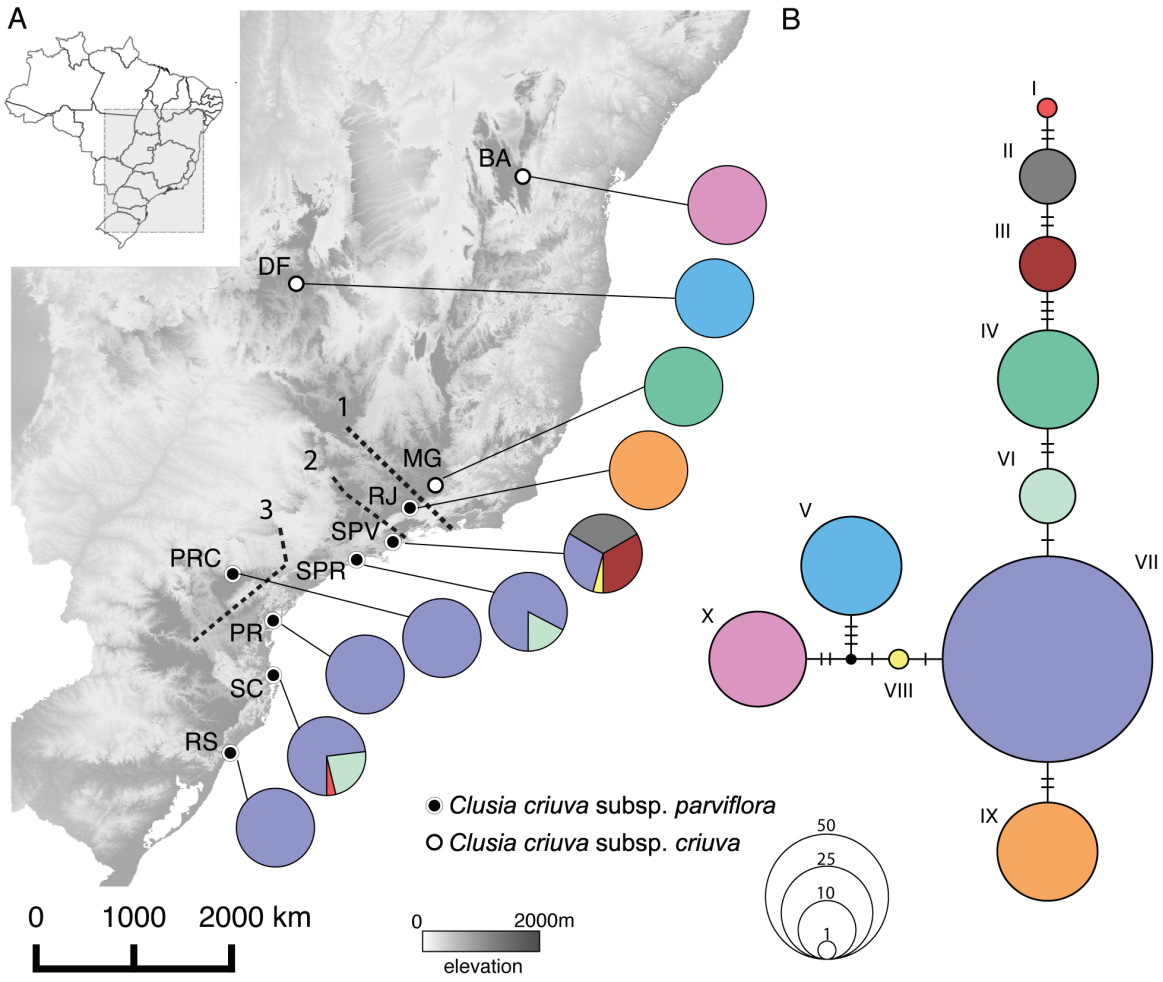

Figure 2. A, Topographical map showing locations and plastid SSR haplotype frequencies of populations of *C. criuva*. Circles are not proportional to sampling sizes, and each colour represents different plastid SSR haplotypes. Dashed lines indicate the three identified breaks for *C. criuva* based on nuSSR allele frequencies. Break 1 – between populations MG and RJ, 2 – isolating population RJ and 3 – isolating population PRC. B, Haplotype network obtained with the software NETWORK 5.001 showing the ten plastid SSR haplotypes identified by distinct colours. Circles are proportional to sampling sizes, each transverse line indicates 1-bp mutation between haplotypes, and the black dot represents an unobserved haplotype. Sampling locality codes follow Table 1.

Genetic differentiation was estimated through pairwise F_{ST} values and confidence intervals with 1000 bootstraps, using diveRsity (Keenan *et al.*, 2013) for R (R Development Core Team, 2016). This package implements the analysis of variance approach (Weir & Cockerham, 1984). A dendrogram based on Nei's genetic distance (Nei, 1972) matrices was generated with the software TFPGA v.1.3 (Miller, 1997) following the unweighted pair group method using arithmetic averages (UPGMA) clustering method with 10 000 bootstrap permutations. We tested for genetic structure within and among populations using the Bayesian inference clustering method implemented in the software Structure v.2.3.4 (Pritchard, Stephens & Donnelly, 2000). We used the 'admixture' model as we suspected that there might be gene flow between sampling locations, and allele frequencies were considered to be correlated (Falush, Stephens & Pritchard, 2003). We ran 500 000 iterations of MCMC with a burn-in of 150 000. Simulations were performed for number of groups (K) varying from 1 to 15, with 20 independent runs for each K value. Using the online software Structure Harvester (Earl & vonHoldt, 2012), we obtained ΔK (Evanno, Regnaut & Goudet, 2005), which estimates the most likely K considering the uppermost hierarchical level, among the sampling locations analyzed. Using the same model (assuming an admixture model and correlated allele frequencies), we assessed the number of putative hybrids in both subspecies through the ancestry model

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

(GENSBACK = 3 and MIGPRIOR = 0.05). We also ran 500 000 MCMC iterations with a burn-in of 150 000, assuming $K = 2$, based on the ΔK results obtained in the previous analyses.

To evaluate significant genetic separation of the *C. criuva* complex into subspecies or other levels of classification, we employed a hierarchical analysis of molecular variance (AMOVA; Excoffier, Smouse & Quattro, 1992). We tested three a priori hypotheses based on subspecies identification and on the biome and vegetation from which samples were obtained. The criterion we used to identify the grouping that best explained our nuSSR data was the maximum variance among groups (Φ_{CT} values) that significantly departed from random distributions.

Demographic changes and phylogeographic structure

To identify genetic barriers, we used the method implemented in the software Barrier v.2.2 (Manni, Guerard & Heyer, 2004), which correlates geographical and genetic distances between populations using Monmonier's maximum difference algorithm (Monmonier, 1973). We imported into this software the coordinates from all sampling locations and Nei's genetic distance (Nei, 1972) matrices based on the nuSSRs produced in GenAlEx v.6.5. The first three barriers suggested by the software were plotted on the topographical map of Brazil, enabling us to compare their exact locations to the terrain where they occurred. We used the software SPAGeDi v.1.4 (Hardy & Vekemans, 2002) to test for phylogeographic structure and calculated R_{ST} (Slatkin, 1995), an F_{ST} analogue that assumes the stepwise mutation model (SMM), considering the genetic structure we observed (see Results). When R_{ST} is significantly higher than F_{ST}, it is possible to detect phylogeographic structure. Permutation tests provide a break in the structure due to allelic sizes but still maintain allelic identity. Therefore, if R_{ST} is still higher than F_{ST} after the permutation tests, then stepwise mutations are contributing more to phylogeographic structure than are migration rates (Hardy, 2003). We performed 20 000 allele permutations to obtain the values of R_{ST} using SPAGeDi. We also used the software Bottleneck version v.1.2.02 (Cornuet & Luikart, 1997) to assess whether any of the sampled populations had experienced the effects of recent bottlenecks. These demographic retractions can be detected when the heterozygosity expected under mutation-drift equilibrium is lower than the heterozygosity observed in a population considering the size of the sample and the number of alleles. We adopted the SMM and the two-phase model (TPM) (15 variations and 95% of SMM) and later evaluated them with the Wilcoxon signed-rank test.

Using the package 'adegenet' v.2.1.0 (Jombart, 2008) in the statistical software R (R Development Core Team, 2016), we assessed how the distribution of *C. criuva* in space affects the genetic variability observed using spatial principal components analysis (sPCA; Jombart *et al.*, 2008). This analysis uses Moran's *I* (Moran 1948, 1950) algorithm to establish a comparison between the allelic frequency of two neighbouring sites. To determine the latter, a connection network is used to define which sites can be considered neighbours (Jombart *et al.*, 2008). Here, we opted for the inverse distance algorithm as a measure of connectivity, determining the exponent as 2 and the minimal distance as 0.001. As an output of this type of analysis, two possible outcomes are predicted. The first is the presence of 'global structure,' which allows for a complete or gradual differentiation between two groups, indicating a positive spatial correlation. The second is a 'local structure' that reflects stronger genetic differences among closely located individuals, indicating for this scenario a negative spatial correlation (Jombart *et al.*, 2008).

Haplotypes and molecular variance analyses

The haploid data obtained with plastid SSRs was used to calculate the haplotype frequencies and to assess whether a higher percentage of molecular variance was present within our sampling localities, using an AMOVA (Excoffier *et al.*, 1992) test. These estimations were performed with GenAlEx 6.5. After 10 000 permutations of the full data set, we also obtained the F_{ST} analogue index Φ_{PT}. We described the plastid SSR haplotype phylogenetic relationships by reconstructing a median-joining network (Bandelt, Forster & Röhl, 1999) with NETWORK 5.001 (http://www.fluxus-engineering.com/, last accessed December 2017).

Niche modelling

We used the software MAXENT (Phillips, Anderson & Schapire, 2006), a reliable method to model and project species distributions (Elith *et al.* 2011), to predict the current and past [6000 before present (BP) – Mid-Holocene and 21 000 BP – Last Glacial Maximum (LGM)] geographical ranges of both subspecies of *C. criuva*. Niche models were developed based on 22 sampling locations of *C. criuva* subsp. *criuva* and 30 of *C. parviflora* subsp. *parviflora*. Using a set of 19 bioclimatic variables available from WorldClim v.1.4 (Hijmans *et al.*, 2005), we calculated pairwise Pearson's correlations to exclude the effects of collinearity from the generated models. We generated

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

a random set of 1000 points in space and extracted the corresponding values from each environmental variable in the present. One of the variables per pair with high correlation (> 0.8) was discarded based on the percentage of contribution from each layer in previous models tested with the same data set. Our final set of non-collinear variables included mean diurnal range (bio2), isothermality (bio3), temperature seasonality (bio4), mean temperature of warmest quarter (bio10), annual precipitation (bio12), precipitation seasonality (bio14) and precipitation of driest month (bio15), of wettest quarter (bio16), of warmest quarter (bio18) and of coldest quarter (bio19). We performed two types of analyses [(1) considering only *C. criuva* subsp. *parviflora* and (2) only *C. criuva* subsp. *criuva*] to assess possible differences between them. We used the area under the receiver operating characteristic curve (AUC; Hanley & McNeil, 1982) to evaluate the level of randomness of the model generated. We present the result here as a consensus of two global climate models (Model for Interdisciplinary Research on Climate, MIROC, and Center for Climate System Research, CCSM) that were used to predict the distribution of *C. criuva* in the past. For quantification of niche equivalency between species niches, we performed pairwise comparisons between models generated for *C. criuva* subsp. *criuva* and *C. criuva* subsp. *parviflora*. Schoener's *D* (Schoener, 1968) and Warren's *I* (Warren, Glor & Turelli, 2008), two indices of niche similarity, were calculated using the ENMtools R package (Warren, Glor & Turelli, 2010). Both indices compare the per-cell suitability between two species inside a geographical area, but the latter incorporates a Hellinger distance-based correction (Warren *et al.*, 2008). We estimated the statistical significances of these indices by comparing the observed values for *I* and *D* with the values calculated for a null distribution of 99 random models generated with random subsets within the total occurrence points for the two subspecies. With this approach, we addressed the probability that the observed similarity values between models for niches of subspecies were random while calculating a *P* value for each comparison.

RESULTS

NUCLEAR MICROSATELLITES

Genetic diversity and population structure

We identified 41 alleles across ten loci in the 289 samples of *C. criuva*, with an average of 4.1 alleles per locus and 23.8 alleles per population (Supporting Information, Tables S2 and S3). Population BA showed the lowest number of alleles (20) and the lowest allelic richness (18), whereas population SPV displayed the highest number of alleles (29) and the greatest allelic richness (26). We identified a total of six private alleles distributed in four different populations: three in the state of SP (two in SPV and one in SPR), one in BA and two in DF. Population MG had the lowest observed (H_O = 0.19) and expected (H_E = 0.18) heterozygosities. The highest H_O was 0.56 (RS), whereas the highest H_e was 0.40 (SPV). The average H_O across populations was higher than the average H_E: 0.46 and 0.33, respectively (Supporting Information, Table S2). For nine out of ten populations, the null hypothesis of Hardy–Weinberg equilibrium was rejected ($P < 0.05$). For the fixation index (F_{IS}), all sampling locations showed negative values, indicating an excess of heterozygotes in these populations. The estimates ranged from −0.73 (SPR) to −0.15 (MG) and resulted in an average value across loci and populations of −0.37. We observed that only one locus had a high null allele frequency ($N > 0.20$) for one population (RJ) (Supporting Information, Table S4). The remaining values fell into the categories of moderate (12%, $0.05 \leq N < 0.19$) or low frequencies (87%, $N < 0.05$). Locus C18 showed a majority of moderate values (seven out of ten) and was the only locus to display this pattern. The value obtained for F_{ST} using the software FreeNA was 0.329 without the ENA correction described in Chapuis & Estoup (2007). With the correction, the value obtained was 0.321.

The shortest distance between two populations was between RJ and SPV (89 km), composed only of *C. criuva* subsp. *parviflora*, with an F_{ST} of 0.1, which indicates an intermediate level of genetic differentiation (Balloux & Lugon-Moulin, 2002). The greatest distance was between BA (*C. criuva* subsp. *criuva*) and RS (*C. criuva* subsp. *parviflora*) (1599 km), and the F_{ST} observed was 0.525, showing an extremely high level of genetic structure between these two populations (Balloux & Lugon-Moulin, 2002) (Table 2; see Table S5 in Supporting Information for confidence intervals). It was important to contrast sampling locations that are geographically close but that belong to different subspecies, as is the case for RJ and MG. These constituted the second closest pair of populations (95 km) and exhibited an extremely high estimate of F_{ST} (0.448). Only two pairwise F_{ST} estimates were not significant (PR/SPR – 0.014 and SPR/SC – 0.007)

The UPGMA dendrogram based on Nei's genetic distances (1972) (Fig. 1) grouped the populations into two well-supported distinct clusters. The first cluster (*criuva*) encompasses the *C. criuva* subsp. *criuva* populations (BA, DF and MG), and the second cluster (*parviflora*) consists of the seven *C. criuva* subsp. *parviflora* sampling locations (PR, SPR, SPV, RJ, RS, SC and PRC). The *criuva* cluster presented high genetic distances and geographical distances among populations, whereas the *parviflora* cluster

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

Table 2. Pairwise F_{ST} estimates (lower diagonal) and corresponding geographical distances in kilometres (upper diagonal) among populations of *Clusia criuva*. Green – low genetic structure (F_{ST} = 0.00 – 0.05), yellow – intermediate genetic structure (F_{ST} = 0.05 – 0.15) and red – high genetic structure (F_{ST} > 0.15). Populations marked with an '*' correspond to *C. criuva* subsp. *criuva*. The shortest (89 km) and the longest (1599 km) distances between populations are in bold, as well as the only two estimates that are not significative according to the bootstrap confidence interval (0.014 and 0.007)

	PR	SPR	RS	SPV	RJ	SC	PRC	BA*	MG*	DF*
PR		302	321	407	486	128	149	1280	578	715
SPR	**0.014**		587	106	187	373	397	1040	281	639
RS	0.102	0.045		685	772	214	375	**1599**	866	1016
SPV	0.091	0.046	0.058		**89**	473	499	969	183	650
RJ	0.147	0.102	0.111	0.100		559	565	884	95	622
SC	0.027	**0.007**	0.077	0.074	0.100		247	1390	654	843
PRC	0.077	0.087	0.162	0.120	0.242	0.118		1279	650	644
BA*	0.549	0.522	0.525	0.453	0.486	0.506	0.567		803	752
MG*	0.447	0.422	0.450	0.379	0.448	0.410	0.503	0.387		622
DF*	0.517	0.486	0.481	0.423	0.446	0.471	0.547	0.328	0.454	

presented lower genetic and geographical differences than the first cluster, although it also included pairs of populations that are quite geographically distant from one another, e.g. RJ and RS, which are 772 km apart. This distance is even greater than those between pairs of *C. criuva* subsp. *criuva* populations; MG and DF are 622 km apart, and DF and BA are 752 km apart. Within the second cluster, RJ is slightly isolated from the rest of the populations, even though it is geographically close to some of them.

The results obtained through the software Structure were similar to those observed in the analyses of pairwise F_{ST} and Nei's genetic distances. We observed a clear separation of individuals in two different clusters that correspond precisely to each *C. criuva* subspecies (Fig. 1) (see Supplementary Material, Fig. S1 for mean Ln estimate probability and Fig. S2 for delta *K*). Only two hybrids were detected, one in the MG population (0.096) and the other in the PRC population (0.464).

The AMOVA results for nuSSRs based on the hierarchical a priori hypothesis (subspecies, biomes and vegetation) indicated that the greatest significant variation was retained for the subspecies level (Φ_{CT} = 0.33). The other two levels, biome and vegetation, displayed significant but lower values (Φ_{CT} = 0.29 and 0.27, respectively) (Supplementary Material, Table S7).

Demographic contraction and phylogeographic structure

The software Barrier was used to identify putative geographical breaks based on geographical and genetic information. These breaks should be able to prevent gene flow, increasing the genetic differences between populations. The results showed the most probable barrier (barrier 1, Fig. 2) between sampling locations RJ and MG, exactly where the subspecies *C. criuva* subsp. *parviflora* ceases to exist and gives way to *C. criuva* subsp. *criuva*. This method allows the identification of as many breaks as the number of intersections between populations, but it ranks the possibilities according to their likelihood. The second most probable barrier (barrier 2) isolated population RJ, and the third most probable barrier isolated sampling location PRC (barrier 3), which is currently the most affected by human activities (Fig. 2). Based on barrier and genetic structure analyses (Figs 1, 2), we grouped sampling locations according to the first putative break: the southern group (S) encompasses *C. criuva* subsp. *parviflora* populations, whereas the northern group (N) is composed of *C. criuva* subsp. *criuva* populations. To verify phylogeographic structure between the N and S groups, we compared the results of F_{ST} and R_{ST} obtained using SPAGeDi. The structure is evident when the value of R_{ST} is found to be significantly greater than the value of F_{ST}. First, we tested the break using all sampling locations and considering the two groups created as two distinct populations. The F_{ST} value was 0.353 and the R_{ST} value was 0.255. We also tested the same break but included a smaller number of populations to minimize the effects of geographical distance. When we considered six populations, one of *C. criuva* subsp. *criuva* and five of *C. criuva* subsp. *parviflora,* we observed an increase in the value of R_{ST} (0.335) compared to F_{ST} (0.345). Considering only the two populations physically separated by the barrier (RJ and MG), F_{ST} (0.446) continued to be greater than R_{ST} (0.425) (Supporting Information, Table S6). The occurrence of a bottleneck was detected for two populations (RJ and RS, $P < 0.05$).

The first sPCA eigenvalue (Fig. 1) retained a global structure ($P < 0.05$) through a more gradual separation between populations of *C. criuva* subsp. *criuva* (white

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

and light grey) and *C. criuva* subsp. *parviflora* (black and dark grey). In addition, the three populations that are located more closely presented similar values of the first eigenvalue: MG, RJ and SPV (Fig. 1).

Plastid microsatellites

Haplotype distributions and molecular variance analyses

For plastid SSR analyses, we considered only the 244 individuals with no missing data. The number of individuals per population varied from 19 to 28. We obtained ten different haplotypes, distributed unevenly across sampling locations (Fig. 2). All *C. criuva* subsp. *criuva* populations (BA, MG and DF) and the *C. criuva* subsp. *parviflora* population RJ had exclusive haplotypes, whereas the other populations (RS, PRC, PR, SPV, SPR and SC) shared one haplotype. In addition, SPV had three other haplotypes, SPR had one more haplotype and SC had two more (Fig. 2). Plastid data supported the nuclear SSR results; we observed a high level of structure for *C. criuva* subsp. *criuva* and a much lower level for *C. criuva* subsp. *parviflora*. We also observed that population RJ stands out with a unique genetic composition.

The haplotype network shows a similar pattern to that observed through the nuclear SSR analyses. Overall, the haplotypes corresponding to the subspecies *C. criuva* subsp. *criuva* are more closely related than those corresponding to the subspecies *C. criuva* subsp. *parviflora*. However, the haplotype of population MG, which belongs to the subspecies *C. criuva* subsp. *criuva*, is more closely related to the haplotypes of the other subspecies (Fig. 2).

We obtained a high Φ_{PT} (0.821, $P < 0.05$), indicating that the genetic structure is greater among the ten populations than within them. This pattern can be easily verified by the AMOVA percentages found for differentiation among populations (82%) and within populations (18%).

Niche modelling

Under current and past climatic conditions, we observed high AUC values (> 0.963) when considering both paleoclimatic models tested (MIROC and CCSM) and the two past historical periods selected (6000 BP and 21 000 BP), indicating that our model accuracy was high. The results showed distinct trends for both subspecies: we observed a drastic retraction in climatically suitable areas since the LGM for *C. criuva* subsp. *criuva*, whereas after this reduction, between the present and the mid-Holocene, the possible habitats of subspecies *C. criuva* subsp. *parviflora* underwent a slight expansion. The niche equivalency test showed no statistically significant overlap between the niches determined for each subspecies using either of the similarity indices tested: Schoener's *D* (observed $D = 0.219$; $P > 0.01$) and Warren's *I* (observed $I = 0.438$; $P > 0.01$) (Supporting Information, Fig. S3).

DISCUSSION

Genetic differences between subspecies and other important traits to be considered

Based on the clustering analyses (Fig. 1) and on the pairwise F_{ST} values (Table 2), both subspecies of *C. criuva* formed distinct genetic clusters. The identification of only two individuals as putative hybrids is also evidence supporting restricted gene flow between populations of both subspecies. This possibility is corroborated by the lack of morphological intermediates observed between them, although *Clusia* L. is known to produce viable offspring after artificial hybridization even between distantly related species (V. Bittrich, personal observation). The phylogenetic relationships recovered through the haplotypes also indicates a separation between subspecies (Fig. 2). That the haplotype of population MG (*C. criuva* subsp. *criuva*) is more closely related to the haplotypes of *C. criuva* subsp. *parviflora* possibly indicated that spatial proximity among these populations has enabled gene flow in the recent past, as plastid DNA is more conserved than nuclear DNA (Wolfe, Li & Sharp, 1987) and is thus able to recover a more ancient evolutionary history than nuclear SSRs. Additionally, the fact that plastid material can only be maternally inherited indicates that the bird dispersion observed for *C. criuva* (Passos & Oliveira, 2002) might prove more effective than pollination in reaching longer distances.

Although investigating the morphological differences between *C. criuva* subsp. *criuva* and *C. criuva* subsp. *parviflora* was beyond our objective, the stamen serves, essentially, as the only proper morphological diagnosable character, coinciding perfectly with the two genetic clusters identified in the samples (Bittrich, 2003). Considering other traits, such as floral scents, both subspecies (at the time, *C. criuva* subsp. *parviflora* was considered a distinct species) differ substantially in their floral scent compositions (Nogueira *et al.*, 2001). This trait, coupled with the possibility of hybridization between distantly related *Clusia* spp. (V. Bittrich, personal observation), may be a powerful driver of speciation since it is closely connected with the pollinators a species is able to attract (Nogueira *et al.*, 2001).

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

Disperser and pollinator characteristics might also be responsible for the relatively low degree of polymorphism observed within populations. Seeds from the same locule are united by their arils in small packages (diaspores) that are swallowed whole by birds (Passos & Oliveira, 2002), allowing them to sprout in a restricted area. However, ants have been observed secondarily dispersing the seeds from fallen fruits and bird faeces, contributing to the germination of various seeds from the same diaspore in different locations. Thus, the ants decrease sibling competition by separating sister seeds, even though they end up promoting only local movements, in contrast to birds (Passos & Oliveira 2002). At the same time, the nocturnal beetle reported to be responsible for pollination of *C. criuva* subsp. *parviflora* (Correia *et al.*, 1993 *sub nomen C. criuva*; their photographs clearly show that they studied subsp. *parviflora*) is extremely small and has limited flying capacity, restricting its visits to nearby individuals. Since these individuals have a high chance of being closely related to each other, due to limited seed dispersal, the beetles may frequently promote gene flow between closely related plants. The pollination biology of *C. criuva* subsp. *criuva* has not been investigated yet.

Habitat influence, phylogeographic breaks and niche modelling

According to the ecological species concept proposed by Van Valen (1976), an adaptive zone may have fixed natural or unnatural boundaries or it can be unimpeded by any physical frontier, which does not mean that subdivisions cannot develop (Van Valen, 1971). Expansion and fragmentation events have proven to influence differentiation among plant populations in forests and grasslands (Leal, da Silva & Pinheiro, 2016). Consequently, different abiotic conditions displayed by the two domains where plants of the *C. criuva* complex occur may also have played an import role in defining the genetic structure of both subspecies. Analyses performed with the software Barrier suggested three distinct breaks (Fig. 2). The first (and therefore main) break was inferred between populations MG and RJ, coinciding with subspecies divergence and suggesting that this level is indeed the uppermost hierarchical level of genetic structure observed with STRUCTURE and through the relationship between R_{ST} (0.255) and F_{ST} (0.353) values (Supporting Information, Table S6). This result is also reinforced by the AMOVA performed with nuclear SSR data, which shows that the 'subspecies' level is responsible for explaining the largest variation compared to 'types of vegetation' or 'biomes' (Supporting Information, Table S7). The gradient shown by the first sPCA eigenvalue also coincides with the region where the first and main break was identified (Fig. 1). Thus, it is possible to infer that this region, where populations MG, RJ and SPV are located, has probably undergone recent interruption in gene flow while still sharing genetic variability to some extent. The pairwise F_{ST} estimates between populations MG and RJ (0.448) also point to a great differentiation between them, despite their being one of the closest pairs of populations (95 km), indicating again that geographical distances are not solely responsible for the observed differences between the two subspecies.

The second break, however, isolated population RJ, which is the only population of *C. criuva* subsp. *parviflora* situated in the mountain chain Serra da Mantiqueira (IBDF and FBCN, 1982). The remaining populations of this subspecies, except for PRC, all occurred either in restinga forests in plains close to the sea or at Serra do Mar, which is a coastal mountain range in Brazil (SP) but currently not connected to Serra da Mantiqueira (Fig. 2). This distribution agrees with the finding of a phylogeographic break associated with Serra da Mantiqueira for the orchid species *Epidendrum denticulatum* Barb.Rodr., characterizing this part of RJ as a transition zone between lineages (Pinheiro *et al.*, 2013). In addition, the high level of diversity found in the Atlantic Forest could be explained by speciation processes originating from restrictions in gene flow between different mountain ranges (Scarano, 2002, Leal *et al.*, 2016).

The third break isolated another population of *C. criuva* subsp. *parviflora*, PRC, from the remaining populations (RS, SC, PR, SPR and SPV). We suspect that this isolation has to do with the fact that PRC is located in the topographic region called Segundo Planalto Paranaense (IAP, 2002). This region is surrounded by Serra Geral and Escarpas Devonianas, which are areas at higher elevations, therefore isolating this population from other sampling locations of *C. criuva* subsp. *parviflora*.

Nonetheless, notwithstanding the remarkable differences between cerrado and Atlantic Forest, there is an astounding gradient of vegetation variation within these domains (MMA/IBAMA, 2012). Scarano (2002) noted that the Atlantic Forest along the Brazilian coast should in fact be considered a great mosaic of distinct vegetation types. In the eastern portion of the Atlantic Forest, the dense ombrophilous forest is predominant and is divided into five different altitude ranges, with *C. criuva* subsp. *parviflora* occupying three of them (lowland dense ombrophilous forest, submontane dense ombrophilous forest and montane dense ombrophilous forest), thus showing great adaptability (Veloso, Rangel Filho & Lima, 1991). Conversely, *C. criuva* subsp. *criuva* occupies a more restricted habitat, occurring in only two of the many distinct and broadly recognized types of vegetation in

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

the cerrado (gallery forests and rocky fields) commonly found, respectively, along riversides and on high plateaus of the cerrado. The global structure identified for the *C. criuva* complex through sPCA demonstrates a strong correlation between the genetic variation and spatial distribution. However, a clear correspondence to a specific type of vegetation could not be identified. Therefore, it remains important to consider ecological effects when considering the genetic variability of this species, especially since *C. criuva* exhibits a cryptic pattern of genetic structure.

Climatic oscillations during the Quaternary have impacted the phylogeographic structure of certain species from the cerrado in different ways. Whereas tree species have usually shown a retraction in range during the LGM, plants occurring in rocky fields experienced an opposite pattern of range retraction during the Last Interglacial (LIG) (Leal *et al.*, 2016). This phenomenon has been considered to be likely due to the shift in climatic conditions between the LGM and LIG, which enabled a warmer and wetter climate to be initiated, possibly restricting cerrado vegetation to restricted areas in south-eastern Brazil (Behling, 2003). Our models of potential distribution indicated a strong habitat retraction during the LIG for *C. criuva* subsp. *criuva* (Fig. 3). Although it can be considered a tree species, it has also been found in rocky fields, making it possible to extend this pattern to our own study.

However, *C. criuva* subsp. *parviflora* seems to have experienced a slight range retraction during the LGM, which agrees with the suggestion that humid forests have retracted during drier and colder periods (Behling & Negrelle, 2001, Turchetto-Zolet *et al.*, 2012). Such retraction probably influenced the genetic diversity of this subspecies, leading to specific genetic patterns that are typically observed for populations that underwent bottleneck events. For *C. criuva*, we identified the occurrence of a bottleneck for only two populations (RJ and RS), although most of our microsatellite markers presented HWE deviations, so these results should be interpreted with caution (see Luikart *et al.*, 1998). Both populations are situated on the periphery of the distribution of *C. criuva* subsp. *parviflora* (northern – RJ; southern – RS), suggesting that they were more prone to the effects of Quaternary climatic oscillations, as peripheral populations are susceptible to sub-optimal conditions that negatively affect survival, reproduction and population growth (Brown, 1984).

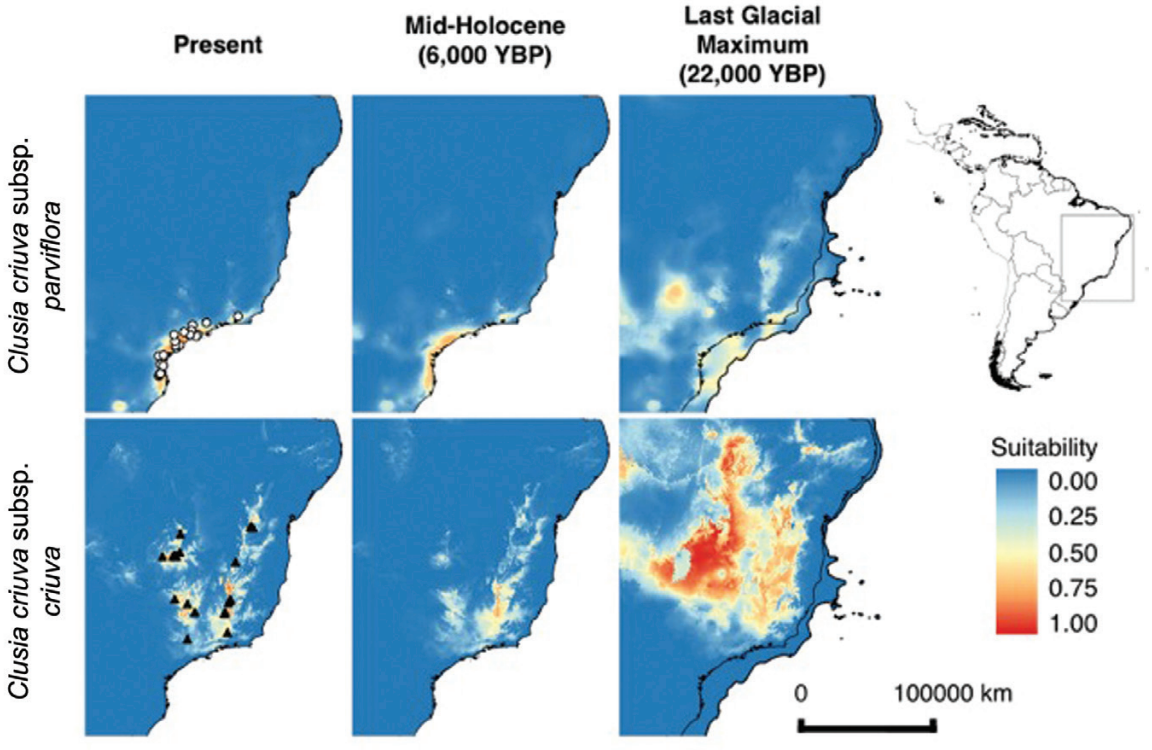

Figure 3. Models of habitat distribution considering the two subspecies separately (*C. criuva* subsp. *parviflora* and *C. criuva* subsp. *criuva*). Analyses were conducted for three temporal periods: Present, Mid-Holocene (6000 BP) and Last Glacial Maximum (22,000 BP). White circles indicate occurrence points of *C. criuva* subsp. *parviflora*, and black triangles indicate occurrence points of *C. criuva* subsp. *criuva* used to generate the models.

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

The niche equivalency test indicated that the niches are substantially different between the two subspecies (Supporting Information, Fig. S3). The environmental differentiation between *C. criuva* subsp. *criuva* and *C. criuva* subsp. *parviflora* can be interpreted as evidence of the evolutionary independence of the lineages. Moreover, closely related species, although expected to share some degree of niche similarity due to phylogenetic niche conservatism (Crisp & Cook, 2012), rarely occupy the exact same niche (Warren *et al.*, 2008). This divergence may be related to historical demographic processes, since this species apparently underwent significant range shifts over the last 22 000 years (Fig. 3).

Polymorphism of the nuclear microsatellites and intrapopulation estimates

According to Petit & Hampe (2006), allogamous tree species usually have large population dimensions, outcrossing mating systems and a long life cycle. However, although *C. criuva* fits the previous description, we unexpectedly observed a relatively low level of intrapopulation polymorphism for both types of molecular markers, especially for the nuclear SSRs. Nonetheless, this expectation might be proved wrong when populations are undergoing disturbances or habitat contraction (Ward *et al.*, 2005). As far as mating systems are concerned, self-incompatible and outcrossing self-compatible species suffer the most through habitat contraction because of their reliance on animal pollination and/or seed dispersion, which makes them dependent on the behaviour and abundance of existing pollinators and dispersers (Aguilar *et al.*, 2008). Overall, the values of observed heterozygosity (H_O) were higher than those obtained for expected heterozygosity (H_E) (Supplementary Material, Table S2), which resulted in negative values of F_{IS}. These findings are expected since the species presents an obligatory outcrossing mating system, which makes it more difficult to observe homozygotes caused by inbreeding. We observed an extremely high value of F_{ST} across loci and, even when considering the ENA correction for null alleles (Chapuis & Estoup, 2007), F_{ST} results showed no pronounced alterations: 0.329 without correction and 0.321 with correction. This finding is in accordance with the observed null allele frequencies, 87% of which were considered low and with only 1% presenting high frequency (Supplementary Material, Table S4).

CONCLUSIONS AND FUTURE DIRECTIONS

We obtained results that clearly show the two subspecies of the *C. criuva* complex as separately evolving lineages. Combined with unambiguous identifiability of fertile plant material, this evidence strongly suggests that a taxonomic review is required to recognize *Clusia criuva* subsp. *criuva* and *C. criuva* subsp. *parviflora* as distinct species. These lineages form distinct genetic clusters that are geographically isolated from each other. Two of the three breaks suggested in the analysis, which isolate two *C. criuva* subsp. *parviflora* populations (RJ and PRC), coincide precisely with geographical barriers. In addition, we observed genetic differentiation even between geographically close populations of the distinct subspecies. Niche modelling results revealed that *C. criuva* subsp. *criuva* may have inhabited a much broader area during the LGM than it now occupies. This finding corroborates our genetic findings; a great reduction of the distribution may have contributed to the further isolation of the two subspecies. Such an event probably greatly diminished gene flow between them, contributing to the formation of distinct genetic clusters. We suggest that further analysis should be carried out, including the use of molecular markers capable of capturing an older evolutionary history.

FUNDING

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (2012/51781-0), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES (grant to MBSC) and CAPES-Computational Biology Program (grant to APS) and the National Council for Scientific and Technological Development – CNPq (grants to APS 309661/2014–5 and MCEA 312479/2013-1).

ACKNOWLEDGEMENTS

This study was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, number 2012/51781-0, which is gratefully acknowledged. MCEA and APS thank CNPq for research grants (312479/2013-1 and 309661/2014–5). MBSC thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for an MS fellowship. AAP thanks FAPESP for a PhD scholarship (2013/11137-7) and CAPES - Computational Biology Program for a post-doctoral fellowship. GMM thanks FAPESP for post-doctoral fellowships (13/08086-1 and 14/22821–9). The authors thank Dr Miklos Maximiliano Bajay, Dr Alexandre Rizzo Zuntini and Prof. Dr Fábio Pinheiro for their valuable input. We thank Nathália Streher for help with the herbarium material. We are grateful to Dr Mariana Barreto, Dr Fernanda Ancelmo and Dr André Conson for help with the nuclear SSR library.

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

We thank all the national and state parks visited and the employees for their great help.

REFERENCES

Adams M, **Raadik TA**, **Burridge CP**, **Georges A. 2014.** Global biodiversity assessment and hyper-cryptic species complexes: more than one species of elephant in the room? *Systematic Biology* **63:** 518–533.

Aguilar R, **Quesada M**, **Ashworth L**, **Herrerias-Diego Y**, **Lobo J. 2008.** Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Molecular Ecology* **17:** 5177–5188.

Alvarado-Serrano DF, **Knowles LL. 2013.** Ecological niche models in phylogeographic studies: applications, advances and precautions. *Molecular Ecology Resources* **14:** 233–248.

Balloux F, **Lugon-Moulin N. 2002.** The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11:** 155–165.

Bandelt H-J, **Forster P**, **Röhl A. 1999.** Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37–48.

Behling H. 2003. Late glacial and Holocene vegetation, climate and fire history inferred from Lagoa Nova in the southeastern Brazilian lowland. *Vegetation History and Archaeobotany* **12:** 263–270.

Behling H, **Negrelle RRB. 2001.** Tropical rain forest and climate dynamics of the Atlantic lowland, southern Brazil, during the Late Quaternary. *Quaternary Research* **56:** 383–389.

Billotte N, **Lagoda PJL**, **Risterucci A-M**, **Baurens F-C. 1999.** Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. *Fruits* **54:** 277–288.

Bittrich V. 2003. Clusiaceae. In: Wanderley MGL, Shepherd GJ, Giulietti AM, Melhem TS, eds. *Flora Fanerogâmica do Estado de São Paulo*. São Paulo: FAPESP: RiMa, 45–62.

Brown JH. 1984. On the relationship between abundance and distribution of species. *The American Naturalist* **124:** 255–279.

Caddah MK, **Campos T**, **Zucchi MI**, **de Souza AP**, **Bittrich V**, **do Amaral MCE. 2013.** Species boundaries inferred from microsatellite markers in the *Kielmeyera coriacea* complex (Calophyllaceae) and evidence of asymmetric hybridization. *Plant Systematics and Evolution* **299:** 731–741.

Chapuis M-P, **Estoup A. 2007.** Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24:** 621–631.

Chung SM, **Staub JE. 2003.** The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. *Theoretical and Applied Genetics* **107:** 757–767.

Cornuet JM, **Luikart G. 1997.** Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144:** 2001–2014.

Correia MCR, **Ormond WT**, **Pinheiro MCB**, **Lima HA. 1993.** Estudo da biologia floral de *Clusia criuva* Camb. Um caso de mimetismo. *Bradea* **6:** 209–219.

Creste S, **Neto A**, **Figueira A. 2001.** Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Molecular Biology Reporter* **19:** 299–306.

Crisp MD, **Cook LG. 2012.** Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes. *New Phytologist* **196:** 681–694.

De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* **56:** 879–886.

Dempster AP, **Laird NM**, **Rubin DB. 1977.** Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* **39:** 1–38.

Doyle JJ, **Doyle JL. 1987.** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, Botanical Society of America* **19:** 11–15.

Earl DA, **vonHoldt BM. 2012.** STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359–361.

Edwards CE, **Judd WS**, **Ionta GM**, **Herring B. 2009.** Using population genetic data as a tool to identify new species : *Conradina cygniflora* (Lamiaceae), a new, endangered species from Florida. *Systematic Botany* **34:** 747–759.

Elith J, **Phillips SJ**, **Hastie T**, **Dudík M**, **Chee YE**, **Yates CJ. 2011.** A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17:** 43–57.

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

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

Falush D, **Stephens M**, **Pritchard JK. 2003.** Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164:** 1567–1587.

Goudet J. 2001. FSTAT, a program to estimate and test gene diversity and fixation indices, version 2.9.3.2. Available at: http://www2.unil.ch/ popgene/softwares/fstat.htm, last accessed March 2018.

Hanley AJ, **McNeil JB. 1982.** The meaning and use of the area under a receiver operating characteristic (ROC) Curve. *Radiology* **143:** 29–36.

Hardy OJ. 2003. Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Molecular Ecology* **12:** 1577–1588.

Hardy OJ, **Vekemans X. 2002.** SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2:** 618–620.

Heringer EP, **Barroso GM**, **Rizzo JA**, **Rizzini CT. 1977.** A Flora do Cerrado. In: Ferri MG, ed. *IV Simpósio sobre o Cerrado*. São Paulo: Editora Universidade de São Paulo, 211–232.

Hickerson MJ, **Carstens BC**, **Cavender-Bares J**, **Crandall KA**, **Graham CH**, **Johnson JB**, **Rissler L**, **Victoriano PF**, **Yoder AD. 2010.** Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution* **54:** 291–301.

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

Hijmans RJ, **Cameron SE**, **Parra JL**, **Jones PG**, **Jarvis A. 2005.** Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965–1978

IAP. **2002.** *Plano de manejo do Parque Estadual do Cerrado*. Curitiba: Instituto Ambiental de Paraná.

IBDF, FBCN. **1982.** *Plano de Manejo: Parque Nacional do Itatiaia*. Brasília: IBDF, FBCN.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403–1405. #8232;

Jombart T, **Devillard S**, **Dufour A-B**, **Pontier D. 2008.** Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101:** 92–103. #8232;

Keenan K, **McGinnity P**, **Cross TF**, **Crozier WW**, **Prodöhl PA. 2013.** diveRsity: an R package for the estimation of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4:** 782–788.

Leal BSS, **da Silva CP**, **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.

Luikart G, **Sherwin WB**, **Steele BM**, **Allendorf FW. 1998.** Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Molecular Ecology* **7:** 963–974.

Manni F, **Guerard E**, **Heyer E. 2004.** Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology* **76:** 173–190.

Miller M. 1997. *Tools for population genetic analyses (TFPGA) 1.3: a Windows program for the analysis of allozyme and molecular population genetic data*. Department of Fisheries and Wildlife. Available at: http://bioweb.usu.edu/mpmbio/index.htm.

MMA/IBAMA. **2011.** *Monitoramento do desmatamento nos biomas brasileiros por satélite. Acordo de cooperação técnica MMA/IBAMA. Monitoramento do bioma cerrado 2009- 2010*. Brasília: MMA/IBAMA.

MMA/IBAMA. **2012.** *Monitoramento do desmatamento nos biomas brasileiros por satélite. Acordo de cooperação técnica MMA/IBAMA. Monitoramento do bioma Mata Atlântica 2008 a 2009*. Brasília: MMA/IBAMA.

Monmonier M. 1973. Maximum-difference barriers: an alternative numerical regionalization method. *Geographical Analysis* **3:** 245–261.

Moran P. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society B (Methodological)* **10:** 243–251.

Moran P. 1950. Notes on continous stochastic phenomena. *Biometrika* **37:** 17–23.

Nei M. 1972. Genetic distance between populations. *The American Naturalist* **106:** 283–292.

Nogueira PC de L, **Bittrich V**, **Shepherd GJ**, **Lopes AV**, **Marsaioli AJ. 2001.** The ecological and taxonomic importance of flower volatiles of *Clusia* species (Guttiferae). *Phytochemistry* **56:** 443–452.

Padial JM, **Miralles A**, **De la Riva I**, **Vences M. 2010.** The integrative future of taxonomy. *Frontiers in Zoology* **7:** 16.

Passos L, **Oliveira PS. 2002.** Ants affect the distribution and performance of seedlings of *Clusia criuva*, a primarily bird-dispersed rain forest tree. *Journal of Ecology* **90:** 517–528.

Peakall R, **Smouse PE. 2012.** GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28:** 2537–2539.

Pessoa EM, **Alves M**, **Alves-Araújo A**, **Palma-Silva C**, **Pinheiro F. 2012.** Integrating different tools to disentangle species complexes: a case study in *Epidendrum* (Orchidaceae). *Taxon* **61:** 721–734

Petit RJ, **Hampe A. 2006.** Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics* **37:** 187–214.

Phillips SJ, **Anderson RP**, **Schapire RE. 2006.** Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190:** 231–259.

Pinheiro F, **Cozzolino S**, **de Barros F**, **Gouveia TMZM**, **Suzuki RM**, **Fay MF**, **Palma-Silva C. 2013.** Phylogeographic structure and outbreeding depression reveal early stages of reproductive isolation in the Neotropical orchid *Epidendrum denticulatum*. *Evolution* **67:** 2024–2039.

Powell W, **Morgante M**, **Andre C**, **McNicol JW**, **Machray GC**, **Doyle JJ**, **Tingey SV**, **Rafalski JA.1995.** Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Current Biology* **5:** 1023–1029.

Pritchard JK, **Stephens M**, **Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

R Development Core Team. **2016.** *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rousset F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8:** 103–106.

Scarano FR. 2002. Structure, function and floristic relationships of plant communities in stressful habitats marginal to the Brazilian Atlantic Rainforest. *Annals of Botany* **90:** 517–524.

Schoener TW. 1968. *Anolis* lizards of Bimini: resource partitioning in a complex fauna. *Ecology* **49:** 704–726.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139:** 457–462.

Turchetto-Zolet AC, **Pinheiro F**, **Salgueiro F**, **Palma-Silva C. 2012.** Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology* **22:** 1193–1213

Van Valen L. 1971. Adaptative zones and the orders of mammals. *Evolution* **25:** 420–428.

Van Valen L. 1976. Ecological species, multispecies, and oaks. *Taxon* **25:** 233–239.

Vaasen A, **Scarano FR**, **Hampp R. 2007.** Population biology of different *Clusia* species in the state of Rio de Janeiro. In: Lüttge U, ed. Clusia, *: a woody Neotropical genus of remarkable plasticity and diversity. Ecological Studies*, **Vol. 194**. Heidelberg: Springer, 117–127.

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019

Veloso HP, **Rangel Filho ALR**, **Lima JCA. 1991.** *Classificação da vegetação brasileira adaptada a um sistema universal*. Rio de Janeiro: IBGE.CDDI.

Ward M, **Dick CW**, **Gribel R**, **Lowe AJ. 2005.** To self, or not to self... a review of outcrossing and pollen-mediated gene flow in Neotropical trees. *Heredity* **95:** 246–254.

Warren DL, **Glor RE**, **Turelli M. 2008.** Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62:** 2868–2883.

Warren DL, **Glor RE**, **Turelli M. 2010.** ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **33:** 607–611.

Weir BS, **Cockerham CC. 1984.** Estimating F-statistics for the analysis of population structure. *Evolution* **38:** 1358–1370.

Weising K, **Gardner RC. 1999.** A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* **42:** 9–19.

Wolfe KH, **Li WH**, **Sharp PM. 1987.** Rates of nucleotide substitution vary greatly among plant mithocondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America* **84:** 9054–9058.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Overview table with specific details for each pair of primer developed for the species *C. criuva* and used in the present investigation.

Table S2. Number of individuals sampled per population (*N*), number of alleles (*A*), allelic richness (A_{ri}), number of private alleles (A_p), observed heterozygosity (H_o), expected heterozygosity (H_e), the difference between H_o and H_e for each population, and Wright's intrapopulation fixation index (F_{IS}). Grey shading indicates Hardy–Weinberg equilibrium deviation ($P < 0.05$). All statistics were based on information from nuSSR.

Table S3. Descriptive statistics per loci and per population of the microsatellite markers described in this paper. H_O: observed heterozygosity, H_E: expected heterozygosity; F_{IS}: inbreeding coefficient, *A*: allelic richness, *N*: number of alleles.

Table S4. Null allele frequencies obtained with the software FreeNA using the algorithm proposed by Dempster *et al.* (1977). Values in red indicate potantial null alleles (null frequency > 0.20). Low frequency values are shown in white.

Table S5. Pairwise F_{ST} estimates (PWF_{ST}) between all the different populations sampled and the correspondent Lower Bootstrap Confidence Interval (LBCI) and Upper Bootstrap Confidence Interval (UBCI) of 95%. Highlighted in red are the non-significative values.

Table S6. Potential phylogeographic break between MG and RJ with the number of populations included in the analysis in parenthesis. F_{ST} and R_{ST} values calculated through SPAGeDi across each of the three barriers, with the lower and upper confidence intervals for R_{ST}.

Table S7. AMOVA table for *C. criuva* based on the variation of nuclear microsatellites. The software ARLEQUIN was used to perform this analysis considering three different hierarchical levels: 'subspecies', 'biome' and 'vegetation'.

Table S8. List of the vouchers collected for each sampling location and deposited at the herbarium UEC.

Figure S1. Mean of estimate of Ln probability of data output obtained with the software STRUCTURE and analysed in the online software Structure Harvester.

Figure S2. Delta K estimate of data output obtained with the software STRUCTURE and analysed in the online software Structure Harvester.

Figure S3. Niche equivalence tests (Warren *et al.*, 2008) for *Clusia criuva* subsp. *parviflora* and *Clusia criuva* subsp. *criuva*. On the left side (I): Warren's *I* values distribution for 99 random repetitions. On the right side (D): Schoener's *D* values distribution for 99 random repetitions. Dashed lines represent observed values. Values of *D* and *I* statistics for all 99 random simulations are available in Supplementary Material (Table S2).

Downloaded from https://academic.oup.com/botlinnean/article-abstract/190/1/67/5480525 by Universidade Estadual de Campinas user on 05 December 2019