Guidelines for the restoration of the tropical timber tree Anacardium excelsum: first input from genetic data

Kelly Bocanegra-González, Marie-Laure Guillemin

Tree Genetics & Genomes

Guidelines for the restoration of the tropical timber tree Anacardium excelsum: first input from genetic data

--Manuscript Draft--

Manuscript Number:	TGGE-D-17-00267R2
Full Title:	Guidelines for the restoration of the tropical timber tree Anacardium excelsum: first input from genetic data
Article Type:	Original Article
Funding Information:	Universidad del Tolima (043-16) — Mrs Kelly Tatiana Bocanegra González
Abstract:	Translocation of trees has been used as a common method to mediate genetic conservation and restoration of forests. However, very few programs include strategies developed to recover or maintain the genetic diversity of the translocated species. Anacardium excelsum is a tree native to the tropics of America that is extensively used in forestry. In Colombia, restoration of forests through the translocation of native species has regained importance, and A. excelsum has been recently included in the National Strategy for Plant Conservation. Thus, in order to define the level of genetic structure and the level of genetic diversity within certain regions where remnants of the seasonally dry tropical forests (SDTF) of Colombia have been retained, we genotyped 106 trees using nuclear Inter-Simple Sequence Repeats (ISSR) and sequenced two non-coding chloroplast loci for these specimens. Our ISSR dataset revealed, the existence of a gradient in genetic diversity within A. excelsum with the most diverse remnants encountered in the south of the country, while the localities sampled in the Caribbean coast and in the Chicamocha canyon were less diverse. Chloroplast loci also pointed out the very low genetic diversity of A. excelsum from the Chicamocha canyon and we propose to prioritize this area within future conservation programs. Both chloroplast and nuclear markers supported the existence of genetic divergence between distinct regions of Colombia, uncovering genetic differences between inter-Andean, Caribbean and Chicamocha canyon A. excelsum remnants. Hence, we advise to choose the provenance of seeds or plants carefully before translocation, and to consider minimal mixing of material from different regions when initializing restoration programs for A. excelsum, in Colombia.
Corresponding Author:	Kelly T. Bocanegra-González, M.Sc Universidad del Tolima Ibagué, Tolima COLOMBIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universidad del Tolima
Corresponding Author's Secondary Institution:	
First Author:	Kelly Tatiana Bocanegra González, M.Sc
First Author Secondary Information:	
Order of Authors:	Kelly Tatiana Bocanegra González, M.Sc Marie-Laure Guillemin, Ph.D
Order of Authors Secondary Information:	
Author Comments:	

Guidelines for the restoration of the tropical timber tree *Anacardium excelsum*: first input from genetic data

Kelly T. Bocanegra-González[1] & Marie-Laure Guillemin[2, 3]

[1]Grupo de Investigación en Biodiversidad y Dinámica de Ecosistemas Tropicales, Facultad de Ingeniería Forestal, Universidad del Tolima. Ibagué 730001, Colombia.

ktbocanegrag@gmail.com

+52 55 68663854

[2]Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Casilla 567, Valdivia 5090000, Chile.

[3]CNRS, Sorbonne Universités, UPMC University Paris VI, UMI 3614, Evolutionary Biology and Ecology of Algae, Station Biologique de Roscoff, CS 90074, Place G. Teissier, 29680 Roscoff, France.

marielaure.guillemin@gmail.com

KT Bocanegra-González and M-L Guillemin contributed equally to this manuscript

Corresponding author: ktbocanegrag@gmail.com

Abstract

Translocation of trees has been used as a common method to mediate genetic conservation and restoration of forests. However, very few programs include strategies developed to recover or maintain the genetic diversity of the translocated species. *Anacardium excelsum* is a tree native to the tropics of America that is extensively used in forestry. In Colombia, restoration of forests through the translocation of native species has regained importance, and *A. excelsum* has been recently included in the National Strategy for Plant Conservation. Thus, in order to define the level of genetic structure and the level of genetic diversity within certain regions where remnants of the seasonally dry tropical forests (SDTF) of Colombia have been retained, we genotyped 106 trees using nuclear Inter-Simple Sequence Repeats (ISSR) and sequenced two non-coding chloroplast loci for these specimens. Our ISSR dataset revealed, the existence of a gradient in genetic diversity within *A. excelsum* with the most diverse remnants encountered in the south of the country, while the localities sampled in the Caribbean coast and in the Chicamocha canyon were less diverse. Chloroplast loci also pointed out the very low genetic diversity of *A. excelsum* from the Chicamocha canyon and we propose to prioritize this area within future conservation programs. Both chloroplast and nuclear markers supported the existence of genetic divergence between distinct regions of Colombia, uncovering genetic differences between inter-Andean, Caribbean and Chicamocha canyon *A. excelsum* remnants. Hence, we advise to choose the provenance of seeds or plants carefully before translocation, and to consider minimal mixing of material from different regions when initializing restoration programs for *A. excelsum,* in Colombia.

Key words: National Strategy for Plant Conservation of Colombia, translocations, seasonally dry tropical forest, molecular markers, ISSR, cpDNA

Introduction

During the last decades, increased efforts to restore habitats degraded by human activities throughout the world have been made (Meli et al. 2017). In the case of tropical forests, most of these efforts have focused on restoring species diversity and forest structure post-logging, while only a limited number have focused on controlling the erosion of genetic diversity or help restore this diversity among species of ecological and

economical importance (Ang et al. 2017). It is clear that, from a holistic point of view, the focus of ecological restoration should consider different levels of organization from genes to ecosystems (Hughes et al. 2008; Aerts and Honnay 2011) and that ecosystem restoration should begin with or at least include a strategy developed to recover or maintain genetic diversity of key stone species (Lesica and Allendorf 1999). According to McKay (2005), a better understanding of how genetic variation is distributed within and among populations (connected by varying degrees of gene flow) is essential for professionals in charge of restoration programs to maintain historic gene flow and local adaptation. Such reforestation programs that take into account the existence of regional genetic differentiation and possible local adaptations are, then, more likely granted to restore ecosystem services in the long term (Menges 2008).

Knowledge about the distribution of genetic variation is key when strategical decisions have to be taken in reforestation programs and should be used to assess, for example, the provenance of seeds or plants and the degree to which material of different populations is mixed before restoration (Hughes et al. 2008; Aerts and Honnay 2011). To mediate genetic conservation and restoration, the practice commonly referred as "translocation" is frequently used. Translocation consists in the transport of individuals, usually from healthier and bigger populations, to other areas of the species' distribution where populations are extinct or struggling. Translocation has been key in establishing new populations or re-establishing extirpated populations and to increase numbers of individuals in critically small populations (Menges 2008; Weeks et al. 2011). On the one hand, as in the case of the palm tree *Pseudophoenix sargentii* in Florida (Maschinski and Duquesnel 2007), this technique has proven to be a very successful mean to counteract genetic erosion, allowing to preserve the evolutionary potential of this endangered species. On the other hand, the unsuccessful restoration experiments through translocations of seeds from *Pinus pinaster* in Western Europe highlight the importance of acquisition of previous knowledge about the genetic structure of the species of interest (Benito-Garzón et al. 2013). Indeed, transplanted populations of *P. pinaster* have been unable to settle in sites characterized by climatic conditions different from the ones encountered in the seedlings' area of origin; a result that has been related to the existence of local adaptation across the species' natural range (Benito-Garzón et al. 2013). Unfortunately, it seems that the case of *P. pinaster* is not unique and that more negative than positive experiences are reported in the process of restoration through translocations (Keller et al. 2000; Menges et al. 2016). It is possible that a failure at integrating knowledge not only on the species

taxonomy, reproductive biology, demography, and ecology but also on the distribution of genetic diversity and functional adaptive traits between populations before choosing "whom to translocate where", account for the majority of unsuccessful restoration programs (Lesica and Allendorf 1999; Menges 2008; Aerts and Honnay 2011).

Moreover, restoring ecosystem functions and biological diversity often requires the use of species specific to that habitat (Lesica and Allendorf 1999; McKay 2005; Benito-Garzón et al. 2013), and the need to use native species in ecological restoration processes has been much emphasized lately. Although most processes of reclamation of degraded land have been accomplished using exotic species (Lamb 1998; D'Antonio & Meyerson 2002), in recent years, projects that include species native to the ecosystems to be recovered have increased. For example, in Colombia, there have been recent initiatives to restore the seasonally dry tropical forests (SDTF) using native species (Thomas et al. 2017; Galindo-Rodríguez and Roa-Fuentes 2017).

Anacardium excelsum is a native tree of the tropics of America, and can be found in riparian forests of the dry tropical and premontane areas from Guatemala to the northern part of South America, including the Guianas (Bernal et al. 2015). This species is used extensively in forestry and, although it has been described as having medium to slow growth rates, a recent study showed that it can reach the desired dimensions and characteristics for logging in approximately 20 years (Lozano et al. 2012). This feature, together with its capacity to grow in a wide range of soils and climatic conditions, makes *A. excelsum* an ideal native candidate for ecological restoration programs of riparian and lowland forests of Central and South America (Lamb 1998; Garen et al. 2011). In spite of these interesting features, habitat loss and overexploitation are now threatening *A. excelsum* populations, especially in Panamá and Colombia (Condit et al. 1995; Cárdenas and Salinas 2007). Recent interest in the restoration of *A. excelsum* populations through translocation has emerged, and the species has been included in the National Strategy for Plant Conservation of Colombia (NSPC) (Pizano and García 2014). Indeed, *A. excelsum* is part of the red list of threatened Colombian timber trees (Cárdenas and Salinas 2007) and is moreover characteristic to the SDTF, a habitat prioritized in the NSPC (Pizano and García 2014). The high fragmentation of SDTFs and the isolation of *A. excelsum* populations located in these remnants suggest that genetic differentiation could exist within the country. However, no information about the genetic structure and diversity of *A. excelsum* is currently available, and no specific molecular tools have been developed to work on these trees.

In the American tropics, floristic differences (Banda-R et al. 2016) and the genetic structure of species of trees commonly encountered in SDTF remnants (Caetano et al. 2008; Thomas et al. 2017) have been shown to be determined by historical (i.e. Pleistocene and Holocene) as well as more current changes affecting the ecosystem. Indeed, the Pleistocene Arc theory proposes that species characteristic to SDTF ecosystems had largely contiguous distributions during the late Pleistocene's cool and dry glacial period, and that contractions in present-day SDTF remnants correspond to a more recent change due to climate warming during the Holocene (Prado and Gibbs 1993; Pennington et al. 2000). *Anacardium excelsum* populations could thus have been much larger and highly connected in the past. Moreover, from the 1500's onwards, anthropogenic disturbance linked to human settlement after European colonization have severely degraded SDTF remnants (Valencia-Duarte et al. 2012; Pizano and García 2014). This could have led to further restriction in gene flow and loss of genetic diversity through bottlenecks in *A. excelsum*.

In order to define the level of genetic structure between regions, and the level of genetic diversity within regions and within remnants of *A. excelsum* in Colombia, we genotyped 106 trees collected from the department of Bolívar in the north, down to the department of the Valle del Cauca in the south using nuclear Inter-Simple Sequence Repeats (ISSRs), and also sequenced two non-coding chloroplast (cpDNA) loci. ISSR and cpDNA loci represent molecular markers with highly distinct evolutionary rates. ISSRs are highly polymorphic since their primers anneal directly on the SSR repeats distributed in the genome, which exhibit a high rate of mutation classically linked to additions and losses of repeat units due to DNA slippage (Tuisima et al. 2016). Conversely, the low rate of nucleotide substitutions, characteristic to the chloroplast genome in plants, has been used to develop molecular markers well suited for phylogenetic and phylogeographic studies (Hamilton 1999). Thus, we hypothesize that cpDNA loci will provide information about ancient gene flow (as their sequences are rather conserved over time), while our ISSR results will reflect more recent patterns of connectivity between *A. excelsum* remnants in Colombia. We further investigate population genetic footprints of past climatic changes and current anthropogenic disturbances in a tree species characteristic to the Colombian SDTFs. Ultimately, the aim of our study is to directly contribute to an effective and sustainable restoration of *A. excelsum* by providing the first genetic information available for this native species in Colombia.

Materials and methods

1. Field sampling

Young and healthy leaves of 106 reproductive individuals of *Anacardium excelsum* were sampled between September 2016 and March 2017, at 13 sites located in different fragments of Colombia's SDTF (see Table 1 and Results section for more details on region of origin and position of the sampling sites). Individuals separated by distances of at least 50 meters were sampled in order to reduce the sampling of genetically closely related individuals. *A. excelsum* is generally characterized by an outcrossing breeding system but the species also partially displays flower that are self-compatible (Ghazoul and McLeish 2001). Flowers are mainly pollinated by small bees (i.e. *Trigona* bees) and seed dispersal is mediated by bats. Sampling localities were selected based on occurrence data of *A. excelsum* obtained from botanical collection records. The locality of BOL, situated in the Caribbean coast (Table 1), was sampled from a botanical garden. Even if the botanical garden Guillermos Piñeres (BOL) includes a four-hectare patch of SDTF forest, which has been mostly dedicated to the study and protection of plants native to the Caribbean region of Colombia, no information about the origin of its *A. excelsum* trees is available (e.g., the possibility of seedlings or transplants originating from other regions). All dried tissue samples were stored in silica gel for posterior DNA extraction.

2. DNA extraction, PCR amplification and sequencing

Total genomic DNA was extracted using the protocol described by Doyle (1987).

Ten ISSRs primers were used for genotyping (Table S1). Amplification cycles consisted of a 5 minutes initiation step at 94°C, followed by 40 amplification cycles consisting of a denaturation step at 94°C for 1 minute, annealing at 46°C- 57°C for 45 seconds (Table S1) and a 72°C extension step for 2 minutes. Each PCR was set up in a volume of 20µl using a PCR mix of: 3µl of DNA (corresponding to more or less 10ng); 2µl of PCR 10X buffer; 2µl of 2.5mM dNTP; 1.8µl of 25mM $MgCl_2$; 0.9µl of 2.5mg/mL BSA; 0.4µl of 5U/µl Taq DNA polymerase (Invitrogen, Carlsbad, USA), 0.8µl of 2µM primer, and the remaining volume being completed by pure water. The amplification products were visualized on a 2% agarose gel (w/v) including 2µl of GelRed™ (Biotium, Fremont, USA) and separated by horizontal electrophoresis for 1h20. To ensure consistent scoring of the bands, all 106 individuals were scored on the same gel for each ISSR.

GeneRuler 100 bp (Thermo Scientific, Massachusetts, USA) was used as molecular weight marker. After electrophoresis, agarose gels were photographed by Gel Doc NEQ-11 (Neon Creative Labware, Santiago, Chile) under ultraviolet light.

Two chloroplast intergenic spacers, *trn*L-*trn*F (Taberlet et al. 1991) and *rpl20-rps12* (Hamilton, 1999) were sequenced. For *trn*L-*trn*F we used the primers B49317 5'-CGAAATCGGTAGACGCTACG-3' and A50272 5'-ATTTGAACTGGTGACACGAG-3' described in Taberlet et al. (1991). For *rpl20-rps12*, the primers *rpl20* 5'-TTTGTTCTACGTCTCCGAGC-3' and *rps12* 5'-GTCGAGGAACATGTACTAGG-3' were used (Hamilton 1999). Fragments were amplified by PCR in a 30µl volume using a mix similar to the one described above for the ISSRs but with a final concentration of 0.2mM $MgCl_2$, 0.4mM of each primer, and 0.2mM of each dNTP. A volume of 0.2µl of 5U/µl Taq DNA polymerase (Invitrogen, Carlsbad, USA) was used for each PCR. For both cpDNA loci, the amplification was done following the PCR cycling parameters proposed by Taberlet et al. (1991). Products were purified using MoBio UltraClean® DNA (Solana Beach, USA) and sequenced using both PCR primers, at Macrogen (https://dna.macrogen.com). The distinct haplotypes detected were deposited in GENBANK under accession numbers MG309720-MG309723 and MG309724-MG309726 for *rpl20-rps12* and *trn*L-*trn*F, respectively.

3. Statistical analysis

ISSR amplified fragments were scored manually as absent (0) or present (1) to build a binary matrix, and amplification failure of a sample was treated as missing data (-1). All ISSRs were amplified and scored twice in order to limit the number of missing data. Each band was assumed to represent the phenotype at a single biallellic locus (Crema et al. 2009). Classical parameters of genetic diversity suitable for dominant molecular markers, calculated using GenAlEx 6.5 (Peakall & Smouse 2012), were: average number of effective alleles (Ne), number of private bands (Pa), Shannon index (I), and expected heterozygosity (He). GenAlEx 6.5 was further used to calculate Nei's genetic distance (Nei 1978) in order to conduct a Principal Coordinates Analysis (PCoA), and to construct a dendrogram using Sequential Agglomerative, Hierarchical, and Nested Clustering (SAHN, in NTSYS-pc; Rohlf 1992). A population genetic structure analysis was carried out based on the model of Bayesian grouping implemented in STRUCTURE (Pritchard et al. 2000). The LOCPRIOR option was implemented to detect weak signals of population structure (i.e. includes a priori

sampling locations as prior information in the clustering algorithm; Hubisz et al. 2009). The admixture model was run, assuming correlated allele frequencies (Falush et al. 2003), using an initial burn-in of 500,000 followed by 1,000,000 Markov chain Monte Carlo (MCMC) iterations. Seven independent runs were performed with the tested number of clusters (K) varying from 1 to 13 (i.e., the number of sites with ISSR data + 1). Values obtained for the mean posterior probability of the data (Ln Pr(X|K)) and calculated for each value of K (Janes et al. 2017) were plotted in STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/structureHarvester/). The clear maximum of the Ln Pr(X|K) values was used to define the optimal K. CLUMPP (Jakobsson and Rosenberg, 2007) was used to combine the results of the seven independent runs, applying the greedy algorithm with 100,000 random input orders. To estimate the partitioning of genetic variability among regions and among individuals within regions (Excoffier et al. 1992), an analysis of molecular variance (AMOVA) was performed with GenAlEx 6.5 by resampling 9,999 times. Pairwise values of Φ_{ST} among regions were calculated with GenAlEx 6.5 and their significance was tested using 9,999 permutations. For both the AMOVA and Φ_{ST} calculations, regions were defined as in Table 1 (i.e. Magdalena river valley, Cauca river valley, Caribbean coast and Chicamocha canyon) and all individuals from distinct sampling sites within each region were pooled. A Mantel test was performed with GENETIX 4.05 (Belkhir et al. 1996-2004) to examine the relationship between Nei's genetic distance and geographic distance (estimated as ln of the geographical distance; Rousset 1997), and its significance was tested using 10,000 permutations. ISSR data from BOL (i.e., the botanical garden) were not included when performing the AMOVA among regions, the calculation of Φ_{ST} among regions, and the test for isolation by distance. Chloroplast sequences were edited in CHROMAS (Conor McCarthy, Griffith University, Australia, http://www.technelysium.com.au/chromas.html) and aligned in MEGA7 (Kumar et al. 2016). The number of haplotypes (nH), the number of polymorphic sites (S), gene diversity (Hd, Nei 1978) and nucleotide diversity (π, Nei and Li 1979) were computed using DnaSP (Rozas et al. 2003). Haplotypic richness (Re) was estimated for each region using CONTRIB and the rarefaction procedure of Petit et al. (1998). Haplotypic richness estimates reduce the effect of variation in sample size of *A. excelsum* among regions, with rarefaction standardizing it to the smallest sample size. A rarefaction size of N = 6 was used, which corresponded to the number of samples of the region with the lowest number of cpDNA sequences obtained (the Chicamocha canyon; Table 1). For both chloroplast loci, a median-joining network was constructed using NETWORK v

4.5 (Bandelt et al. 1999). A rooted UPGMA tree was constructed in MEGA7 (Kumar et al. 2016) using the concatenated *trn*L-*trn*F and *rpl20-rps12* sequences with 1,000 bootstrap replicates. Only one of each distinct sequences was used and gaps were coded by hand as a 5th character. *Anacardium occidentale* (Genbank accession number: KY635877) was used as an outgroup. To estimate the partitioning of genetic variability among regions and among individuals within region (Excoffier et al. 1992), analyses of molecular variance (AMOVA) were performed with ARLEQUIN v. 3.5 (Excoffier et al. 2005) by resampling 10,000 times. Pairwise values of Φ_{ST} among regions were calculated with ARLEQUIN v. 3.5 and their significance was tested using 10,000 permutations. For both, the AMOVA and Φ_{ST} calculations, regions were defined as in Table 1 and sequences from BOL were not included. As for ISSRs, all individuals from distinct sampling sites within each region were pooled.

Results

Of the ten ISSR primers used for genotyping, three showed a high level of missing data (>5%) and were removed before analyses. Three individuals of the SUC population (Caribbean coast region; Table 1) also showed a high level of missing data (i.e. more than half the ISSR primers did not amplify), and these individuals were removed before analyses. For six individuals showing extremely poor DNA quality, no PCR amplifications were possible. Thus, our ISSR analyses were performed on a total of 97 individuals, using seven primers. Additionally, 89 sequences of *trn*L-*trn*F of 725-729bp and 92 sequences of *rpl20-rps12* of 769bp were obtained.

1. Genetic diversity

The seven ISSR primers selected generated 57 fragments. Total number of fragments scored per primer ranged from 6 to 10 (Table S1) with an average of 69.9% polymorphic loci within locality (Table 1). Shannon index (I) and Nei's gene diversity (He) values ranged between 0.23-0.52 and 0.16-0.35, respectively, and the two estimators showed the same pattern within Colombia. The highest values of I and He were observed in the localities of the Magdalena river valley (i.e. ARM, MAR, IBG, VEN and CHA; Table 1). I and He decreased then from the Magdalena river valley to the Cauca river valley, to the Caribbean coast and, at last, to the Chicamocha canyon (Table 1). The lowest values were encountered in SAN (I = 0.23, He = 0.16; Table

1). Among the ISSR markers, only one private fragment was observed for the primer UBC-809 (Table S1) in VEN. When estimated by region, He was 0.35 in the Magdalena river valley, 0.32 in the Cauca river valley, 0.25 in the Caribbean coast and 0.18 in the Chicamocha canyon.

Three haplotypes were observed for *trn*L-*trn*F with five polymorphic sites that included a gap of 4bp (Table 2). For *rpl20-rps12*, four haplotypes and three polymorphic sites were encountered (Table 2). Haplotypic diversity within locality was generally very low and only one haplotype was observed for both loci in most localities (i.e. ARM, IBG, CHA, CLI, QUI and SAN; Table 2). The highest values of diversity were observed in the Caribbean coast (BOL: Hd = 1.00 and 0.67, $\pi = 4.57\ 10^{-3}$ and $2.60\ 10^{-3}$ for *trn*L-*trn*F and *rpl20-rps12*, respectively; CES: Hd = 1.00, $\pi = 2.60\ 10^{-3}$ for *rpl20-rps12*; Table 2). Intermediate values were observed in MED and MAR for both loci and in VEN for *rpl20-rps12* (Table 2). When estimated by region, haplotypic richness (Re) was 2.0 and 2.0 in the Caribbean coast, 0.9 and 0.8 in the Cauca river valley, 0.1 and 0.4 in the Magdalena river valley, and 0.0 and 0.0 in Chicamocha canyon (values given for *trn*L-*trn*F and *rpl20-rps12*, respectively).

2. Genetic structure

The first two axes of the PCoA explained 42.97% and 23.62% of the genetic variability observed with the ISSR data set for *A. excelsum* in Colombia (Fig.1). All localities from the Magdalena and Cauca river valleys and the botanical garden sampled within the region of the Caribbean coast (BOL), clustered together. COR and CES (Caribbean coast), and SAN (Chicamocha canyon) were isolated from the cluster of Magdalena, Cauca and BOL (i.e. along axis 1 for the first two localities and along axis 2 for SAN). The closest localities to SAN in the PCoA were CES and MED. The dendrogram based on Nei's genetic distance grouped the localities from the south/center of the country (ARM, MAR, IBG, VEN, CLI, QUI, CHA) and BOL (Fig. 2C). High genetic distances were observed between this group and COR and CES from the Caribbean coast, SAN from the Chicamocha canyon and MED located at the far north of the Cauca river valley (Fig. 2). Of all these four last localities, only SAN and MED clustered together. The Bayesian analysis carried out in STRUCTURE for all localities shows the existence of three predominant genetic clusters (Fig. 3A). Most individuals displayed varying degrees of mixed ancestry, and two of the three clusters present a rather restricted geographic distribution (Fig. 3B and C). Individuals assigned, with a membership of 0.60 or higher,

to the red genetic cluster were only observed in the Magdalena river valley (Fig. 3C) while individuals assigned to the yellow genetic cluster were observed in the Magdalena river valley, the northern part of the Cauca river valley and Chicamocha canyon (Fig. 3C). Interestingly, all trees sampled in SAN were assigned to the yellow genetic cluster and presented a very low level of admixture (Fig. 3B). Individuals assigned to the blue genetic cluster with a membership of 0.60 or higher were observed in all the sampled localities excepting SAN (Fig. 3C). Geographic isolation strongly influenced genetic differentiation among localities, as demonstrated by the significant correlation observed between genetic distance and ln of geographic distance ($r = 0.749$, $p < 0.01$) based on the Mantel test (Fig. S1). Significant genetic differences among regions were detected by AMOVA, however, only 6% of the total of genetic variation was attributed to differences among regions (Table 3). Significant genetic differentiation (Φ_{ST}) was observed between all pairs of regions, except between the Cauca river valley and the Caribbean coast (Table S2).

Due to the low level of haplotypic diversity, the structuration pattern observed with the two chloroplastic loci was not clear-cut (Table 2, Fig. 2B). For *trn*L-*trn*F, haplotype H2 was observed in all regions and H1 in all regions except in the Chicamocha canyon. Haplotype H3 was less frequent and encountered only in the Caribbean coast. For *rpl20-rps12*, H1 was encountered in all regions except the Chicamocha canyon and H2 in all regions except in the Caribbean coast. Haplotypes H3 and H4 were less frequent and only observed in the Magdalena river valley and the Caribbean coast. The UPGMA tree based on the concatenated sequences of *trn*L-*trn*F and *rpl20-rps12* revealed two branches supported by bootstrap values >75% (Fig. 4). With the exception of two individuals (one from BOL and one from MAR), the tree retrieved a geographical pattern with the first group corresponding to the Caribbean coast and the Chicamocha canyon and the second group to the Magdalena and the Cauca river valleys (Fig. 4). For both loci, AMOVAs attributed a majority of the variance to partition among geographic regions (75% and 57% for *trn*L-*trn*F and *rpl20-rps12*, respectively; Table 3). As for ISSR data, significant genetic differentiation (Φ_{ST}) was also observed between all pairs of regions, except between the Cauca river valley and the Caribbean coast, for the two chloroplastic loci (Table S2).

Discussion

For ISSR markers, our study revealed the existence of a gradient of genetic diversity within *Anacardium excelsum* in Colombia, with the most diverse remnants encountered in the south (i.e. the region of the Magdalena river valley and of the Cauca river valley), while the localities sampled in the Caribbean coast and in the Chicamocha canyon were less diverse. Except for the Chicamocha canyon where the lowest haplotypic richness was also observed for both *trn*L-*trn*F and *rpl20-rps12*, this pattern was not retrieved for the two chloroplastic loci. Both chloroplast and nuclear markers supported the existence of genetic divergence between distinct regions of Colombia. Caribbean coast exhibited unique haplotype diversity for *trn*L-*trn*F chloroplastic locus while two ISSR genetic clusters seem mostly restricted to the southeast of the country. Some of our results pointed to the existence of a pattern of spatial genetic structure in *A. excelsum*. Indeed, the dendrogram obtained using the ISSR data grouped most of the localities of the inter-Andean valleys (except MED; Fig. 2). In the same way, the concatenated sequences of the two chloroplastic loci characterizing samples from the Magdalena and the Cauca river valleys form a well-supported branch in the UPGMA tree (Fig. 4). Moreover, PCoA based on the ISSR dataset isolated SAN (i.e. Chicamocha canyon) and CES and COR (i.e. Caribbean coast) from the rest of the populations sampled (Fig. 1). Taken together, our results uncovered a slight genetic divergence between inter-Andean, Caribbean and Chicamocha canyon *A. excelsum* remnants. Interestingly, for the two chloroplastic loci and the ISSR data, significant genetic differentiation was observed between all region pairs except between the Cauca river valley and the Caribbean coast. We propose that the existence of a past and/or current corridor of dispersal between these two regions may explain our results. Higher genetic differences among regions were detected by AMOVAs based on the two choloroplastic loci than when calculated with the ISSR data. The different evolutionary dynamics (i.e. differences in strength of the genetic drift and migration and mutation rates) of nuclear and chloroplastic loci could explain these contrasting results. Indeed, chloroplast DNA (cpDNA) is uniparentally inherited and haploid and effective population size of chloroplast loci is four times lower than the one of nuclear loci. Difference in type of inheritance could also lead to contrasting migration rates between loci since cpDNA is only dispersed via seeds while nuclear loci are transmitted and dispersed by both pollen and seeds. Moreover, choloroplastic loci are characterized by a low rate of nucleotide substitution (Hamilton 1999), while ISSR data present a high mutation rate (Tuisima et al. 2016). Because of these differences among marker types, chloroplastic loci have been suggested to better reveal

historical changes, especially the population's response to Pleistocene glacial climatic variations, than nuclear loci (Ennos et al. 1999). Contrasting with values of Φ_{ST} between regions reaching up to 0.969 for the choloroplastic loci, our ISSR data showed a high level of admixture between the three genetic groups detected by Bayesian clustering (except in SAN, COR and CES; Fig. 3) and an increasing genetic differentiation among populations with increasing geographic distance (i.e., a pattern of isolation by distance; Fig. S1). We propose that these slight differences in patterns of genetic divergence detected by our two types of molecular markers could be related, in part, to the complex scenario of a combination of historical isolation followed by more substantial gene flow after secondary contact.

Historical separation between inter-Andean and Caribbean SDTF remnants has been previously supported by the existence of floristic differences (Pizano and García 2014) and by a strong genetic structure within species distributed in the Caribbean coast and the Cauca river valley (e.g. the tree *Enterolobium cyclocarpum*; Thomas et al. 2017). The range and distribution of SDTF fragments, and concomitantly of the *A. excelsum* populations, has been molded by the history of the region. Recent studies suggest that genetic isolation between SDTF tree populations predate the late Pleistocene (Caetano et al., 2008; Thomas et al., 2017), with genetic differentiation building up in SDTF remnants isolated during the Neogene (Burnham and Carranco 2004; Côrtes et al., 2015). Secondary contact between differentiated genetic groups could have occurred during the Pleistocene, a period during which the SDTF range is proposed to have been much larger (Pennington 2009). Indeed, the Pleistocene Arc theory suggests that the present-day SDTF patches represent only relics of a much more contiguous formation extending all around the Amazon basin from the northeast of Brazil to Paraguay, including the inter-Andean valleys of Bolivia, Peru, Ecuador and Colombia, up to the Caribbean coast of Colombia and Venezuela (Prado and Gibbs 1993; Pennington et al. 2000). Even if the Pleistocene Arc theory has been questioned (Werneck et al. 2011), SDTFs are still considered to have occupied a much larger area in the past especially in the north of South America, the Andean region from Bolivia to Peru and along a northern Amazonian block including areas of Colombia, Venezuela and the Guiana highlands. In Colombia, the historically larger areas of SDTF were likely connecting our four regions under study (Pennington 2009; Thomas et al. 2017). The maximum extension of the SDTFs is suggested to concur with the last Glacial–Holocene transition (Werneck et al. 2011). Additionally, a recent palynological and geochemical study of a site close to Mount Paramillo has estimated that vegetation change linked to

wetter and cooler climate, and then a possible recess of the SDTFs, in the northern termination of the Colombian Western Cordillera could be as recent as 4,000 yr before present (Munoz-Uribe 2012). In general, patterns of isolation by distance equilibrate relatively quickly and rely mostly on recent demographic events (Barton et al. 2013). However, for species with long generation times such as some woody trees (trees are generally considered long-lived perennials with lengthy juvenile period; e.g. in Anacardiaceae in Hormaza and Wünsch 2007 and Dinesh et al. 2011), it is possible that the fragmentation of SDTFs is, in our case, yet too recent (< 4,000 yr) to detect its genetic footprint and explains why a pattern of isolation by distance can still be retrieved in *A. excelsum*. Human activities during the last century, as for example the usage of tree seedlings or transplantation from different localities (e.g. one tree from the botanical garden of BOL located in the Caribbean coast had chloroplastic haplotypes typically encountered in the southern part of the country; Fig. 4) or forestry activity could also have artificially increased gene flow between some isolated *A. excelsum* remnants and led to a recent admixture between differentiated genetic groups.

For both kinds of markers (ISSRs and chloroplast sequences), the locality of SAN, within the Chicamocha canyon, presented the lowest level of genetic diversity for *A. excelsum* when compared to the other remnants of Colombian SDTFs. Our analyses also provide evidence for possible genetic divergence of SAN from the rest of the Colombian SDTF remnants of *A. excelsum*. During the last decades, this area has been shown to suffer from high and persistent anthropogenic disturbance (Valencia-Duarte et al. 2012), a phenomenon classically related to the existence of recurrent bottlenecks and low genetic diversity in plants (Young et al. 1996). We propose that anthropogenic disturbance could have increased the geographic isolations of relics present in the Chicamocha canyon and that this may have ultimately limited gene flow between the natural population of SAN and the rest of the Colombian SDTF remnants. The fragmentation of SDTFs has also led to genetic isolation and inbreeding depression in other tree species in Brazil, such as *Tabebuia ochracea* (Moreira et al. 2009), *T. roseoalba*, *Handroanthus chrysotrichus*, *H. impetiginosus* and *H. serratifolius* (Collevatti et al. 2014). Inbreeding depression in plants could lead to a decrease in fecundity, in the number of viable seeds as well as lower survival or lower resistance to stress. Identifying populations potentially affected by inbreeding can be essential for the development of long-term conservation strategies (Keller and Waller 2002). An effect of forest fragmentation on pollination success and seed production was documented in *A. excelsum* remnants in Costa Rica, a result possibly related to inbreeding depression in this partially self-

incompatible tree (Ghazoul and McLeish 2001). Unfortunately, the markers used in our study are dominant (i.e. ISSRs) or haploid (i.e. chloroplastic loci) and do not allow to test for possible traces of inbreeding in *A. excelsum* populations. Thus, to facilitate the establishment of strategies that could mitigate risks of local extinction in small isolated remnants of *A. excelsum*, complementary studies focusing on co-dominant molecular markers such as microsatellites or SNPs (Single Nucleotide Polymorphisms) will have to be undertaken.

Our study is the first attempt to provide genetic information at the country level for *A. excelsum*, a native species representing a good candidate for local and national restoration programs of SDTFs. *A. excelsum* is generally propagated by seeds since other methods, such as in vitro propagation or cuttings, have been proven remarkably difficult (Barreto et al. 2007). No specific biological knowledge is required to grow *A. excelsum* from seeds since the percentage of germination without any further treatment lies around 50% or higher (Barreto et al. 2007). Restoration of SDTF is included in the National Strategy for Plant Conservation of Colombia (NSPC), a recent environmental legislation aiming at the rehabilitation, recovery and conservation of disturbed areas in Colombia (Pizano and García 2014; MiniAmbiente 2015). Herein, Regional Autonomous Corporations (hereafter noted RAC) are in charge of the restoration and reforestation programs, and have begun with the implementation of such plans for *A. excelsum*, as for example the RAC of Cundinamarca (CAR) and the RAC of Tolima (CORTOLIMA) in the Magdalena river valley, and the RAC of the Cauca Valley (CVC) and the RAC of Quindió (CRQ) in the Cauca river valley (Morales 2016, KT Bocanegra-González pers. obs.). To meet the goals proposed in the NSPC, *A. excelsum* seedlings from various nurseries and from different bioclimatic life zones (as defined by the Holdridge life zones system; Holdridge 1947) are generally mixed up when reestablishing SDTFs, without considering the potential introduction of new genetic variants and genetic pollution in the newly restored region, or the usage of seeds that are poorly adapted to the new local conditions (MiniAmbiente 2015). Seedlings from mixed origins could generate highly genetically diverse *A. excelsum* populations and potentially increase short-term resilience of the reforested patches (Jump et al. 2009). However, this artificial gene flow can contribute to long-term population decline via outbreeding depression and genetic swamping (see Hufford & Mazer 2003 for a review). Because we do not yet know enough to predict which crossing scenarios have a risk for outbreeding depression in *A. excelsum*, we advise against mixing seeds or plants from regions showing genetic divergence,

or using trees from off-regional sources in transplantation programs using *A*. *excelsum*. Finally, because of the apparent genetic isolation and very low genetic diversity of *A*. *excelsum* in the Chicamocha canyon in the state of Santander, we propose to prioritize this area within future conservation programs.

Acknowledgments

We would like to thank Fernando Fernández Méndez, Omar Aurelio Melo, Luis C. Acosta Cadenas, German Urueña Serrano, Carlos A. Morales Carranza, Jeferson D. Galvis Jiménez, Sonia C. Camargo Roa, Jhon J. Borda Velasquez, Ivan D. Vergara Terreros, Vannesa A. Montoya Sánchez, and Raul Rico Molina for their assistance in the field and for providing samples of *Anacardium excelsum* from Colombia. We would like to thank two anonymous reviewers and TGGE associated editor for their insightful and constructive comments. This research was supported by the program "Talento Humano" N°043-16 from the Universidad del Tolima attributed to K. T. B-G.

Data Archiving Statement

The distinct haplotypes detected were deposited in GENBANK under numbers MG309720-MG309723 and MG309724- MG309726 for *rpl20-rps12* and *trn*L-*trn*F, respectively. This information is also available in the materials and methods' section.

References

Aerts R, Honnay O (2011) Forest restoration, biodiversity and ecosystem functioning. BMC Ecology 11:29

Ang CC, O'Brien MJ, Ng KKS et al (2016) Genetic diversity of two tropical tree species of the Dipterocarpaceae following logging and restoration in Borneo: high genetic diversity in plots with high species diversity. Plant Ecol Divers 9:459-469

Banda-R K, Delgado-Salinas A, Dexter KG et al (2016) Plant diversity patterns in neotropical dry forests and their conservation implications. Science 353:1383-1387

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37-48

Barreto D, Valero N, Muñoz A et al (2007) Efecto de microorganismos rizosféricos sobre germinación y crecimiento temprano de *Anacardium excelsum*. Zonas Áridas 11:240-250

Barton N, Etheridge A, Kelleher J, Véber A (2013) Inference in two dimensions: allele frequencies versus lengths of shared sequence blocks. Theor Popu Biol 87:105-119

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) Genetix 4.02, Logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier

Benito-Garzón M, Ha-Duong M, Frascaria-Lacoste N et al (2013) Habitat restoration and climate change: dealing with climate variability, incomplete data, and management decisions with tree translocations. Restoration Ecol 21:530-536

Bernal R, Gradstein SR, Celis M (Eds) (2015) Catálogo de plantas y líquenes de Colombia. Instituto de Ciencias Naturales, Universidad Nacional de Colombia. Available from: http://catalogoplantasdecolombia.unal.edu.co

Caetano S, Prado D, Pennington RT et al (2008) The history of Seasonally Dry Tropical Forests in eastern South America: inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). Mol Ecol 17:3147-3159

Cárdenas LD, Salinas NR (2007) Libro rojo de plantas de Colombia. Vol 4. Especies maderables amenazadas, primera parte. Instituto Amazónico de Investigaciones Científicas (SINCHI), Bogotá, D.C.

Collevatti RG, Estolano R, Ribeiro ML et al (2014) High genetic diversity and contrasting fine-scale spatial genetic structure in four seasonally dry tropical forest tree species. Plant Syst Evol 300:1671-1681

Condit R, Hubbell SP, Foster RB (1995) Demography and harvest potential of Latin American timber species: data from a large, permanent plot in Panama. J Trop For Sci 7:599-622

Crema S, Cristofolini G, Rossi M et al (2009) High genetic diversity detected in the endemic *Primula apennina* Widmer (Primulaceae) using ISSR fingerprinting. Plant Syst Evol 280:29-36

D'antonio C, Meyerson LA (2002) Exotic plant species as problems and solutions in ecological restoration: a synthesis. Restoration Ecol 10:703-713

Dinesh MR, Hemanth KV, Ravishankar KV et al (2011) *Mangifera*. In Kole C (Ed), Wild crop relatives: genomic and breeding resources. Springer, Berlin, Heidelberg

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19: 11-15

Ennos RA (1999) Using organelle markers to elucidate the history, ecology and evolution of plant populations. In: Hollingsworth PM, Bateman RM, Gornall RJ (eds) Molecular systematics and plant evolution, London and New York, pp 1-19

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

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinf Online 1:47-50

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

Galindo-Rodríguez C, Roa-Fuentes LL (2017) Seed desiccation tolerance and dispersal in tropical dry forests in Colombia: Implications for ecological restoration. Forest Ecol Manag 404:289-293

Garen EJ, Saltonstall K, Ashton MS et al (2011) The tree planting and protecting culture of cattle ranchers and small-scale agriculturalists in rural Panama: opportunities for reforestation and land restoration. Forest Ecol Manag 261:1684-1695

Ghazoul J, McLeish M (2001) Reproductive ecology of tropical forest trees in logged and fragmented habitats in Thailand and Costa Rica. Plant Ecol 153:335-345

Holdridge LR (1947) Determination of world plant formations from simple climatic data. Science 105:367-368

Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Mol Ecol 8:521-523

Hormaza JI, Wünsch A (2007) Pistachio. In: Kole C (Ed), Genome mapping and molecular breeding in plants, fruits and nuts, vol 4. Springer, Berlin, Heidelberg

Hubisz MJ, Falush D, Stephens M et al (2009). Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332

Hufford KM, Mazer SJ (2003) Plant ecotypes: genetic differentiation in the age of ecological restoration. Trends Plant Sci 18:147-155

Hughes AR, Inouye BD, Johnson MT (2008) Ecological consequences of genetic diversity. Ecol Lett 11:609-623

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806

Janes KJ, Miller JM, Dupuis JR et al (2017) The K = 2 conundrum. Mol Ecol 26:3594-3602

Jump AS, Marchant R, Penuelas J (2009) Environmental change and the option value of genetic diversity. Trends Plant Sci 14:51-58

Keller M, Kollmann J, Edwards PJ (2000) Genetic introgression from distant provenances reduces fitness in local weed populations. J Appl Ecol 37:647-659

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends Ecol Evol 17:230-241

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870-1874

Lamb D (1998) Large scale ecological restoration of degraded tropical forest lands: the potential role of timber plantations. Restoration Ecol 6:271-279

Lesica P, Allendorf FW (1999) Ecological genetics and the restoration of plant communities: mix or match? Restoration Ecol 7:42-50

Lozano B, Franco O, Bonilla J (2012) Estimación del crecimiento diamétrico, de *Anacardium excelsum* (Kunth) Skeels, por medio de modelos no lineales, en bosques naturales del departamento del Tolima. Bol Cient Mus Hist Nat 16:19-32

McKay JK, Christian CE, Harrison S et al (2005) "How local is local?" - a review of practical and conceptual issues in the genetics of restoration. Restoration Ecol 13:432-440

Maschinski J, Duquesnel J (2007) Successful reintroductions of the endangered long-lived Sargent's cherry palm, *Pseudophoenix sargentii*, in the Florida Keys. Biol Cons 134:122-129

Meli P, Holl KD, Rey Benayas JM et al (2017) A global review of past land use, climate, and active vs. passive restoration effects on forest recovery. Plos One 12:e0171368

Menges ES (2008) Restoration demography and genetics of plants: when is a translocation successful? Aust J Bot 56:187-196

Menges ES, Smith SA, Weekley CW (2016) Adaptive introductions: How multiple experiments and comparisons to wild populations provide insights into requirements for long-term introduction success of an endangered shrub. Plant Diversity 38:238-246

MiniAmbiente (2015) Plan Nacional de Restauración: restauración ecológica, rehabilitación y recuperación de áreas disturbadas, Ministerio de Ambiente y Desarrollo Sostenible de Colombia, Bogotá D.C (http://www.minambiente.gov.co/index.php/bosques-biodiversidad-y-servicios-ecosistematicos/gestion-en-biodiversidad/restauracion-ecologica)

Morales GA (2016) Plan de manejo y conservación del Caracolí (*Anacardium excelsum*) en la jurisdicción CAR. Bogotá, Colombia.

Moreira PA, Fernandes GW, Collevatti RG (2009) Fragmentation and spatial genetic structure in *Tabebuia ochracea* (Bignoniaceae) a seasonally dry Neotropical tree. Forest Ecol Manag 258:2690-2695

Munoz-Uribe PA (2012) Holocene climate variability in tropical South America: case history from a high-mountain wet zone in NW Colombia based on palynology and X-ray microfluorescence. PhD Dissertation, University of Geneva

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269-5273

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

Pennington RT, Prado DE, Pendry CA (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. J Biogeogr 27:261-273

Pennington RT, Lavin M, Oliveira-Filho A (2009) Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. Annu Rev Ecol Syst 40:437-457

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12:844-855

Pizano C, García H (Eds) (2014) El bosque seco tropical en Colombia. Instituto Alexander von Humbolt, Bogotá D.C

Prado DE, Gibbs PE (1993). Patterns of species distributions in the dry seasonal forests of South America. Ann Miss Bot Gard 80:902-927

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

Rohlf FJ (1992) NTSYS-pc: Numerical taxonomy and multivariate analysis system version 1.80. Exeter Software. Department of Ecology and Evolution, State University of New York, New York.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219-1228

Rozas J, Sánchez-DelBarrio JC, Messeguer X et al (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497

Taberlet P, Gielly L, Pautou G et al (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105-1109

Thomas E, Tobón CG, Gutiérrez JP et al (2017). Genetic diversity of *Enterolobium cyclocarpum* in Colombian seasonally dry tropical forest: implications for conservation and restoration. Biodivers Conserv 26:825-842

Tuisima LL, Hlásná-Čepková P, Lojka B, Weber JC, Alves-Milho SF (2016) Genetic diversity in *Guazuma crinita* from eleven provenances in the Peruvian Amazon revealed by ISSR markers. Bosque 37:63-70

Valencia-Duarte J, Trujillo Ortiz LN, Vargas Ríos O (2012) Dinámica de la vegetación en un enclave semiárido del río Chicamocha, Colombia. Biota Colomb13:40-59

Weeks AR, Sgro CM, Young AG et al (2011) Assessing the benefits and risks of translocations in changing environments: a genetic perspective. Evol Appl 4:709-725

Werneck FP, Costa GC, Colli GR et al (2011). Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidence. Global Ecol Biogeogr 20:272-288

583 Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for
584 plants. Trends Ecol Evol 11:413-418

Table 1. Sampling localities of *Anacardium excelsum* and associated genetic diversity for seven inter-simple sequence repeat (ISSR) primers. For each locality, the abbreviation (code), geographic coordinates, number of samples/number of samples genotyped with the seven ISSRs (Ns/Nsp), number of effective alleles (Ne), percentage of polymorphic loci (%P), Shannon's index (I), expected heterozygosity (He) are displayed.

Region	Locality	Code	Geographic coordinates	Ns/Nsp	Ne	%P	I	He
Magdalena river valley	Armero	ARM	5º00'42.59"N 74º54'39.35"W	15/14	1.50	89.9	0.43	0.28
	Mariquita	MAR	5°14'49.05"N 74°53'20.60"W	15/13	1.53	80.7	0.43	0.29
	Ibagué	IBG	4°25'35.47"N 75°12'46.66"W	16/16	1.52	91.2	0.46	0.30
	Venadillo	VEN	4°40'31.14"N 74°49'20.06"W	15/15	1.62	94.7	0.52	0.35
	Chaparral	CHA	3°45'10.53"N 75°32'26.29"W	6/4	1.45	71.9	0.39	0.26
Cauca river valley	Medellín	MED	6°19'35.09"N 75°30'44.34"W	5/5	1.36	56.1	0.31	0.20
	Cali	CLI	3°26'52.31"N 76°33'30.44"W	10/10	1.50	85.9	0.45	0.30
	Armenia	QUI	4°30'43.23"N 75°39'03.07"W	5/5	1.43	75.4	0.39	0.25
Caribbean coast	Pueblo nuevo	COR	8°28'22.14"N 75°30'36.71"W	3/3	1.42	57.8	0.34	0.23
	Turbaco	BOL	10°33'44.54"N 75°22'30.71"W	3/3	1.39	59.6	0.33	0.22
	Manaure	CES	10°13'56.89"N 73°02'03.08"W	3/3	1.30	40.3	0.24	0.16
	Corozal	SUC	9°17'16.62"N 75°13'44.68"W	4/0	-	-	-	-
Chicamocha canyon	Bucaramanga	SAN	7°07'14.40"N 73°08'39.89"W	6/6	1.30	40.3	0.23	0.16
Mean						69.9	0.38	0.23

Table 2. Genetic diversity of *Anacardium excelsum* for two chloroplastic loci (*trn*L-*trn*F and *rpl20-rps12*). For each locality, the abbreviation (code, as in Table 1), number of sequences (N); number of haplotypes (nH); gene diversity (Hd); nucleotide diversity (π); number of polymorphic sites (S) and a list of haplotypes (Hap) are given. In the haplotype list, the haplotype's name is directly followed by the number of sampled individuals carrying that haplotype, in parentheses.

		*trn*L-*trn*F						*rpl20-rps12*					
Region	Code	N	nH	Hd	π (.10^{-3})	S	Hap.	N	nH	Hd	π (.10^{-3})	S	Hap.
Magdalena river valley	ARM	15	1	0	0	0	H1(15)	14	1	0	0	0	H1(14)
	MAR	14	2	0.14	0.98	5	H1(13), H2(1)	14	2	0.14	0.37	2	H1(13), H2(1)
	IBG	16	1	0	0	0	H1(16)	14	1	0	0	0	H1(14)
	VEN	12	1	0	0	0	H1(12)	14	4	0.40	1.00	3	H1(11), H2(1), H3(1), H4(1)
	CHA	4	1	0	0	0	H1(4)	5	1	0	0	0	H1(5)
Cauca river valley	MED	5	2	0.40	2.74	5	H1(1), H2(4)	5	2	0.40	1.04	2	H1(1), H2(4)
	CLI	7	1	0	0	0	H1(7)	8	1	0	0	0	H1(8)
	QUI	3	1	0	0	0	H1(3)	5	1	0	0	0	H1(5)
Caribbean coast	COR	1	1	_	_	_	H2(1)	0	_	_	_	_	_
	BOL	3	3	1.00	4.57	5	H1(1), H2(1),	3	2	0.67	2.60	3	H1(2), H4(1)

							H3(1)						
	CES	2	1	0	0	0	H1(1), H3(2)	3	3	1.00	2.60	3	H1(1), H3(1), H4(1)
	SUC	1	1	_	_	_	H3(1)	1	1	_	_	_	H3(1)
Chicamocha canyon	SAN	6	1	0	0	0	H2(6)	6	1	0	0	0	H2(6)
	Total *A. excelsum*	89	3	0.33	1.36	5		92	4	0.34	0.86	3	

Table 3. Analysis of molecular variance (AMOVA) for *Anacardium excelsum* distributed across four Colombian regions for: A) seven nuclear inter-simple sequence repeat (ISSR) primers, B) *trn*L-*trn*F chloroplastic sequences and C) *rpl20-rps12* chloroplastic sequences. Regions were defined as in Table 1: Magdalena river valley, Cauca river valley, Caribbean coast, and Chicamocha canyon.

Source of variation	d.f.	Sum of squares	Variance components	% total variance
ISSR				
Among regions	3	62.417	0.633	6%
Within regions	90	959.370	10.660	94%
Φ_{ST}	0.056 ($p = 0.012$)			
*trn*L-*trn*F				
Among regions	3	36.415	0.868	75%
Within regions	83	23.585	0.284	25%
Φ_{ST}	0.753 ($p < 0.0001$)			
rpl20-rps12				
Among regions	3	11.460	0.254	57%
Within regions	85	16.046	0.189	43%
Φ_{ST}	0.573 ($p < 0.0001$)			

Levels of significance are based on 10,000 permutations

Note that the locality of BOL, corresponding to a botanical garden, was not included in the analysis.

Figure legends

Figure 1. Principal co-ordinates analysis (PCoA) of *Anacardium excelsum* populations based on seven nuclear inter-simple sequence repeat (ISSR) primers. Shades of grey correspond to the four sampled regions (Magdalena river valley: white, Cauca river valley: black, Caribbean coast: dark grey and Chicamocha canyon: light grey; see Table 1 for more information on the localities and the abbreviation codes).

Figure 2. Genetic differentiation among populations of *Anacardium excelsum*. A map including all the sampled localities is given on the left panel (A) and shades of grey correspond to the four sampled regions (Magdalena river valley: white, Cauca river valley: black, Caribbean coast: dark grey and Chicamocha canyon: light grey). The codes of each locality correspond to those noted in Table 1. B) Median-joining networks for the *trn*L-*trn*F and *rpl20-rps12* chloroplastic loci. In the networks, the size of each circle represents the occurrence rate of each haplotype and the shades of grey the corresponding sampling regions. Substitutions are indicated by a short line and the gap by a black square. C) Dendrogram computed according to the genetic similarity (Nei, 1978) through the SAHN (Sequential Agglomerative, Hierarchical and Nested Clustering) method using the NTSYS software. Data from seven nuclear inter-simple sequence repeat (ISSR) primers was used for tree reconstruction.

Figure 3. Bayesian assignment of individuals of *Anacardium excelsum* into genetic clusters using the program STRUCTURE based on seven nuclear inter-simple sequence repeat (ISSR) primers. A) Values obtained for the mean posterior probability of the data (Ln Pr(X|K)) and plotted against various values of K tested, suggesting K= 3 as the most likely number of clusters. B) Bayesian assignment of individuals into K = 3 genetic clusters. Each individual is represented by a vertical line, and each color represents a single cluster. The height of each color denotes the probability of an individual being assigned to the respective cluster. Individuals are arranged into populations from which they were sampled. C) Distribution of individuals with distinct level of admixture between genetic groups across sampled localities of *A. excelsum*. The pie charts' color-code is: red = individuals assigned to the red genetic group with cluster membership of 0.60 or

higher, yellow = individuals assigned to the yellow genetic group with cluster membership of 0.60 or higher, blue = individuals assigned to the blue genetic group with cluster membership of 0.60 or higher and white with black stripes = admixed individuals. Circles sizes are proportional to number of individuals. The code of each locality corresponds to those noted in Table 1.

Figure 4. UPGMA tree obtained for the concatenated chloroplastic data set (i.e., joining *trn*L-*trn*F and *rpl20-rps12*) of *Anacardium excelsum*. *Anacardium occidentale* (KY635877) was used as an outgroup. Bootstrap values (1,000 replicates) are given when superior to 75. For each unique concatenated sequence, localities where encountered - directly followed by the number of sampled individuals presenting this sequence between parentheses - are given. Correspondence with haplotypes presented in Table 2 for the *trn*L-*trn*F and *rpl20-rps12* is given on the right for each unique concatenated sequence. Abbreviations of locality names are noted in Table 1.

Supporting Information

Table S1. ISSR's primers used in Colombian *A. excelsum* populations, with their respective sequences, annealing temperatures and number of bands.

Primer	Sequence (5'-3')	Anneling temperature (°C)	Bands	References
UBC-809	AGA GAG AGA GAG AGA GG	48	9	Soares et al 2016; Tuisima et al, 2016
UBC-810	GAG AGA GAG AGA GAG AT	49	6	Soares et al 2016; Tuisima et al, 2016
UBC-811	GAG AGA GAG AGA GAG AT	53	10	Zhou et al, 2014; Duarte et al, 2015; Soares et al 2016
UBC-823	TCT CTC TCT CTC TCT CC	57	6	Hu et al 2014; Soares et al 2016
UBC-825	ACA CAC ACA CAC ACA CT	52.2	9	Hu et al 2014; Zhou et al, 2014; Duarte et al, 2015; Soares et al 2016
UBC-835	AGA GAG AGA GAG AGA GYC	48	9	Zhou et al, 2014; Duarte et al, 2015
UBC-841*	GAG AGA GAG AGA GAG AYC	55	-	Hu et al 2014; Zhou et al, 2014; Duarte et al, 2015; Tuisima et al, 2016
UBC-844*	CTC TCT CTC TCT CTC TRC	46	-	Hu et al 2014; Duarte et al, 2015
Chris*	CACACACACACACAYG	50	-	Guler, 2012
Jhon	AGA GAG AGA GAG AGYC	53	8	Duarte et al, 2015
Mean			8.14	

* Not kept for genetic analyses due to high numbers of missing data, see Results section for details.

Duarte JF, de Carvalho D, de Almeida VF (2015) Genetic conservation of *Ficus bonijesulapensis* RM Castro in a dry forest on limestone outcrops. Biochem Sys Ecol 59:54-62

Guler Y (2012) Population dynamics and sexdetermining mechanisms in the marine amphipod, *Echinogammarus marinus*. PhD Dissertation, University of Portsmouth

Hu ZY, Lin L, Deng JF et al (2014) Genetic diversity and differentiation among populations of *Bretschneidera sinensis* (Bretschneideraceae), a narrowly distributed and endemic species in China, detected by inter-simple sequence repeat (ISSR). Biochem Sys Ecol 56:104-110

Soares ANR, Vitória MF, Nascimento ALS et al (2016) Genetic diversity in natural populations of mangaba in Sergipe, the largest producer State in Brazil. Genet Mol Res 15:1503-8624

Tuisima-Coral LL, Hlásná-Čepková P, Lojka B et al (2016) Genetic diversity in *Guazuma crinita* from eleven provenances in the Peruvian Amazon revealed by ISSR markers. Bosque 37:63-70

Zhou TH, Wu KX, Qian ZQ et al (2014) Genetic diversity of the threatened Chinese endemic plant, *Sinowilsonia henryi* Hemsi.(Hamamelidaceae), revealed by inter-simple sequence repeat (ISSR) markers. Biochem Sys Ecol 56:171-177

Table S2: Matrices based on pairwise Φ_{ST} distance between sampling regions of *Anacardium excelsum* using molecular markers (values below the diagonal and significance, p<0.05, in bold). Levels of significance are based on 10,000 permutations (values above diagonal). Results are given for seven nuclear inter-simple sequence repeat (ISSR) primers, and two chloroplastic loci: *trn*L-*trn*F and *rpl20-rps12*. Regions were defined as in Table 1. Note that the locality of BOL, corresponding to a botanical garden, was not included in the analysis.

ISSR				
Region	Magdalena river valley	Cauca river valley	Caribbean coast	Chicamocha canyon
Magdalena river valley	_	0.006	0.005	0.001
Cauca river valley	**0.025**	_	0.374	0.003
Caribbean coast	**0.072**	0.005	_	0.002
Chicamocha canyon	**0.121**	**0.101**	**0.263**	_
***trn*L-*trn*F**				
Region	Magdalena river valley	Cauca river valley	Caribbean coast	Chicamocha canyon
Magdalena river valley	_	0.005	0.000	0.000
Cauca river valley	**0.355**	_	0.111	0.003
Caribbean coast	**0.898**	0.295	_	0.016
Chicamocha canyon	**0.969**	**0.619**	**0.417**	_
rpl20-rps12				
Region	Magdalena river valley	Cauca river valley	Caribbean coast	Chicamocha canyon
Magdalena river valley	_	0.043	0.005	0.000
Cauca river valley	**0.105**	_	0.069	0.002
Caribbean coast	**0.626**	0.200	_	0.004
Chicamocha canyon	**0.882**	**0.669**	**0.503**	_

653

Figure 1

Click here to download Figure new Figure 1.pdf

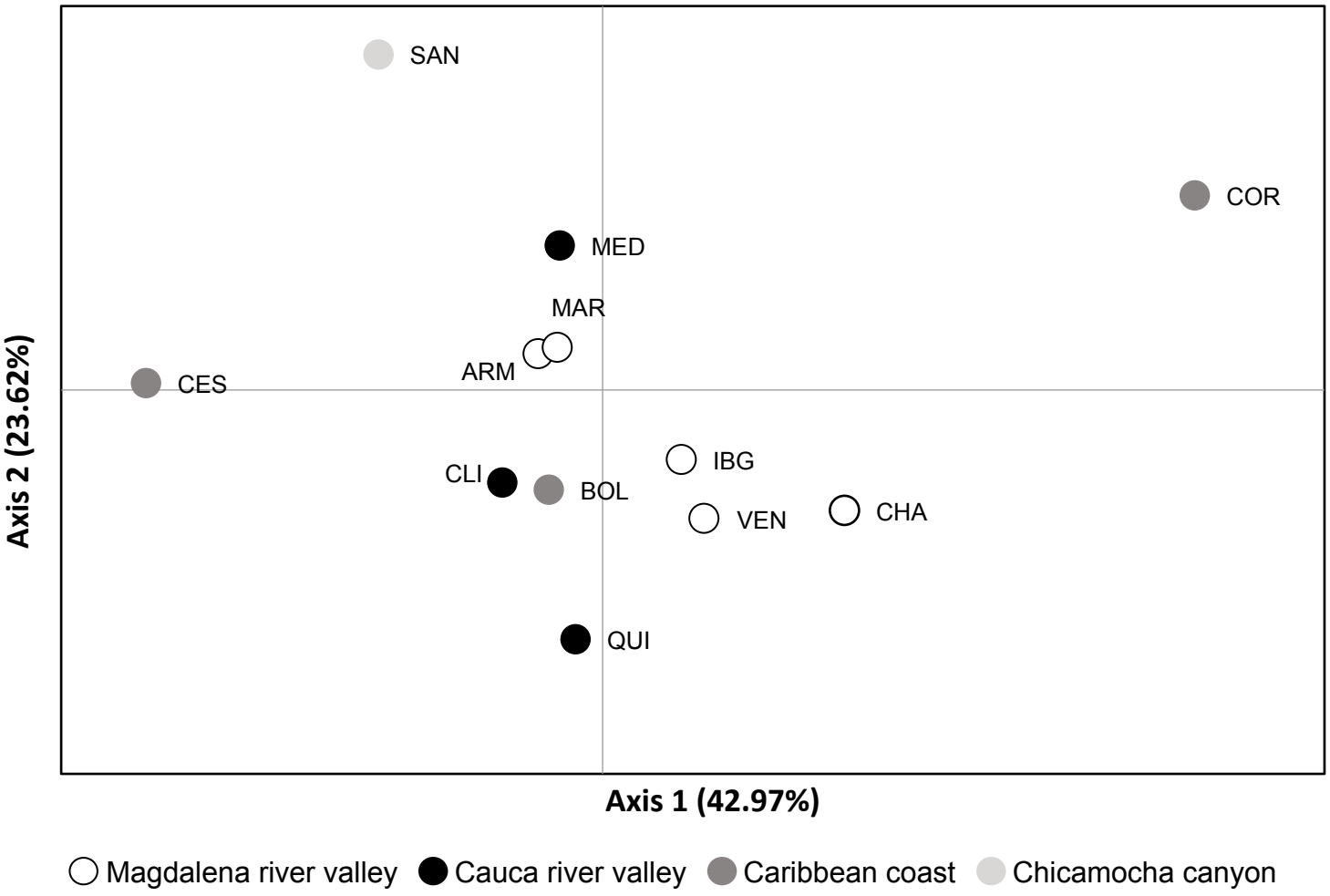

Figure 2

Click here to download Figure new Figure 2.pdf

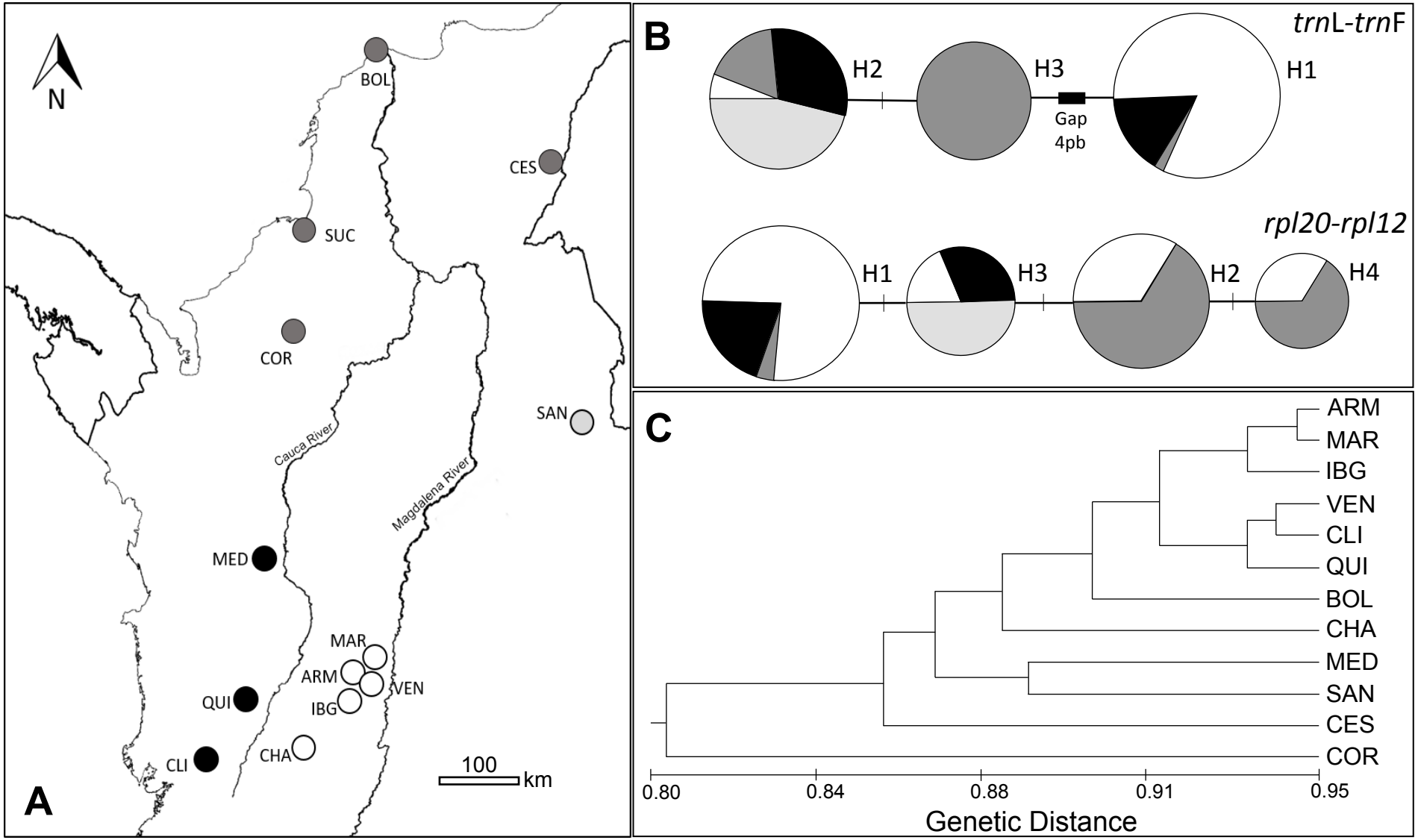

Figure 3

Click here to download Figure new Figure 3.pdf

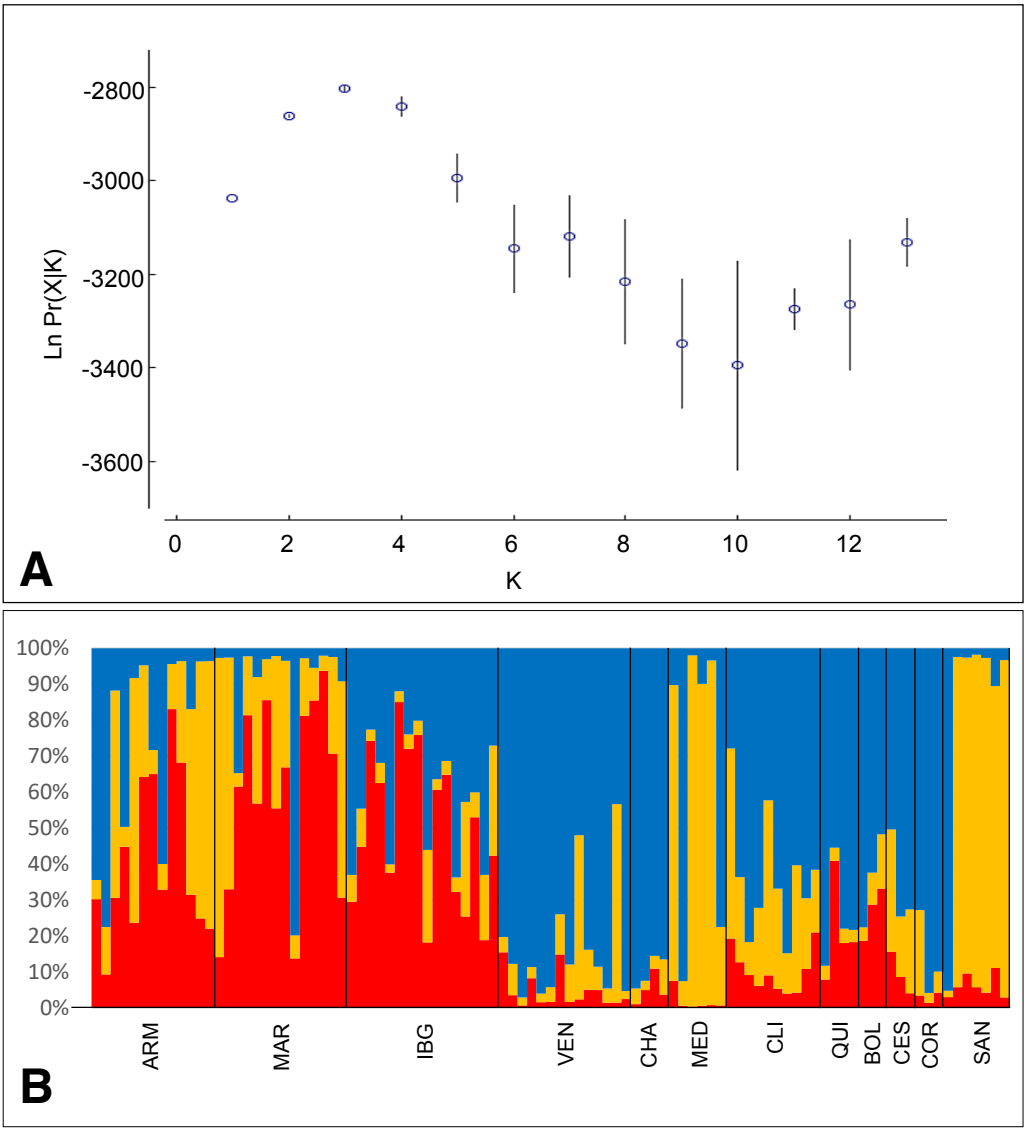

Figure 4

Click here to download Figure new Figure 4.pdf