REVISTA DE
Biología Tropical

DNA barcoding for molecular identification of *Gynerium sagittatum* (Poales: Poaceae): genetic diversity in savannah genotypes from Córdoba, Colombia

Hernando J. Rivera-Jiménez[1]*, Bruno C. Rossini[2], Alicia Del Carmen Humanez Alvarez[3], Saura R. Silva[4], Juan Yepes Escobar[1] & Celso L. Marino[5, 2]

1. Universidad de Córdoba. Departamento de Biología. Grupo de Investigación BIODIMARC. Código Postal: 230002. Montería, Colombia; hriveraj@gmail.com
2. Universidade Estadual Paulista, Instituto de Biotecnologia (IBTEC), UNESP, Botucatu, São Paulo, Brasil; bruno.rossini@unesp.br
3. Universidad del Sinú "Elías Bechara Zainum", Facultad de Salud y de Ciencias e Ingenierías, Montería, Córdoba, Colombia; ahumanez5@yahoo.com
4. Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, Departamento de Biologia Aplicada à Agropecuária, São Paulo, Brasil; saura.silva@gmail.com
5. Universidade Estadual Paulista, Departamento de Genética, Instituto de Biociências, Botucatu, São Paulo, Brasil; clmarino@ibb.unesp.br
    *Correspondance

ABSTRACT. Introduction: The fiber of the *Gynerium sagittatum* Aubl. P. Beauv is raw material for the elaboration of several handcrafts, which are symbols of Colombian cultural identity. In the manufacture process, different genotypes are used according to the fiber quality and the type of craftsmanship, but it is believed that *Gynerium* is a complex species, and to date, there is no agreement on whether these genotypes belong to the same species or to different species. **Objective:** The aim of this study was to quickly and accurately identify wild cane plants using the nuclear ribosomal internal transcribed spacer (ITS1+ITS2), three chloroplast regions (*mat*K*, rbc*L*, ycf*1), and their combinations. **Methods:** Different tests were used for discrimination: (1) inter and intraspecific distances, (2) Best Match (BM), Best Close Match (BCM), and tree-based method (3) Neighbor Joining (NJ) and (4) maximum likelihood and bayesian inference in molecular data. **Results:** The results showed that BM and BCM approaches revealed the low rate of correct species identification for ITS+matK (33.3 %) and ITS (28.6 %) loci, showing similarity among sequences. These results were further supported by tree-based analyses, where all individual regions and the different gene combinations had a zero discrimination rate. **Conclusions:** all genotypes belong to the same species of wild cane, therefore existing morphological differences can be related to phenotypic plasticity.

Key words: *Gynerium saccharoides*; wild cane; DNA barcode; nuclear ribosomal internal transcribed spacer; chloroplast genes.

The wild cane plant fiber (*Gynerium sagittatum* Aubl. P. Beauv.) is the essential raw material for the elaboration of several handcrafts, where the "vueltiao" hat stands out, being a symbol of Colombian cultural identity. Unfortunately, this genetic resource is disappearing over time leading to scarcity and increasing the costs of the raw material

for the artisanal sector. This compromises in the near future the economy, labor, social stability, life quality, tradition, folklore, and the identity of one of the most unprotected social sectors in the country, the indigenous people (Zorro & Prieto, 1999). In the elaboration of handcrafts, artisans show that there are differences among the three genotypes called "Criolla", "Martinera" and "Costera". The "Martinera" and "Costera" genotypes make the manufacturing process hard, diminishing the efficiency in the artisan's work; whereas the "Criolla" genotypes are easy to manipulate, therefore they are used in the elaboration of finer and of greater cost crafts. The biodiversity of wild cane (Fig. 1) in Colombia is threatened by the risk of losing its genetic resource due to habitat destruction with anthropogenic activities such as cattle and agriculture. Anthropogenic activities lead to fragmentation and consequently to the reduction of the genetic diversity of the species (Aramendiz-Tatis, Espitia-Camacho, & Cardona-Ayala, 2009). Currently, there are scarce studies based on molecular characterizations for the different wild cane specimens. The lack of knowledge of the species diversity with increased genetic vulnerability to pests, diseases and other adverse factors due to genetic uniformity, poses a high risk of losing this plant genetic resource (Zorro & Prieto, 1999).

Characterization studies that described the morphological variation were performed by Aramendiz-Tatis et al. (2009), detecting five phenotypic classes. Through genetic diversity studies using the AFLP technique among populations from different regions of Colombia, multiple correspondence analysis distinguished three groups. Traits with desirable agronomic attributes were identified for handcrafts, however a low correlation was observed between the geographic distance and the genetic differentiation level of the species (Rivera-Jiménez, Suárez-Padrón, & Palacio-Mejía, 2009). Therefore, at present, there is doubt whether the genotypes used in the elaboration of these handcrafts belong to the same species or if there have been changes at the genetic level

that can separate these genotypes in different *Gynerium* species.

One of the most promising techniques for identifying biological units is the use of DNA standard regions, called DNA barcoding for the species molecular identification (Hebert, Cywinska, Ball, & deWaard, 2003). DNA barcoding can be used as a powerful tool in species that are difficult to identify based on morphological traits. Barcoding can also be used as a taxonomic support tool in delimiting and describing complex species (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). In addition, this technique helps to identify species with phenotypic plasticity, where there is the possibility of classifying them erroneously (Heinrichs et al., 2011). Therefore, DNA barcodes have the potential to become an important support for the assessment of biodiversity conservation, and they can help to increase flora descriptions (Hartvig, Czako, Kjaer, Nielsen, & Theilade, 2015). Significant advances have been made in the implementation of DNA barcoding for molecular

Fig. 1. Wild cane plant in savannahs of Cordoba, Colombia.

identification in plants, as the nuclear internal transcribed spacer (ITS1 and ITS2) and other regions of the plastid genome such as *psb*K-*psb*I, *trn*H- *psb*A, *atp*F-*atp*H, *mat*K, *rbc*L, *rpo*C1, *rpo*B, *rbc*L, *rpo*B, and *ycf*1, but it has been shown that combining some regions has more discriminative power than using a single region, and currently the most accepted combination are the standardized regions *mat*K and *rbc*L (CBOL Plant Working Group, 2009; Xu et al., 2015; Dong et al., 2015). Nuclear (ITS) and chloroplast (*rbc*L, *mat*K) regions were used to identify specimens of grasses (Poaceae tribe *Poeae*), the ITS marker being the region that achieved the highest success rate for the identification of specimens (Birch, Walsh, Cantrill, Holmes, & Murphy, 2017). In previous studies, the ITS2 nuclear region was identified as the best option for the identification of Poaceae medicinal species (Tahir, Hussain, Ahmed, Ghorbani, & Jamil, 2018). We here evaluated the performance of the regions (ITS, ITS2, *mat*K, *rbc*La, and *ycf*1) as standard barcodes for the identification of wild cane plants in the savannahs of Córdoba, Colombia.

MATERIALS AND METHODS

Plant material, DNA extraction, PCR amplification and sequencing: A total of 35 foliar tissue samples were included in the study. Seven specimens of each were analyzed, from "Criolla 0", "Criolla 1", "Criolla 2", "Martinera", and "Costera", genotypes that belong to the *Gynerium sagittatum* species. The samples of "Criolla 0", "Criolla 1" and "Criolla

2" were collected in the Cuatro Vientos, San Andres de Sotavento, located at 50 m.a.s.l. The samples of "Martinera" and "Costera" were collected in the Los Vidales, Tuchín, located at 70 m.a.s.l. Collections were made in the Zenues indigenous reservation of the Córdoba department, Colombia (Table 1).

Young leaves (without midrib) were detached from each plant, and a 1 cm^2 piece of leaf tissue was desiccated in an airtight plastic bag containing silica gel. Genomic DNA was isolated using a modified CTAB 2 % protocol (Doyle & Doyle, 1990). PCR products for two nuclear ribosomal internal transcribed spacer (ITS1 and ITS2) and three plastid barcodes (the coding genes *mat*K, *rbc*L and *ycf*1) were amplified and sequenced using universal primers (Kress et al., 2009; Soltis & Soltis, 2009; Chen et al., 2010; Dong et al., 2015; Table 2). The PCR reaction (10 μL) contained approximately 50 ng (1 μL) of template DNA according to the protocols of (CBOL Plant Working Group, 2009; Rivera-Jiménez et al., 2017). The sequencing reactions were performed in both directions according to the specifications of the BigDye Terminator Cycle Sequencing Kit from Applied Biosystems V 3.1 (Applied Biosystems). All sequences have been deposited in GenBank under accession numbers KY549399-KY549417 for the *mat*K, KY549418-KY549441 for the *rbc*L and KY522854-KY522874 for the ITS regions. We also included public sequences of the *mat*K gene (GenBank: HE586080.1) and of the *rbc*L gene (GenBank: U31105.1) from *Gynerium sagittatum* species.

TABLE 1

Genotypes and collection site analyzed in this study

Sample ID	Genotype	Genus	Total no. of plants	Collection site
C-0	Criolla 0	*Gynerium*	7	Cuatro vientos[1]
C-1	Criolla 1	*Gynerium*	7	Cuatro vientos[1]
C-2	Criolla 2	*Gynerium*	7	Cuatro vientos[1]
M	Martinera	*Gynerium*	7	Los Vidales[2]
Ct	Costera	*Gynerium*	7	Los Vidales[2]

[1] San Andrés de Sotavento- Latitude: 9º15´2"N - Longitude: 75º32´6"W: 50m. Department of Córdoba. Colombia.
[2] Tuchín- Latitude: 9º14´22"N - Longitude 75º32´9"W: 70m. Department of Córdoba. Colombia.

Rev. Biol. Trop. (Int. J. Trop. Biol.) • Vol. 68(4): 1049-1061, December 2020

1051

TABLE 2

A list of primers used for PCR and sequence in this study

Region	Primer	Sequence 5′-3′	Tm (ºC)	Reference
ITS1	5a fwd	CCTTATCATTTAGAGGAAGGAG	50 ºC	(Chen et al., 2010)
	4 ver	TCCTCCGCTTATTGATATGC		
ITS2	S2F	ATGCGATACTTGGTGTGAAT	56 ºC	(Chen et al., 2010)
	S3R	GACGCTTCTCCAGACTACAAT		
*mat*K	1RKIM-f	ACCCAGTCCATCTGGAAATCTTGGTTC	52.2 ºC	International Barcode of Life (iBOL)
	3FKIM-r	CGTACAGTACTTTTGTGTTTACGAG		
*rbc*L	rbcLa-F	ATGTCACCACAAACAGAGACTAAAGC	62 ºC	(Kress et al., 2009)
	rbcLa-R	GTAAAATCAAGTCCACCRCG		
*ycf*1	ycf1bF	TCTCGACGAAAATCAGATTGTTGTGAAT	57 ºC	(Dong et al., 2015)
	ycf1bR	ATACATGTCAAAGTGATGGAAAA		

Data analysis: DNA barcodes candidates were edited using the BioEdit software, version 7.0.9.0 (Hall, 1999). Also manual adjustments were made. All sequences were deposited in the GenBank (Appendix 1). Informative polymorphic characters were identified by MEGA6 (Tamura et al., 2013). Alignment of the sequences was performed using the MUSCLE alignment tool (Edgar, 2004). The different locus combinations were taken into account for the evaluation of the model independently in each marker. The barcode analysis was calculated using the Kimura 2-Parameter (K2P) model (Kimura, 1980). The effectiveness of the regions (ITS, ITS2, *mat*K, *rbc*La, *ycf*1) and their combinations as barcodes was evaluated using three different methods.

Genetic Distance-Based Method: The program TaxonDNA (Meier, Shiyang, Vaidya, & Ng, 2006) was used to test the accuracy of the species assignments, the cluster analysis, and the distribution of interspecific and intraspecific distances in the dataset. The best match (BM) and the best close match (BCM) were taken into account, as well as the formation of groups determined by similar sequences for each region evaluated, by thresholds of 1 % and 5 %.

Tree-Based Method: A tree-based method was used to evaluate the species resolution degree (identification). Each barcode region

and possible combinations of the regions were evaluated by the species resolution degree they provided. The analyzes were performed according to the consensus parameters of Neighbor-Joining (NJ), Kimura 2-Parameter (K2P) and cluster analysis by sequence divergence between genotypes, using MEGA6 (Tamura et al., 2013). Node support was assessed using the bootstrap resampling (1 000 replicates) (Felsenstein, 1985). The species were determined by analyzing the lengths of the branches in pairs and if two species diverge, they must be separated by a branch length greater than zero and a bootstrap greater than 50 %, under these criteria they are considered separate species.

Phylogenetic analysis: The phylogenetic analysis used concatenated and individual genes datasets. The best-fitting substitution model was calculated using jModeltest (Darriba, Taboada, Doallo, & Posada, 2012) according to the AIC criterion (Akaike, 1973). The selection of best-fitting model for each region were: for *rbcl* (TIM2), *mat*K (HKY), ITS (HKY-G) and for combined dataset (GTR+G+I). The matrixes were analyzed using Bayesian Inference (BI) in MrBayes V.3.2 (Ronquist et al., 2012) and Maximum Likelihood (ML) with RAxML version 8 on XSEDE (Stamatakis, 2014). For the BI, 500 000 000 generations were iterated, and sampled every 1 000 generations, for two runs each with four chains. The first 25 % trees were discarded as burn-in.

For the ML analyses, the GTRGAMMA model was performed and bootstrap values were obtained using rapid bootstrapping with 1 000 replicates. Trees were edited using TreeGraph2 beta version 2.0.52-347 (Stöver & Müller, 2010). Same loci from the species *Zea mays*, *Piptatherum miliaceum,* and *Pennisetum purpureum* were used as outgroup. Phylogenetic analyses, as well as "best of fit" modeling test, were performed at CIPRES Science Gateway (www.phylo.org).

RESULTS

PCR amplification and sequencing: The sequence information of five candidate DNA barcode markers, ITS*, mat*K*, rbc*L, and *ycf*1, are provided in Table 3. Sequencing success rates were 92.6 % (ITS), 100 % (*mat*K), and 100 % (*rbc*L). The complete ITS region (*ITS1-ITS2*) was used as a single barcode locus. Unfortunately for *ycf*1, the universal primer proposed by Dong et al. (2015) did not returned a great amplification success. The present study submitted 64 new sequences to NCBI, which included 21 sequences of nrITS1+ nrITS2; 19 sequences of *mat*K, and 24 sequences of *rbc*L

(Appendix 1). Using BLAST analysis, all the loci correctly identified a 100 % of the samples at species level (*G. sagittatum*); while ITS1 had an identification rate of 96 % at the family level (*Poaceae*). The absence of species-level identification using the region barcode nrITS1 is due to the lack of sequence records in the NCBI database.

Intra- and interspecific diversity: The aligned sequence lengths were amplified between 865 bp for ITS to 584 bp for *rbc*L, the ITS region showed the most variable and informative sites of parsimony, followed by *mat*K (Table 4). The intraspecific distances in pairs in the seven bar codes varied from a minimum of 0 to a maximum of 17.9 %. The mean intraspecific distances were minimum for *rbc*L (0.01 %) and maximum for ITS (5.8 %). Interspecific distances in pairs varied from a minimum of 0 % to a maximum of 18.3 % (Table 4). The mean interspecific distances were minimum for *rbc*L (0.01 %) and maximum for ITS (5.9 %), therefore, ITS showed the highest intra and interspecific average distance. The combination of sequences from different barcode regions increased intraspecific and interspecific mean distances. The data showed overlap

TABLE 3
Characteristics of the eight wild cane barcodes evaluated in this study, including global percentage PCR for each genotype

Parameter	C-0	C-1	C-2	M	Ct	ITS	*mat*K	*rbc*L	*ycf*1	ITS + *mat*K	ITS + *rbc*L	matK + *rbc*L	ITS + matK + *rbc*L
Universality of the primers	-	-	-	-	-	Yes	Yes	Yes	No	-	-	-	-
Percentage PCR success (%)	100	100	100	85	90	91.8	88.5	95.1	0	-	-	-	-
Percentage sequencing success (%)	98	93	100	100	95	92.6	100	100	-	-	-	-	-
Length of aligned sequence (bp)	-	-	-	-	-	865	790	584	-	1 655	1 449	1 374	2 239
No. of parsimony informative sites/ variable sites	-	-	-	-	-	85/248	4/17.0	0/1	-	89/265	85/249	4/17.0	89/266
No. of species samples (individuals)	-	-	-	-	-	20	19	24	-	21	21	19	21
Percentage ability to discriminate (NJ)	-	-	-	-	-	0	0	0	0	0	0	0	0

Genotype: C-0= Criolla 0; C-1= Criolla 1; C-2= Criolla 2; M= Martinera; Ct=Costera.

Barco de locus	Intraspecific distances (%)			Interspecific distances (%)			Intra-/interspecific distance overlap with 5 % error margin on both sides	
(no. of sequences)	Min	Max	Mean	Min	Max	Mean	Overlapping distance range	Intra-/interspecific sequences in the overlap
ITS (21)	0.7	17.9	5.8	0.53	18.3	5.89	1.1-17.9	94.4
*mat*K (20)	0	1.4	0.4	0	1.3	0.24	0-1.40	100
*rbc*L (25)	0	0.2	0.01	0	0.2	0.01	0-0.17	99.3
ITS+*mat*K (21)	0.5	12.9	3.6	0.4	18.3	4.09	0.82-12.94	91.3
ITS+*rbc*L (21)	0.4	17.9	3.7	0.3	14.5	3.21	0.52-17.93	95.5
*mat*K+*rbc*L (20)	0	1.4	0.3	0	1.3	0.19	0-1.44	100
ITS+*mat*K+*rbc*L (21)	0.4	6.6	2.3	0.2	12.8	2.5	0.48-6.60	91.8

between intraspecific and interspecific distances of the individual or concatenated sequences. The minimum overlap percentage was 91.4 % in combination sequences using two genes (ITS+*mat*K) and the maximum 100 % using the matK region and the combination sequences of two regions (*mat*K+*rpo*C1) (Table 4).

Species discrimination: The identification of species through the use of BM or BCM was deficient for the three loci and their combinations, because in all cases the identification success was < 40 % (Table 5). The analysis based on TaxonDNA software showed that the concatenated region ITS+*mat*K had the highest rate for the correct identification of species (BM: 33.3 %; BCM: 33.3 %;) followed by ITS, ITS+*mat*K+*rbc*L, ITS+*rbc*L (Table 5) and *mat*K, *rbc*L, *mat*K+*rbc*L had the lowest discrimination rate (BM: 0 %; and BCM: 0 %). To evaluate the efficiency of genes to produce specific groups of species, we use the "group" function of TaxonDNA at two different thresholds, 1 % and 0.5 %. With a threshold of 1 %, ITS worked best by producing 15 groups, and 21 of those groups included only one species (Table 6). With a threshold of 0.5 %, the ITS region also produced the maximum number of groups (21), with only one species of equal value (Table 6).

Tree based analyses: NJ trees were constructed for each individual gene and the

different gene combinations based on K2P. One of this research objectives was to test whether DNA regions barcodes could discriminate among wild cane species. The ITS region (Fig. 2) that showed a higher polymorphism at sequence level did not show defined clusters that would allow to separate the genotypes according to the phenotypic and/or genetic characteristics reported by Aramendiz-Tatis et al. (2009) and Rivera-Jiménez et al. (2009) in previous studies. In the same way, it obtained the same results in the other trees generated by other locus and regions combinations. None of the sequences showed intraspecific variations (data not shown), these genotypes shared the same cluster for each individual gene and the different gene combinations, and some of the species positions were within the other species clades.

The evaluation of barcoding sequences based on phylogenetic trees was established by the usage of individual regions and their combinations (Appendix 2, Appendix 3, Appendix 4). The most informative tree (representing the most well-resolved tree) was the BI tree using the ITS region (Fig. 3). However, the analyses showed that there was no formation of defined groups when using morphology and genotypes previous classification through Bayesian Inference analysis (Fig. 3), thus showing a similar behavior compared to the other methods. The non-existence of well-differentiated groups in the analyzed taxa, suggests that all the studied

TABLE 5

Identification based on the "best match" and "best close-match" functions of the TaxonDNA program

Barcode locus	Best match		Incorrect %	Best close match		Incorrect %
	Correct %	Ambiguous %		Correct %	Ambiguous %	
ITS (21)	28.6	0	71.4	28.6	0	71.4
*mat*K (20)	0	90	10	0	90	10
*rbc*L (25)	0	96	4	0	96	4
ITS+*mat*K (21)	33.3	0	66.6	33.3	0	66.6
ITS+*rbc*L (21)	23.8	0	76.2	23.8	0	76.2
*mat*K+*rbc*L (20)	0	85	15	0	85	15
ITS+*mat*K+*rbc*L (21)	28.6	0	71.4	28.6	0	71.4

TABLE 6

Cluster analysis of wild cane based on the plant barcodes

Barcode locus	At 1 % threshold			Clusters with only one species	At 0.5 % threshold			Clusters with only one species
	No. of clusters	% of clusters with threshold violation	Largest pairwise distance (%)		No. of clusters	% of clusters with threshold violation	Largest pairwise distance (%)	
ITS (21)	15	0	0	21	21	0	0	21
*mat*K (20)	2	50	1.3	1	3	0	0.4	2
*rbc*L (25)	1	0	0.2	0	1	0	0.2	0
ITS+*mat*K (21)	11	9.1	8.8	10	18	5.5	0.6	16
ITS+*rbc*L (21)	10	10	7.1	9	15	6.7	1.1	14
*mat*K+*rbc*L (20)	2	50	1.2	1	3	0	0.2	2
ITS+*mat*K+*rbc*L (21)	5	20	7.1	4	14	7.1	4.8	13

genotypes belong to the same species and that some morphological variations can be the result of environmental factors, which suggests the existence of phenotypic plasticity in this species.

DISCUSSION

In the present study, four plant loci (ITS, *mat*K, *rbc*L and *ycf*1) were evaluated as DNA barcoding for the differentiation of possible wild cane species. These regions had already been tested as DNA barcoding regions in terrestrial plants (Hollingsworth, 2014). The *ycf*1 locus used in this research was proposed by Dong et al. (2015) for its potential use as DNA barcoding in different plant groups. The efficiency of the DNA barcoding in this study was justified based on its potential use to differentiate species between angiosperm plants and

systematic studies in Poaceae (Barker, Linder, & Harley, 1995; Grass Phylogeny Working Group II, 2011; Hollingsworth, 2014). Some authors such as Neubig and Abbott (2010) have shown low PCR success of this region for plants, like in this study, especially in the Lauraceae and Annonaceae families. On the other hand, through the use of the ITS region as a DNA barcode, genetic diversity studies have been carried out in plants of the same family of Poaceae, as in sugarcane (*Saccharum*), separating accessions of *S. spontaneum* from *S. officinarum, S. barberi, S. sinense, S. robustum* (Yang et al., 2016). Many efforts have been made to discover DNA barcoding regions that are more variable and capable to identify taxa of terrestrial plants.

According to our results, ITS has more parsimonious informative sites and better discriminatory power among the proposed loci,
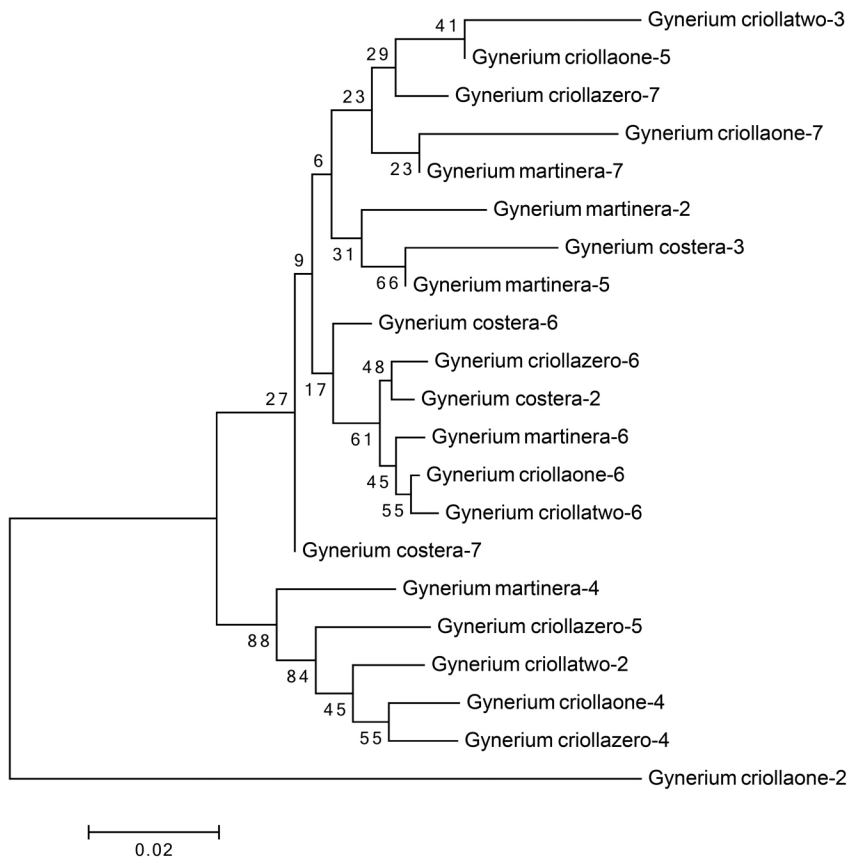
Rev. Biol. Trop. (Int. J. Trop. Biol.) • Vol. 68(4): 1049-1061, December 2020

1055

Fig. 2. Neighbor-joining K2P tree based on analysis of the ITS region. The numbers above the nodes correspond to bootstrap values > 50 %.

i.e., *mat*K and *rbc*L, which is consistent with the results of many previous studies (Ashfaq, Asif, Anjum, & Zafar, 2013; Hartvig et al., 2015; Xu et al., 2015; Tahir et al., 2018). The analysis of intra and interspecific distances showed that ITS had the highest sequence divergence (Table 4). However, according to the Neighbor-joining (NJ) tree, the ITS region and the different gene combinations had a zero-discrimination rate on the wild cane specimens (0 %) (Table 4). This ITS nuclear region has been used for a long time to study the phylogeny, taxonomy, and the species identification in plants (Barker et al., 1995; Grass Phylogeny Working Group II, 2011; Ashfaq et al., 2013; Hollingsworth, 2014). Studies carried out by Tahir et al. (2018) conclude that the ITS2

region showed the highest percentages of intra- and interspecific divergences, followed by the *mat*K and *rbc*L regions for the identification of medicinal species of Poaceae.

Several combinations of two or three locus have been proposed as barcodes, but a consensus on the usefulness of these barcodes has not been achieved (Xu et al., 2015). The analyses conducted in our research indicated that these sequences have high values in the intraspecific average distances of some genotypes and the interspecific distance between them, overlapping their intra- and interspecific distances without differences of DNA barcoding, thus reducing the identification rate of species. The combination of *mat*K+*rbc*L is proposed by CBOL Plant Working Group
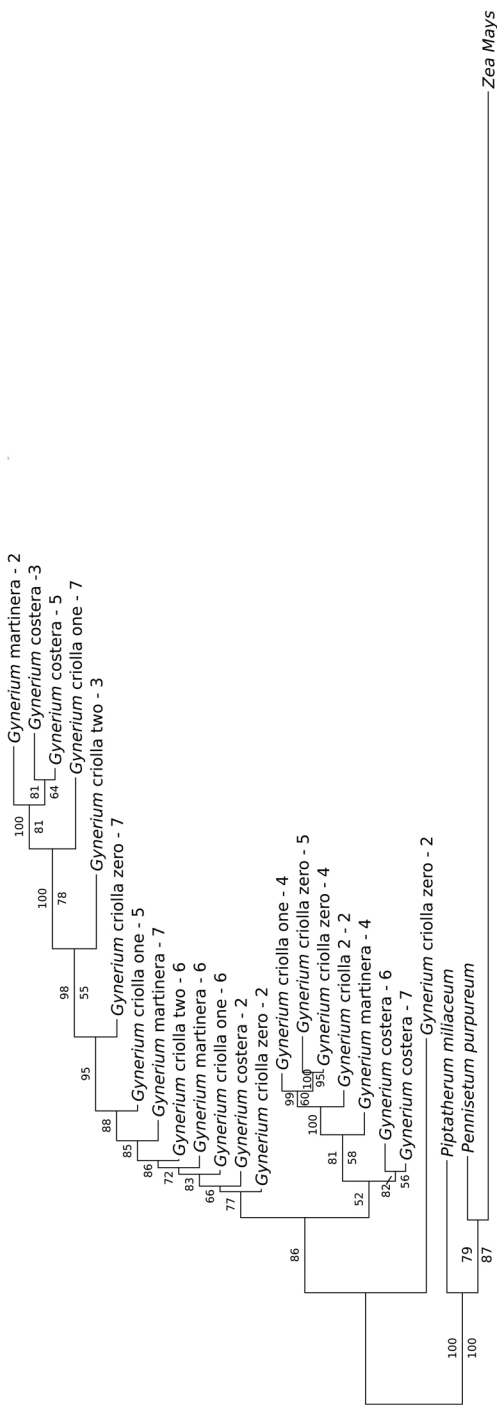
Fig. 3. ITS gene tree for *Gynerium sagittatum* from Bayesian Inference approach. Label values above the branches correspond to posterior probability values, and below them to maximum likelihood bootstrap.

(2009) as a universal DNA barcoding for all terrestrial plants, however, in this research, it had the lowest discrimination resolution (0 %) (Table 5) among the four evaluated combinations, due to the low variability of these coding genes. In contrast, the combination of ITS+*mat*K had the highest percentage (33.3 %) species identification compared to the other DNA barcoding candidate regions or combinations (Table 5). According to the distances results, "BM," "BCM," and the analysis of Neighbor-joining (NJ) trees, we can predict that all the evaluated specimens can be the same species. However, a strategy to identify the species of this genus would be having a better understanding of the specimens geographic information, an approach that has been used in DNA barcoding (Parveen, Singh, Raghuvanshi, Pradhan, & Babbar, 2012). According to our results, the highest identification criteria were for 'BM' and 'BCM' compared to the other two evaluation criteria, showing different values for the loci of the evaluated barcode regions. Surprisingly, a nuclear and plastid gene sequences combination reduced the identification success according to the TaxonDNA program "all species barcodes".

We noticed that in the majority of the wild cane specimens, not only the intraspecific and interspecific distances were very large, but there was also a visible distance overlap in the barcode sequences that ranged from 91.4 to 100 % for the different loci or their combinations. In addition, the reduced identification success in combined sequences may be explained due to the increased level of sequence overlaps or by incongruence between the plastid and nuclear genes (Ashfaq et al., 2013). In a previous study on DNA barcoding of *Dendrobium* species, Xu et al. (2015) have reported a successful identification of the taxon through TaxonDNA, showing greater success in multilocus regions, based on different program criteria. In another study on the discrimination of cotton species, Ashfaq et al. (2013) have reported low percentages of taxon identification, based on BM and BCM criteria, both in single regions and multilocus regions. Copaci, Pocol, Căprar, and Sicora,

(2015) tested DNA barcoding to differentiate species of *Calluna vulgaris*, they found a lack of intraspecific variability for the *mat*K and *rpo*C1 markers. In another study, Selvaraj et al. (2012) used ITS region to identify *B. diffusa* from the other three species, despite the fact that they share many morphological similarities. Previously, Singh, Parveen, Raghuvanshi, and Babbar, (2012) in a study on *Dendrobium* species determined that the ITS region provided the highest resolution, which allowed to identify species. However, Awad, Fahmy, Mosa, Helmy, and El-Feky (2017) could not discriminate Triticum species through the use of chloroplast genes and their combinations.

We performed cluster analysis to evaluate the barcodes efficiency to separate the species. As a single locus, ITS was the region that discriminated the best, producing 15 clusters and 21 clusters with a single species at a threshold level of 1 %. The loci combination reduced the number of cluster and cluster single-species. None of the regions were tested to produce clusters with single species profiles. At the same threshold, the combination of all three loci (ITS+*mat*K+*rbc*L) produced 5 clusters, and 4 of those clusters included single species. At a 0.5 % threshold, the number of clusters in ITS were equal to the 1 % threshold, the combination of two loci (ITS+*mat*K) increased the number of clusters from 11 to 18, and cluster single-species also increased (from 10 cluster at 1 % to 16 cluster at 5 %).

The "cluster" analysis and the "clusters included single species" function of the Taxon-DNA program allow the efficiency of the DNA barcoding to separate species. This grouping function showed that ITS region was the DNA barcode that gave the best result, producing a larger number of groups. Although this analysis helps us to understand the resolution power taking into account the low viability of the threshold values, the NJ tree analysis shows that the resolution degree to separate the species was of 0.0 %. Interestingly, the five genotypes ("Criolla 0", "Criolla 1", "Criolla 2", "Martinera" and "Costera") do not form a specific cluster.

Morphological characterizations performed in these genotypes were reported by Aramendiz-Tatis et al., (2009) identifying genotype groups that showed attributes such as a soft texture fiber, a scanty pubescence pod, a thick stem wall, and a thin stem diameter, this group was colloquially called "criolla". Second and third groups were characterized by having rough texture fibers, pods with abundant pubescence, thin stem walls, and slightly thick stem diameters, and were denominated "Martinera" and "Costera" respectively, however, the author reports little genetic variability in the accessions studied. Later (Rivera-Jiménez et al., 2009) performed multiple correspondence analysis, using AFLP type markers in the same species, showing a low correlation between the geographical distance and the level of genetic differentiation. According to Kalliola, Puhakka and Salo, (1992) wild cane is characterized by being invasive, rustic, and fast-growing, although it is native from the west of India, it has a wide range of distribution that goes from Mexico, Central America, and all South America to Paraguay, showing that the species is a complex of variants in its morphology and ecology; identifying two types of plants, some with a large stem and others with a short stem in the western Amazon. Several researchers propose that invasive species are highly plastic (Hulme, 2008); consequently, it has been pointed out that the ecological amplitude is correlated to the plasticity in some species (Sultan, 2001). The phenotypic plasticity concept in invasive species has been discussed since the middle of the last century. Baker (1965) reports it as a means that guarantees the adaptive success of this type of plants. In this research, the unique regions of ITS nuclear ribosomal DNA and ITS+*mat*K concatenated regions were the most variable regions and presented all the desired characteristics of a DNA barcoding that meets the requirements for amplification and sequencing. Considering the genetic method based on distance, the tree-based method, and the phylogenetic analysis methods, there was no discrimination of the evaluated genotypes, showing little genetic variability, therefore,

it is suggested that all the evaluated individuals ("Criolla", "Martinera", and "Costera") belong to the same species and that the existing morphological differences may be related to phenotypic plasticity.

ACKNOWLEDGMENTS

RESUMEN

Código de barras de ADN para la identificación molecular de *Gynerium sagittatum* (Poales: Poaceae): diversidad genética de genotipos de las sabanas de Córdoba, Colombia. Introducción: La fibra de *Gynerium sagittatum* Aubl. P. Beauv, es materia prima esencial para la elaboración de varias artesanías, que son símbolos de la identidad cultural colombiana. En el proceso de fabricación, se utilizan diferentes genotipos de acuerdo con la calidad de la fibra y el tipo de artesanía, pero se cree que *Gynerium* es una especie compleja y hasta la fecha, no hay un consenso sobre si estos genotipos pertenecen a la misma especie o especies diferentes. **Objetivo:** Identificar de forma rápida y precisa plantas de caña silvestre utilizando el espaciador transcrito interno ribosomal nuclear (ITS1+ITS2), tres regiones de cloroplasto (*mat*K, *rbc*L, *ycf*1) y sus combinaciones. **Métodos:** Se utilizaron diferentes pruebas para la discriminación: (1) distancias inter e intraespecíficas, (2) Prueba Best Match (BM), Best Close Match (BCM) y método basado en árboles (3) Neighbor Joining (NJ) y (4) Probabilidad de inferencia bayesiana mediante datos moleculares. **Resultados:** Los resultados mostraron que los enfoques BM y BCM revelaron una baja tasa de identificación correcta de especies para los loci ITS+*mat*K (33.3 %) e ITS (28.6 %), mostrando similitud entre las secuencias. Estos resultados fueron respaldados por análisis basados en árboles, donde todas las regiones individuales y las diferentes combinaciones de genes tuvieron una tasa de discriminación de cero (0 %). **Conclusiones:** los genotipos evaluados pertenecen a la misma especie de caña flecha y las diferencias morfológicas existentes pueden estar relacionadas con plasticidad fenotípica.

Palabras clave: *Gynerium saccharoides*; caña flecha; DNA código de barras; espaciadores internos transcritos (ITS); genes de cloroplastos.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In S. Kotz & N.L. Johnson (Eds.), *Breakthroughs in Statistics Foundations and Basic Theory* (pp. 199-213). New York, USA: Springer.

Aramendiz-Tatis, H., Espitia Camacho, M., & Cardona Ayala, C. (2009). *Valoración de los recursos fitogenéticos de caña flecha (Gynerium sagittatum Aubl.) en el caribe colombiano* (1a Ed). Bogotá: Produmedios.

Ashfaq, M., Asif, M., Anjum, Z.I., & Zafar, Y. (2013). Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (Gossypium: Malvaceae). *Molecular Ecology Resources*, *13*(4), 573-582.

Awad, M., Fahmy, R.M., Mosa, K.A., Helmy, M., & El-Feky, F.A. (2017). Identification of effective DNA barcodes for Triticum plants through chloroplast genome-wide analysis. *Computational Biology and Chemistry*, *71*, 20-31.

Baker, H.G. (1965). Characteristics and modes of origins of weeds. In H.G. Baker, & G.L. Stebbins (Eds.), *The Genetics of Colonizing Species* (pp. 141-172). USA: Academic Press.

Barker, N.P., Linder, H.P., & Harley, E.H. (1995). Polyphyly of Arundinoideae (Poaceae): Evidence from rbcL Sequence Data. *Systematic Botany*, *20*(4), 423-435.

Birch, J.L., Walsh, N.G., Cantrill, D.J., Holmes, G.D., & Murphy, D.J. (2017). Testing efficacy of distance and tree-based methods for DNA barcoding of grasses (Poaceae tribe Poeae) in Australia. *PLoS ONE, 12*(10), e0186259.

CBOL-Plant Working Group. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America, 106*(31), 12794-12797.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., & Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE, 5*(1), e8613.

Copaci, C.M., Pocol, I., Căprar, M., & Sicora, C. (2015). Evaluating the potential of a few barcode markers in identifying the species Calluna vulgaris (L.) Hull. *Journal of Horticulture, Forestry and Biotechnology, 19*(2), 57-61.

Darriba, D., Taboada, G.L., Doallo, R., & Posada, D. (2012). JModelTest 2: More models, new heuristics and parallel computing. *Nature Methods, 9*(8), 772-772.

Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., & Zhou, S. (2015). ycf1, the most promising plastid DNA barcode of land plants. *Scientific Reports, 5*, 8348.

Doyle, J., & Doyle, J. (1990). Isolation of plant DNA from fresh tissue. *Focus, 12*, 13-15.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797.

Grass Phylogeny Working Group II. (2011). Rapid report New grass phylogeny resolves deep evolutionary relationships and discovers C 4 origins. *New Phytologist, 193*(2), 304-312.

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist, 125*(1), 1-15.

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95-98.

Hartvig, I., Czako, M., Kjaer, E.D., Nielsen, L.R., & Theilade, I. (2015). The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). *PLoS ONE, 10*(9), e0138231.

Hebert, P.D.N., Cywinska, A., Ball, S.L., & deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the royal society B. Biological Sciences , 270* (1512), 313-321.

Heinrichs, J., Kreier, H.P., Feldberg, K., Schmidt, A.R., Zhu, R.L., Shaw, B., & Wissemann, V. (2011). Formalizing morphologically cryptic biological entities:

New insights from DNA taxonomy, hybridization, and biogeography in the leafy liverwort *Porella platyphylla* (Jungermanniopsida, Porellales). *American Journal of Botany, 98*(8), 1252-1262.

Hollingsworth, P.M. (2014). A DNA barcode for land plants. *Molecular Ecology Resources, 14*(3), 437-446.

Hulme, P.E. (2007). Phenotypic plasticity and plant invasions: Is it all Jack? *Functional Ecology, 22*(1), 3-7.

Kalliola, R., Puhakka, M., & Salo, J. (1992). Intraspecific Variation, and the Distribution and Ecology of Gynerium-Sagittatum (Poaceae) in the Western Amazon. *Flora, 186*, 153-167.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution, 16*(2), 111-120.

Kress, W.J., Erickson, D.L., Jones, F.A., Swenson, N.G., Perez, R., Sanjur, O., & Bermingham, E. (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 18621-18626.

Meier, R., Shiyang, K., Vaidya, G., & Ng, P.K.L. (2006). DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology, 55*(5), 715-728.

Neubig, K.M., & Abbott, J.R. (2010). Primer development for the plastid region YCF1 in annonaceae and other magnoliids. *American Journal of Botany, 97*(6), 52-55.

Parveen, I., Singh, H.K., Raghuvanshi, S., Pradhan, U.C., & Babbar, S.B. (2012). DNA barcoding of endangered Indian *Paphiopedilum* species. *Molecular Ecology Resources, 12*(1), 82-90.

Rivera-Jiménez, H., Rossini, B.C., Tambarussi, E.V., Veasey, A., Ibanes, B., & Marino, C.L. (2017). Acta Scientiarum DNA barcode regions for differentiating *Cattleya walkeriana* and *C. loddigesii*. *Acta Scientiarum - Biological Sciences, 39*(2016), 45-52.

Rivera-Jiménez, J.H., Suárez-Padrón, I.E., & Palacio-Mejía, J.D. (2009). Analysis of the genetic diversity of "caña flecha" Gynerium sagittatum Aubl. usign the AFLP technique. *Agricultura Técnica en México, 35*(1), 78-84.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S.,… & Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology, 61*(3), 539-542.

Selvaraj, D., Shanmughanandhan, D., Sarma, R.K., Joseph, J.C., Srinivasan, R.V., & Ramalingam, S. (2012). DNA barcode ITS effectively distinguishes the

medicinal plant *Boerhavia diffusa* from its adulterants. *Genomics, Proteomics & Bioinformatics*, *10*(6), 364-367.

Singh, H., Parveen, I., Raghuvanshi, S, & Babbar, S. (2011). The loci recommended as universal barcodes for plants on the basis of floristic studies may not work with congeneric species as exemplified by DNA barcoding of *Dendrobium* species. *BMC Research Notes*, *5*(1), 1-11.

Soltis, P.S., & Soltis, D.E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology*, *60*, 561-588.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Stöver, B.C., & Müller, K.F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, *11*, 1-9.

Sultan, S.E. (2001). Phenotypic plasticity for fitness components in Polygonum species of contrasting ecological breadth. *Ecology, 82*(2), 328-343.

Tahir, A., Hussain, F., Ahmed, N., Ghorbani, A., & Jamil, A. (2018). Assessing universality of DNA barcoding in geographically isolated selected desert medicinal species of Fabaceae and Poaceae. *PeerJ, 6,* e4499.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725-2729.

Xu, S., Li, D., Li, J., Xiang, X., Jin, W., Huang, W., & Huang, L. (2015). Evaluation of the DNA barcodes in dendrobium (Orchidaceae) from mainland Asia. *PLoS ONE*, *10*(1), 1-12.

Yang, C.F., Yang, L.T., Li, Y.R., Zhang, G.M., Zhang, C.Y., & Wang, W.Z. (2016). Sequence Characteristics and Phylogenetic Implications of the nrDNA Internal Transcribed Spacers (ITS) in Protospecies and Landraces of Sugarcane (*Saccharum officinarum* L.). *Sugar Tech*, *18*(1), 8-15.

Zorro, W.A., & Prieto, F.V. (1999). *Aproximación a la problemática económica productiva de la comunidad indígena Zenú*. Colombia: Universidad Nacional de Colombia.

See Digital Appendix at: / Ver Apéndice digital en:
revistas.ucr.ac.cr

Rev. Biol. Trop. (Int. J. Trop. Biol.) • Vol. 68(4): 1049-1061, December 2020

1061