CROP BREEDING AND APPLIED BIOTECHNOLOGY

ARTICLE

Assessing DNA sequence variation in *Megathyrsus maximus*: an exploratory survey for use in breeding programs

Antonieta A. Jesus[1], Luice G. Bueno[2], Geice R. Silva[3], Diego B. Galvani[2], Liana Jank[2], Lucimara Chiari[2], Mateus F. Santos[2] and Fábio M. Diniz[3]

Crop Breeding and Applied Biotechnology
24(4): e46102447, 2024
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332024v24n4a44

Abstract: *Distinguishing Megathyrsus maximus genotypes in breeding programs based on morphological traits alone can be challenging. Thus, this study proposes the utilization of molecular traits to identify genetic differences and understand the characteristics of these genotypes. Twenty-four genotypes were sequenced, resulting in 341 bases within the trnL-trnF spacer and 459 bases from ITS2 sequences. These sequences were head-to-tail concatenated and analyzed for phylogenetic diversity. Substitution saturation analysis confirmed the sequences' suitability for phylogenetic reconstruction, as ISS value (0.0279) was significantly lower than its critical value (ISS.c = 0.7523). Both Maximum Parsimony and Bayesian Inference analyses yielded consistent results with most nodes showing support above 99%. PCoA analysis demonstrated a clustering pattern consistent with the phylogenetic tree, identifying two distinct groups with clear separation. Approximately 82% of the total variance was explained by the PCoA. The obtained genetic information serves as a valuable resource for developing M. maximus markers.*

Keywords: *Tropical forage, sequencing, phylogenetic reconstruction, promising genotypes*

INTRODUCTION

The grass species *Megathyrsus maximus* (Jacq.) B.K. Simon & S.W.L. Jacobs (syn: *Panicum maximum* Jacq), commonly referred to as "guinea grass", is a tropical forage of Eastern and Southern African origin. The species is among the most cultivated grasses in Brazil due to its desirable attributes such as adaptability, productivity, nutritional value, easy propagation by seeds, and great acceptability for ruminant animals (Jank et al. 1997). Many cultivars of *M. maximus* have been developed for use in different production systems and environmental conditions.

In breeding programs, the differentiation of genotypes may be naturally difficult if only morphological characters are taken into consideration. Phenotypic identification based on traditional morphological methods may not generate reliable classifications at intraspecies-level assessments, as morphological characters are susceptible to environmental and developmental factors (Meiklejohn et al. 2019).

***Corresponding author:**
E-mail: fabio.diniz@embrapa.br
ORCID: 0000-0003-0867-3552

Received: 17 May 2024
Accepted: 20 August 2024
Published: 04 October 2024

[1] Universidade Federal do Piauí, Northeast Biotechnology Network, Renorbio, 64049-550, Teresina, PI, Brazil
[2] Embrapa Gado de Corte, Avenida Rádio Maia, 830, Zona Rural, 79106-550, Campo Grande, MS, Brazil
[3] Embrapa Caprinos e Ovinos, Fazenda Três Lagoas, Estrada Sobral/Groaíras, km 4, 62010-970, Sobral, CE, Brazil

Molecular markers have been extensively used as a tool to address questions related to genetic diversity, identification of plant species and cultivars, conservation of plant genetic resources (Rout and Mohapatra 2006, Elansary et al. 2017) and variability existing in germplasm collections (Costa et al. 2011). Therefore, molecular traits can be useful in providing evidence of intraspecific genetic differences without environmental influences (Meiklejohn et al. 2019) and, hence, are indispensable for plant breeding programs. Particularly, DNA sequences from organelle and/or nuclear genomes have been used for genetic differentiation and phylogenetic studies in plants, depending on their rate of evolution and on the purpose of the performing task (Odintsova and Yurina 2003). Coding and non-coding regions within the chloroplast genome (*rbc*L*, trn*L) and nuclear ribosomal DNA (ITS1 and ITS2) are repeatedly reported to be suitable for genotype characterization and for revealing genetic relationships among infraspecific taxa, such as at cultivar level, in many commercial species of angiosperm families (Lee et al. 2017, Duan et al. 2019, Dhivya et al. 2020, Vieira et al. 2022).

Forage breeding programs can benefit from developing molecular tools to improve breeding efficiency, and therefore, boosting breeding process. In this context, however, molecular-genetic information on *M. maximus* available in the literature is still scarce. Thus, this study is aimed to perform a phylogenetic analyses of *M. maximus* genotypes as a perspective of using the nucleotide variation in the plastome *trn*L-*trn*F and nuclear ITS2 as DNA markers for genetic relationship and characteristics of *M. maximus* genotypes.

MATERIAL AND METHODS

Plant material, DNA extraction and quantification

Seeds of 24 genotypes of *M. maximus* (5 cultivars and 19 accessions) were obtained from the Embrapa Beef Cattle germplasm collection and planted in 11-L plastic pots (2 replicates for genotype), at the experimental station of Embrapa Goats and Sheep, in Sobral, Ceará, Brazil. Fresh young leaves were collected from each genotype, dehydrated and, then, samples of about 100 mg of dried material were manually macerated in liquid nitrogen. Total DNA was isolated from dried leaves using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) according to the manufacturer's instruction. DNA extracts were checked for quality and quantity on a 1% agarose gel electrophoresis with 0.04 µL mL^{-1} of ethidium bromide (10 mg mL^{-1}) and viewed on a UV transilluminator.

DNA amplification and sequencing

Polymerase chain reactions (PCRs) were performed in a 20 µL total volume containing approx. 15 ng of genomic DNA, 0.2 µg mL^{-1} BSA, 2.0 µM of 10× PCR buffer (Thermo Fisher, Waltham, MA, USA), 2.0 µM $MgCl_2$ (Thermo Fisher), 0.5 µM dNTP (New England BioLabs, Ipswich, MA, USA), 0.3 µM of each primer, 1 U *Taq* DNA Polymerase (Invitrogen Life Technologies, Carlsbad, CA, USA) and ultrapure UVed water.

One nuclear (ITS2) and two chloroplast regions (*rbc*L and *trn*L-*trn*F spacer) were analyzed. The PCR amplifications were run in a Veriti 96-well Thermal Cycler (Applied Biosystems, Foster City, CA, USA) with the primers indicated below. The forward and reverse primers were Pla:5'-GGAAGGAGAAGTCGTAACAAGG-3'/P4:5'-TCCTCCGCTCATTGATATGC-3' for all of ITS1, 5.8S, and ITS2 regions (Downie and Katz-Downie 1996, White et al. 1990), *trn*L (UAA):5'-GGTTCAAGTCCCTCTATCCC-3'/ *trn*F (GAA):5'-ATTTGAACTGGTGACACGAG-3' for *trn*L (Taberlet et al. 1991), and *rbc*L1:5'-ATGTCACCACAAACAGARACTAAAGC-3'/ *rbc*L3ambigR:5'-GGCGGACCTTGGAARTATAAG-3' for *rbc*L (Olmstead et al. 1992). *The following PCR cycle conditions* were used: pre-denaturation of 94 °C for 5 min, followed by 30 cycles of 94 °C for 1 min, Ta °C (according to each primer pair) for 1 min, and 72 °C for 1 min, and a final extension of 72 °C for 5 min. Amplicons were checked for correct size and quality by gel electrophoresis with ethidium bromide staining.

PCR products were purified with ExoSAP-IT™ PCR Product Cleanup (Applied Biosystems, Foster City, CA, USA) and quantified using Nanodrop 2000c (Thermo Fisher, Waltham, MA, USA) to remove the excess primers and nucleotides. Sequencing of DNA fragments was performed on the AB 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) by ACTGene Molecular Analyses Ltda (Biotechnology Center, UFRGS, Porto Alegre, RS, Brazil) using 50-cm capillaries and POP7 polymer (Applied Biosystems, Foster City, CA, USA), with 2.5 pmol of the same primer used in PCR reactions and 0.5 mL of BigDye™ Terminator v3.1 (Thermo Fisher Scientific, Waltham, MA, USA) in a final volume of 10 mL. All new unique nucleotide sequences were submitted to the GenBank database.

DNA sequence analyses

DNA sequences from each genotype were inspected with CHROMAS version 2.6.6 (Technelysium, Brisbane, QLD, Australia) and corrected when necessary, followed by unique sequences detection using the software DAMBE v.5.2 (Xia and Xie 2001). Homologous nucleotide sequences from all samples were aligned using the software CLUSTALX version 1.81 (Thompson et al. 1994) on the sequence alignment editor BIOEDIT version 7.2.6.1 (Hall 1999). Finally, DNA sequences of each region were concatenated for further analyses.

Sequence diversity

Genetic diversity indices were estimated by calculating the mean nucleotide composition, indels, haplotype (H) and nucleotide (π) diversities using the software DNASP v.5.10 (Librado and Rozas 2009). The aligned sequences were also imported into MegaX, v.10.2.6 (Kumar et al. 2018) for the estimation of transitions and transversions. Maximum Composite Likelihood estimates of the nucleotide substitution pattern were calculated using the MegaX program. To avoid inconsistencies in phylogenetic reconstruction due to saturation of nucleotide base substitutions, the occurrence of substitution saturation was also investigated with DAMBE v.5.2.

Phylogenetic analyses

Sequences were analyzed using the maximum-parsimony (MP) approach in MegaX (Kumar et al. 2018). A heuristic search with the tree-bisection–reconnection (TBR) branch-swapping algorithm with 1000 random taxon addition replicates was used. Node support was assessed using 1000 bootstrap replicates with 10 random taxon addition replicates. A Bayesian likelihood approach was also used in MrBayes v.3.1 (Ronquist and Huelsenbeck 2003) in a partitioned analysis of each gene. The model of DNA substitution that best fitted the genomic data was selected based on the Akaike Information Criterion (AIC) (Posada and Buckley 2004) using JModeltest v.3.06 (Posada 2008). The Bayesian Inference (BI) analysis was carried out using the default setting of four Markov chains (three heated and one cold) for 1.000.000 generations, sampling once every 100 generations. Four separate analyses were carried out simultaneously starting with different random trees. Post-burn-in trees from all four analyses were used to estimate posterior probabilities. FigTree v1.4.3 software (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize the tree. The genetic similarity among samples was estimated using PAST v1.34 (Hammer et al. 2001). A multidimensional principal coordinate analysis (PCoA) was performed on the data set to reveal the degree of genetic differentiation between sites.

RESULTS AND DISCUSSION

Sequence diversity

No nucleotide polymorphism was detected in the *rbc*L sequences in all studied genotypes and, therefore, this sequencing data has been discarded from further analysis. PCR fragments were successfully amplified for all primer pair combinations. The Pla-P4 primers encompassed all of ITS1, 5.8S, and ITS2 regions, the *rbc*L1-*rbc*L3ambigR primers covered the chloroplast gene encoding the RuBisCO protein (*rbc*L), and the *trnL*(UAA)-*trn*F(GAA) primers targeted the *trn*L spacer (Figure 1).

Targeted sequencing revealed a clear read length of 341 bases within the *trn*L-*trn*F intergenic spacer, and 459 bases from internal transcribed spacer2 sequences. To our knowledge, this is the first evaluation of the power and robustness

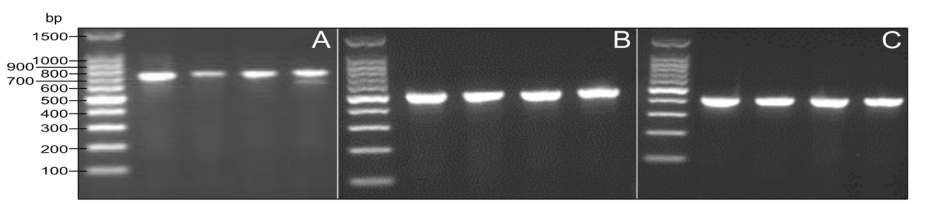

Figure 1. An ethidium-bromide-stained agarose gel showing PCR products from the amplification of *Megathyrsus maximus* DNA extracts using primer pair combinations. (A): Pla and P4 for ITS1-5.8S-ITS2); (B): *rbc*L1 and *rbc*L3ambigR for *rbc*L); and (C): *trn*L (UAA) and *trn*F (GAA) for *trn*L spacer). Molecular marker: 100-bp ladder (*TransGen Biotech Co.* Ltd.).

of these two spacers for *M. maximus*. Sequencing sizes of these DNA regions are similar to those previously described for forage grasses (Taberlet et al. 1991, Baldwin et al. 1992, Stoneberg-Holt et al. 2005). Because the ITS2 region of angiosperms are expected to be less than 300 base pairs (Baldwin et al. 1992), the additional 159 base pairs that we sequenced in this region were due to the inclusion of partial 5.8S and large subunit ribosomal RNA gene sequences neighboring the ITS2.

A nucleotide BLAST analysis (http://www.ncbi.nlm.nih.gov/blast/) of all obtained sequences revealed highly similar sequence identity to *M. maximus* available in GenBank (KU291470.1 and AY129712.1). All samples accumulated percent maximum identity of 100% and 99.6%, for the *trn*L-*trn*F and ITS2 spacers, respectively. DNA sequences from these spacers were head-to-tail concatenated (internal transcribed spacer2 + *trn*L-*trn*F intergenic spacer) to produce a more robust dataset. The unique sequences of these regions were deposited separately into GenBank under accession numbers ON720306-ON720322 (ITS), ON720323-ON720325 (*rbc*L), and ON720326-ON720328 (*trn*L-*trn*F).

Analysis of the concatenated sequences (~760 bases) revealed 37 polymorphic sites (ITS2 = 34 and *trn*L = 3), from which 32 were informative and 5 uninformative positions (Table 1). Although the nucleotide composition across the concatenated sequences were almost evenly distributed, A (24.8%), C (23.1%), T (26.1%), and G (24.9%), the GC- and AT-contents of ITS2 and *trn*L-trnF sequences varied slightly within each spacer, but considerably between the two spacers for all samples. Thus, for ITS2 mean GC- and AT-contents were 58.0% and 41.9%, respectively. For *trn*L these values were 34.8% and 65.1%. The GC-content of the ITS2 and the *trn*L-trnF spacer were typical of Poaceae species (Ghariani et al. 2020, Omelchenko et al. 2022). Overall, the GC- and AT-contents in the concatenated sequences were respectively 47.7% and 51.9%, respectively. The GC-content may affect genome functioning and species ecology (Šmarda et al. 2014).

Table 1. Summary of polymorphic sites in the studied regions of *Megathyrsus maximus*. The polymorphic sites are numbered in vertical format

Genotype/cultivar	Polymorphic sites		Origin
	ITS2	*trn*L	
	0000001111 1111111222 2222222222 3344 7889990112 2222226112 2233556679 8801 8393597451 2345892450 2719340816 5639	445 679 908	
PM20	AATGCTTTCC TATTTGCTTT GAACATTCCA GCCT	CTC	Morogoro, Tanzania
PM44	AATGCCTTCC TATTTGCTTT GGACAATCCA GCCT	AAT	Hybrid (unknown origin)
PM37	AATGCTTTCC TATTTGCTTT GAACAATCCA GCCT	CTC	Dar, Tanzania
PM19	AATGCTTTCC TATTTGCTTT GAACATTCCA GCCT	CTC	Morogoro, Tanzania
PM269	CCTACCACTT GCATCATCCC GGTCAACTCG ATCC	CTC	Nairobi, Kenya
PM122	CCTACTACTT GCATCATCCC GGTCAATTCG ATTC	CTC	Nairobi, Kenya
PM21	CCTATCACTT GCATCATCCC GGTCAACTCG ATCC	AAT	Morogoro, Tanzania
PM294	CCTACTACTT GCATCATCCC GGTCAATTCG ATTC	CTC	Rumuruti, Kenya
PM119	CCTACCTCTT GCATCGTCCT GGTCAACTCA ACCT	AAT	Nairobi, Kenya
PM405	CCTACCACTT GCATCATCCC GGTCAACTCG ATCC	CTC	Dar, Tanzania
PM150	CCTACCACTT GCATCATCCC GGTCAACTCG ATCC	AAT	Fortnall, Kenya
PM22	CCTGCCTCCC TATTTGTCCT GGTCAACTCG GCCT	AAT	Morogoro, Tanzania
PM190	CCTACCACTT GCATCATCCC GGTCAACTCG ATCC	CTC	Rogwe-Tanga, Tanzania
PM406	CCTACTACTT GCATCATCCC GGTCAATTCG ATCC	CTC	L. Lunga-Mombasa, Kenya
PM51	CCTACCACTT GCATCATCCC GGTCAACTCG ATTT	AAT	Brazzaville, Congo
PM13	CCTACCACTT GCATCATCCC AGTCAATTCG ATCC	CTC	Rogwe-Kilosa, Tanzania
PM59	CCCGCTTCTC TCACCGTCCT GGTAAATTCA ACCC	AAT	Daloa, Ivory Coast
PM234	CCTACTTCTC TCATCGTCCC GGTCAACTCA ACCT	AAT	Tuli, Botswana
Aruana	CCTACTACTT GCATCATCCC GGTCAATTCG ATTC	CTC	West and Central Africa
Quenia	CCTACCACTT GCATCATCCC GGTCAACTTG ATCC	CTC	Korogwe-Kilosa, Tanzania
Tamani	CCTACCACTT GCATCATCCT GGTCGATTCG ATCC	AAT	Rogwe-Kilosa, Tanzania
Massai	AATGCCTTCC TATTTGTTTT GAACATTCCA GCCT	CTC	Dar/ Salaam-Bagamoyo, Tanzania
Tanzania	CCCACCTCTC TCATCGTCCT GGTCGACTCA ACCT	CTC	Korogwe, Tanzania

Aligned sequences indicated the presence of insertions or deletions (InDels). The ITS2 spacer showed only one InDel, while the *trn*L intron revealed the presence of five distinct InDels. The diversity of InDels per site presented in the germplasm collection was 0.0014. Even though insertions and deletions are widespread in plant genomes, they occur more often in non-coding regions (Guo et al. 2016). Since insertions/deletions-associated nucleotide polymorphism in chloroplast noncoding regions are frequently used as genetic markers (Hamilton et al. 2003), these InDels were also used in the phylogenetic analysis.

Twenty-four haplotypes were detected, demonstrating a high degree of haplotypic diversity (0.96) among the studied genotypes. The mean nucleotide diversity per site (p) was 0.01231. In general, these values indicate that few sites are polymorphic and few mutations exist to differentiate one haplotype from the other. Transition and transversion ratios were k1 = 2.038 (purines) and k2 = 6.233 (pyrimidines). The overall transition/transversion ratio was estimated to be 2.058. The relatively high content of AT for all concatenated sequences (i.e., 51.9%) may partly explain the high proportion of transitions identified in the current study. Overabundance of transitions relative to transversions may be due to chemical constraints, as AT-rich DNA segments mutate faster than those rich in GC, owing to an extra hydrogen bond. Transition bias has been widely reported among studies of mutation rates (Wolfe et al. 1987).

Phylogenetic inference

In the substitution saturation analysis, the ISS value (0.0279) was significantly lower (P = 0.0000) than its critical value (ISS.c = 0.7523), indicating that the sequences are much less susceptible to substitution saturation and, therefore, useful for phylogenetic reconstruction (Xia et al. 2003). Results from sequence alignment analyses allowed inferring phylogenetic relationships among *M. maximus* accessions and commercial cultivars. Thus, one of the 100 most parsimonious trees (length = 7) had consistency (CI) and retention (RI) indices, which are measures of how well an individual material fits on a phylogenetic tree under a parsimony model, equal to 1.00, with most nodes showing consistency above 99%. Accessions and cultivars were represented in a monophyletic group.

In the MP tree, a small subgroup formed by the accessions PM120, PM269, and PM122 clustered together, probably due to a common origin (Figure 2A). The Bayesian analysis, however, did not support such grouping. Different analytical methods can lead to different conclusions, and given the ongoing debate over the superiority of one of these phylogenetic reconstruction methods (i.e., Parsimony versus Bayesian methods), it remains valid to consider that parsimony trees might be more suitable for less complex, smaller data sets or for preliminary exploratory analyses because of their speed and simplicity. In contrast, Bayesian methods are better for large quantities of complex data and for analyses where a robust statistical modeling is required (Sansom et al. 2018). These slightly diverse genotypes can promptly be used for further improvements (Kumar et al. 2019) in breeding programs for the species.

Overall, the individual phylogenies were essentially consistent with each other. The tree based on Bayesian inference, which holds the maximum *a posteriori* probability for all nodes, showed similar topology and equivalent node supports to the MP method (Figure 2A). Probability values greater than or equal to 0.95 are considered satisfactory and significant for phylogenetic studies (Rannala and Yang 1996). Two well-defined groups were shaped using the BI (Figure 2B), which might be also associated to the pattern observed in the MP tree. Group I included accessions PM120, PM20, PM269, PM122, PM294, PM405, PM190, PM37, PM406, PM13, PM19 and clustered with most of commercial cultivars (Aruana, Quenia, Massai and Tanzania), while group II contained all remaining accessions. High degree of similarity among the studied sequences was observed within each group.

The Massai cultivar, present in group I, showed the greater distance from other accessions. This process may occur due to the high rate of non-synonymous amino acid mutation and small variation found in the second group formed, mainly observed in the accessions PM51, PM59, PM234, and Tamani cultivar, while most materials remain uniform in the phylogenetic tree.

The two-dimensional plot from the PCoA *showed* a clustering pattern synonymous with the BI phylogenetic tree (Figure 2B). Two distinct groups were identified with a neat separation with no overlap, one along the upper quadrants and the other in the lower quadrants, which corroborates the clustering pattern observed in Figure 2. The plotted data indicated the presence of an important variability among plant materials, confirmed by the elongated ellipses.

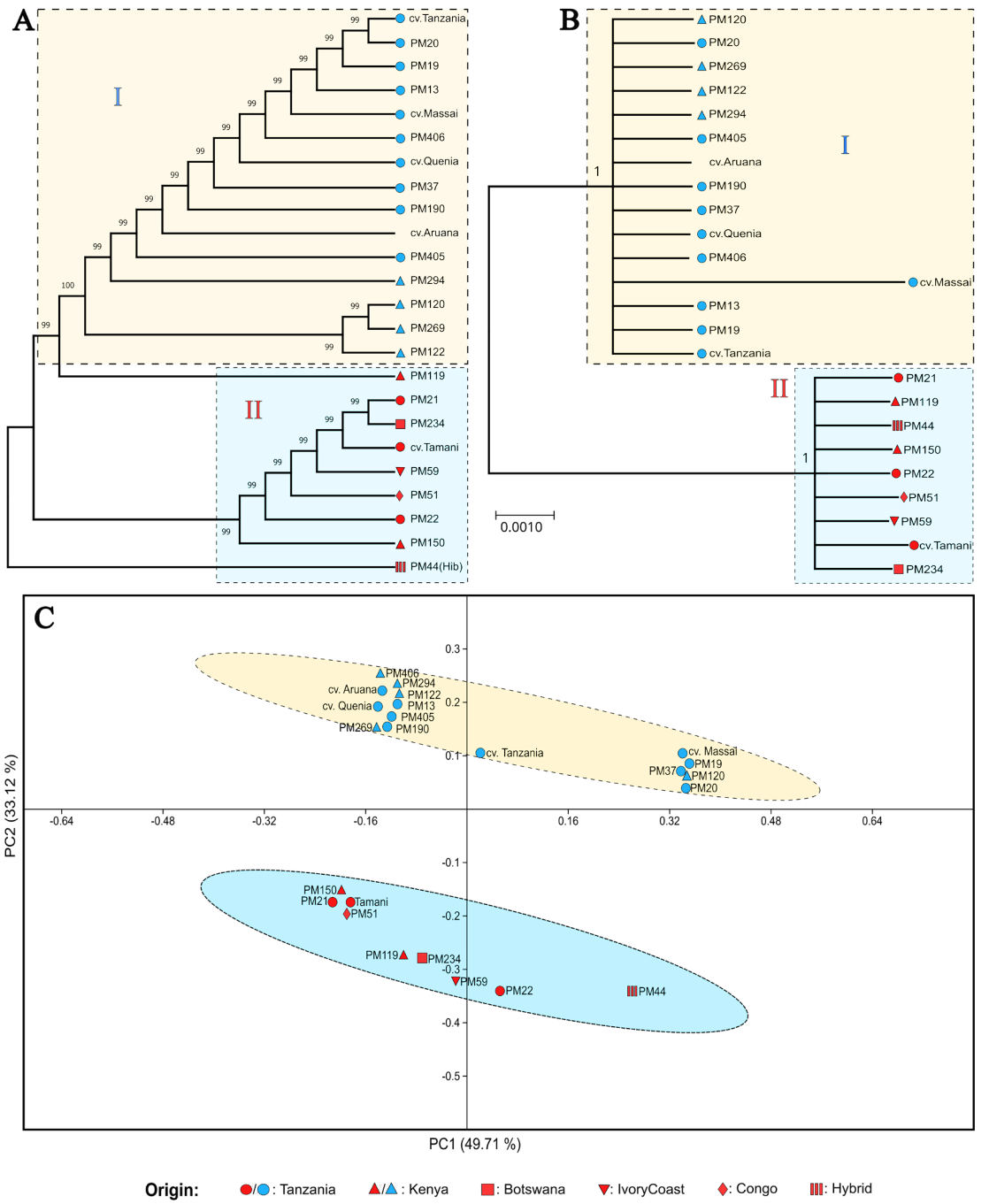

Figure 2. (A-B) Phylogenetic trees representing *Megathyrsus maximus* accessions (PM) and commercial cultivars based on ITS2 and *trn*L-*trn*F spacers. (A) Maximum parsimony (MP) analysis with the numbers at the nodes indicating bootstrap values; (B) Bayesian inference (BI) analysis with posterior probability values. Symbols after accessions and cultivars are associated with the Principal Coordinate Analysis (PCoA); (C) Scatter-plot of the principal coordinate analysis (PCoA) based on the variation of each nucleotide within DNA sequences. Dotted line represents inferred genetic group I and group II. The color of each symbol indicates the inferred genetic group I (blue) and II (red). The symbols represent the predefined origin of the accessions, as illustrated below Figure 2C. Inertia ellipses represent the substructures visible in the graph.

Approximately 82% of the total variance was explained by the PCoA (Figure 2C). The first component (PC1) described 49.71% and the second component (PC2) 33.12% of the total variability among accessions.

Noticeably, although accessions were distributed continuously along principal component axes, a few sub-clusters were also observed. The PCoA revealed no clustering based on the geographic origin of the accessions, indicating that geographic isolation does not impact the genetic diversity of *Megathyrsus maximus*. This suggests that mechanisms other than geographic isolation may contribute to their genetic differentiation.

In each of these groups, however, subgroups were formed by the similarity among some genetic materials, with the isolation of the accession PM44 and the Tanzania cultivar standing out. Accession PM44, being a hybrid, shows a significant genetic distance from the other accessions. As for Tanzania, the difference in adaptation to different ecogeographic conditions and local selection pressure can explain this isolation (Jo et al. 2012). Several factors may be behind the mutations accumulated throughout history, but the greater the diversity, the higher the chance that in the population some individuals have alleles that contribute to facing environmental changes, such as biotic or abiotic stresses (Able et al. 2007).

Based on the maximum parsimony (MP) analysis, accession PM119 appears to be more closely related to the "group I", as the branch length between PM119 and PM21 is slightly longer than the branch between PM119 and PM122, suggesting that PM119 and PM122 have an even closer evolutionary relationship and similarity, which can also be explained by a shared origin in Nairobi, Kenya. However, this closer relationship was not detected by Bayesian inference analysis and PCoA, which clustered these two accessions into "group II". The conflicting grouping of PM119 show how different phylogenetic approaches can lead to different conclusions about evolutionary relationships. This fact emphasizes the importance of considering multiple lines of evidence when interpreting the evolutionary relationships among the accessions.

As the results obtained through PCoA corroborate the clustering of PM119 into group II obtained by the Bayesian analysis, we decided to prioritize this scenario for genetic grouping composition. This decision is also based on the general consensus that Bayesian inference analysis is a more robust and reliable method (Mohammadi and Prasanna 2003).

CONCLUSIONS

1) The *rbc*L gene is not to be recommended for the characterization of *M. maximus* accessions and cultivars;

2) The *trn*L-*trn*F intergenic spacer and the ITS2 region, as a two-sequence multilocus marker, are efficient in the characterization of *M. maximus* accessions and cultivars;

3) The occurrence of insertions and/or deletions in plastid DNA is useful for resolving relationships between closely related sequences, and therefore, increasing the power of analysis in intraspecific studies.

ACKNOWLEDGMENTS

This work was supported by EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária), SEG number: 20.18.01.017.00.00 (Development of cultivars of *Panicum maximum* Jacq), and a fellowship from CAPES (National Council for the Improvement of Higher Education Personnel). We thank Mrs. Lidiane Viana Ximenes for laboratory assistance, and Mr. José Wilson Tavares Bezerra, Mr. Pedro José Correia, and Mr. Francisco de Mesquita Pereira for field assistance.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current research are available from the corresponding author upon reasonable request.

REFERENCES

Able JA, Langridge P and Milligan AS (2007) Capturing diversity in the cereals: many options but little promiscuity. **Trends in Plant Science 12**: 71-79.

Baldwin BG (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. **Molecular Phylogenetics and Evolution 1**: 3-16.

Costa TS, Silva AVC and Lédo AS (2011) Diversidade genética de acessos

do banco de germoplasma de mangaba em Sergipe. **Pesquisa Agropecuária Brasileira 46**: 499-508.

Dhivya S, Ashutosh S, Gowtham I, Baskar V, Baala Harini A, Mukunthakumar S and Sathishkumar R (2020) Molecular identification and evolutionary relationships between the subspecies of *Musa* by DNA barcodes. **BMC Genomics** 21: 659.

Downie SR and Katz-Downie DS (1996) A molecular phylogeny of Apiaceae subfamily apioideae: evidence from nuclear ribosomal DNA internal transcribed spacer sequences. **American Journal of Botany** 83: 234-251.

Duan H, Wang W, Zeng Y, Guo M and Zhou Y (2019) The screening and identification of DNA barcode sequences for *Rehmannia*. **Scientific Reports** 9: 17295.

Elansary HO, Ashfaq M, Ali HM and Yessoufou K (2017) The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry. **PLOS ONE** 12: e0172170.

Ghariani S, Charfeddine A, Amari M, Chakroun M and Neila TF (2020) Interspecific molecular variation of Tunisian complex *Lolium perenne* L. and *Festuca arundinacea* Schreb. based on the internal transcribed spacer locus (ITS). **Physiology and Molecular Biology of Plants 26**: 331-339.

Guo C, Du J, Wang L, Yang S, Mauricio R, Tian D and Gu T (2016) Insertions/Deletions-associated nucleotide polymorphism in *Arabidopsis thaliana*. **Frontiers in Plant Science 7**: 1792.

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

Hamilton MB, Braverman JM and Soria-Hernanz DF (2003) Patterns and relative rates of nucleotide and insertion/deletion evolution at six chloroplast intergenic regions in New World species of the Lecythidaceae. **Molecular Biology and Evolution 20**: 1710-1721.

Hammer O, Harper DAT and Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. **Palaeontologia Electronica 4**: 1-9.

Jank L, Callixto S, Costa JCG, Savidan YH and Curvo JBE (1997) Catálogo de caracterização e avaliação de germoplasma de *Panicum maximum*: descrição morfológica e comportamento agronômico**.** Embrapa Gado de Corte, Campo Grande, 57p. (**Documentos, 68**).

Jo MH, Ham I, Moe KT, Kwon SW, Lu FH, Park YJ, Kim WS, Won MK, Kim T and Lee EM (2012) Classification of genetic variation in garlic (*Allium sativum* L.) using SSR markers. **Crop Science 6**: 625-631.

Kumar M, Sharma VR, Kumar V, Sirohi U, Chaudhary V, Sharma S, Saripalli G, Naresh RK, Yadav HK and Sharma S (2019) Genetic diversity and population structure analysis of Indian garlic (*Allium sativum* L.) collection using SSR markers. **Physiology Molecular Biology Plants 25**: 377-386.

Kumar S, Stecher G, Li M, Knyaz C and Tamura K (2018) Mega X: molecular evolutionary genetics analysis across computing platforms. **Molecular Biology and Evolution 35**: 1547-1549.

Lee S-C, Wang C-H, Yen C-E and Chang C (2017) DNA barcode and identification of the varieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences. **Journal of Food and Drug Analysis 25**: 260-274.

Librado P and Rozas J (2009) DNASP V5: a *software* for comprehensive analysis of DNA Polymorphism Data. **Bioinformatics 25**: 1451- 1452.

Meiklejohn KA, Damaso N and Robertson JM (2019) Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. **PLOS One 14**: e0217084.

Mohammadi SA and Prasanna BM (2003) Analysis of genetic diversity in crop plants. Salient statistical tools and considerations. **Crop Science 43**: 1235-1248.

Odintsova MS and Yurina NP (2003) Plastid genomes of higher plants and algae: structure and functions. **Molecular Biology 37**: 649-662.

Olmstead RG, Michaels HJ, Scott KM and Palmer JD (1992) Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of rbcL. **Annals of the Missouri Botanical Garden 79**: 249-265

Omelchenko DO, Krinitsina AA, Kasianov AS, Speranskaya AS, Chesnokova OV, Polevova SV and Severova EE (2022) Assessment of ITS1, ITS2, 5'-ETS, and *trnL-F* DNA barcodes for metabarcoding of Poaceae pollen. **Diversity 14**: 191.

Posada D (2008) jModelTest: Phylogenetic model averaging. **Molecular Biology and Evolution 25**: 1253-1256.

Posada D and Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. **Systematic Biology 53**: 793-808.

Rannala B and Yang Z (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. **Journal of Molecular Evolution 43**: 304-311.

Ronquist F and Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. **Bioinformatics 19**: 1572-1574.

Rout GR and Mohapatra A (2006) Use of molecular markers in ornamental plants: a critical reappraisal. **European Journal of Horticultural Science 71**: 53-68.

Sansom RS, Choate PG, Keating JN, Randle E. (2018) Parsimony, not Bayesian analysis, recovers more stratigraphically congruent phylogenetic trees. **Biology Letters 14**: 20180263.

Šmarda P, Bureš P, Horová L, Leitch IJ, Mucina L, Pacini E, Tichý L, Grulich V and Rotreklová O (2014) Ecological and evolutionary significance of genomic GC content diversity in monocots. **Proceedings of the National Academy of Sciences USA 111**: e4096-e4102.

Stoneberg-Holt S, Horová L, Bures P, Janecek J and Cernoch V (2005) The trnL-F plastid DNA characters of three *Poa pratensis* (Kentucky bluegrass) varieties. **Plant, Soil and Environment 51**: 94-99.

Taberlet P, Gielly L, Pautou G and Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. **Plant Molecular Biology 17**: 1105-1109.

Thompson JD, Higgings DG and Gibson TJ (1994) CLUSTAL W: improving the sentitive of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weight matrix choice. **Nucleic Acids Research 22**: 4673-4680.

Vieira MB, Faustino MV, Lourenço TF and Oliveira MM (2022) DNA-based tools to certify authenticity of rice varieties - an overview. **Foods 11**: 258.

White TJ, Bruns TD, Lee SB and Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In Innis MA, Gelfand DH, Sninsky JJ and White TJ (eds) PCR Protocols: A guide to methods and applications. Academic Press, New York, p. 315-322.

Wolfe KH, Li W-H and Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. **Proceedings of the National Academy of Sciences USA 84**: 9054-9058.

Xia X and Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. **Journal of Heredity 92**: 371-373.

Xia X, Xie Z, Salemi M, Chen L and Wang Y (2003) An index of substitution saturation and its application. **Molecular Phylogenetics and Evolution 26**: 1-7.