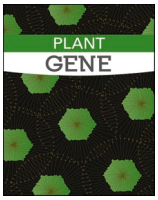
Plant Gene 34 (2023) 100414

Contents lists available at ScienceDirect

Plant Gene

journal homepage: www.elsevier.com/locate/plantgene



Complete genome of the kikuyu grass chloroplast (*Cenchrus clandestinus*) and comparative analysis within the subfamily Panicoideae

Juliana Arango [a,*], Juan F. Alzate [b], Albeiro López [a], Edna J. Márquez [c], José J. Echeverri [a]

[a] Research Group on Biodiversity and Molecular Genetics BIOGEM, Universidad Nacional de Colombia, Medellín, Antioquia, Colombia
[b] Centro Nacional de Secuenciación Genómica-CNSG, SIU, Universidad de Antioquia, Medellín, Colombia
[c] Research Group Animal Biotechnology, Universidad Nacional de Colombia, Medellín, Antioquia, Colombia.

ARTICLE INFO

Edited by A Rooney

Keywords:
Plastome genome
Cenchrus clandestinus
Next generation sequencing
Panicoideae
Poaceae

ABSTRACT

Within the subfamily Panicoideae, the *Cenchrus* genus which is distributed in both tropical and subtropical regions worldwide—is economically important in terms of their production levels. For milkshed areas, one of the most important pastures is kikuyu (*Cenchrus clandestinus*), which represents the basic forage used for feeding in a number of countries. In this study, the kikuyu grass (*Cenchrus clandestinus*) plastome was sequenced, assembled, and annotated to broaden the information and the set of available genomic data. One whole-genome shotgun (WGS) library was constructed using Nextera preparation kits and was sequenced in an Illumina MiSeq platform. In addition, the genomic organization and arrangement of genes as well as their phylogenetic relationship with other species of the family Poaceae were compared using 81 protein-coding genes. The present study characterized and annotated the complete plastome of kikuyu grass, *Cenchrus clandestinus*, as an informative contribution for future studies potentially investigating the evolution of plant genomes and, specifically, aiming to elucidate the phylogenetic relationships within the family Poaceae,Overall, the results indicate that the structure and organization are conserved compared with other references within the family Poaceae. Phylogenetic relationships confirmed the position of kikuyu within the *Cenchrus* genus, and they are consistent with previous results obtained for other species of the subfamily Panicoideae.

s

1. Introduction

The family Poaceae, commonly known as the grass family, consists of about 12 subfamilies divided into two main lineages, namely, the BOP clade (composed of the subfamilies Bambusoideae, Oryzoideae, and Pooideae subfamilies) and the PACMAD clade (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae), and three basal lineages, namely, Anomochlooideae, Pharoideae, and Puelioideae, with around 12,100 species in total. In particular, the subfamilies Pooideae (4234 spp.), Panicoideae (3560 spp.), Bambusoideae (1641 spp.), and Chloridoideae (1601 spp.) are characterized by having the largest number of species described (Kellogg, 2015; Huang et al., 2017).

Within the subfamily Panicoideae, the genera *Cenchrus* and *Pennisetum,* which are distributed in tropical and subtropical regions worldwide, are economically important in terms of their production levels. The kikuyu (aggressive and vigorous perennial that spreads by surface) can be categorised as a grass that has medium–high crude protein content, high fibre and hemicellulose.The genus *Pennisetum* contains about 80–140 species, whereas genus *Cenchrus* includes about 20–25 species, among which are found in forage crops such as *C. setiger, C. ciliaris* (buffel grass), and *C. pennisetiformis* (Donadío et al., 2009; Chemisquy et al., 2010). For milkshed areas, one of the most important pastures is kikuyu (*Cenchrus clandestinus*), which represents the basic forage used for feeding in a number of countries, including Colombia, Costa Rica, Hawaii, New Zealand, and South Africa (Kwazulu-Natal) (Fukumoto and Lee, 2003; García et al., 2014).

The phylogenetic relationships and delimitation between the two genera (*Pennisetum* and *Cenchrus)* is controversial, and some studies have even proposed to modify the nomenclature of several species, considering the inclusion of one genus within the other, and vice versa. Thus, the distinction between the two genera is not clearly defined, and it is the subject of debate in numerous studies (Donadío et al., 2009; Duvall et al., 2001; Aliscioni et al., 2012). By using two plastid markers,

* Corresponding author at: Research Group on Biodiversity and Molecular Genetics BIOGEM, Universidad Nacional de Colombia, Medellín 050001, Colombia.
E-mail address: jarangog@unal.edu.co (J. Arango).

https://doi.org/10.1016/j.plgene.2023.100414
Received 11 August 2022; Received in revised form 4 February 2023; Accepted 9 February 2023
Available online 13 February 2023

Table 1
Species of the subfamily Panicoideae included in the comparative and phylogenetic analyses.

Species	GenBank Accession
Cenchrus americanus	KX756179
Cenchrus purpureus	NC_036384.1
Setaria viridis	NC_028075
Sorghum bicolor	NC_008602.1

a nuclear marker, taxonomic information, and cytogenetics, Chemisquy et al. (2010) determined that *Pennisetum clandestinum* should be renamed as *Cenchrus clandestinus*. They also proposed nomenclature changes for *Panicum americanum* (as *Cenchrus americanus*), *Pennisetum basedowii* (as *Cenchrus basedowii*), *Pennisetum chilense* (as *Cenchrus chilensis*), and *Pennisetum purpureum* (as *Cenchrus purpureus*), among others. However, the use of only partial gene sequences may limit and underestimate the relationship between closely related individuals (Hollingsworth et al., 2011).

At present, the accessibility to state-of-the-art technologies has increased the capability to analyze complete genomes, contributing to the massive collection of information and progress of phylogenetic studies. Particularly, complete plastome analyses, some of which are conducted using next-generation sequencing technologies (NGS), have improved the resolution of the latest studies on phylogenetic relationships, generating results with high and better support values (Huang et al., 2017; Edgar, 2004; Kugita et al., 2003). In the chloroplast, there are multiple copies of circular DNA molecules, varying from 120 to 217 kb, and they are characterized by the maintenance of conserved genetic content between divergent individuals, which has allowed the assessment of phylogenetic relationships at different taxonomic levels; however, changes have also been found in the structure and arrangement of the gene order in related plants (Palmer, 1991; Katayama and Ogihara, 1993).

Investigations based on data available from more than 300 complete chloroplast genomes have revealed the presence of DNA rearrangements within the family Poaceae, and approximately 88% of these genomes correspond to the most used families by humankind (Huang et al., 2017). For example, kikuyu is a perennial grass recognized as a fundamental resource for specialized livestock, although it is considered as a weed on golf courses owing to its invasive growth (Miyasaka et al., 2007; Wilen et al., 1995; Marais, 2001). In the present study, to broaden the information and the set of available genomic data, the plastome of *Cenchrus clandestinus* was sequenced, and the gene arrangements and phylogenetic relationships with other species of the family Poaceae and subfamily Panicoideae were examined.

2. Materials and methods

2.1. Vegetative material of Cenchrus clandestinus

Fresh grass leaves was collected from Paysandú farm of Universidad Nacional de Colombia located in Santa Elena, Municipality of Medellín, Department of Antioquia. The procedures for sample preparation, sequencing, and assembly were performed at the National Center for Genomic Sequencing (CNSG) of Universidad de Antioquia. Genomic DNA was extracted using a PowerPlant® Pro DNA Isolation Kit (Qiagen) according to the manufacturer's protocols. DNA concentration were defined using fluorescence measurements with PicoGreen® (Quant-iT PicoGreen dsDNA Assay Kit ref.: p7589), and DNA purity was defined spectrophotometrically by measuring absorbance at 260, 280, and 320 nm.

2.2. Genome sequencing and de novo assembly

One whole-genome shotgun (WGS) library was constructed using Nextera preparation kits and was sequenced in an Illumina MiSeq platform generating paired-end reads of 250. The quality of sequences was refined using trimmomatic and PRINSEQ to identify and cut out low-quality regions (<Q30) and low-complexity sequences (Schmieder and Edwards, 2011; Bolger et al., 2014). After quality trimming, singleton reads and those with less than 70 bases were excluded from the dataset. The sequences (reads) containing chloroplast genome information were identified using the BLASTN algorithm and were assembled de novo with NEWBLER v. 2.9 (Margulies et al., 2005) (runAssembly mode) using the default options. The contigs obtained were ordered and oriented with the *Cenchrus americanus* reference using the ABACAS script (Perl language) together with manual supervision in ARTEMIS.

To validate the de novo assembly results, a comparison was made with *Cenchrus americanus* references with accession numbers KX756179 and KJ490012. A mapping of the reads was conducted using NEWBLER (Newbler runMapping) and was visualized with TABLET (no inconsistencies were observed). The few remaining gaps were validated via PCR and capillary sequencing. Primers were designed flanking the splice zones of contigs (14 primers), and gaps were manually cured in Artemis using the capillary sequencing results.

2.3. Annotation and visualization of the genome

The annotation process was performed using the automatic transfer RATT tool (Sanger). Subsequently, a manual inspection was performed using the Artemis annotator (Rutherford et al., 2000) based on the references of *Cenchrus americanus* with accession numbers KX756179 and KJ490012. All the tRNA genes were predicted using the tRNAscan-SE Search Server (Schattner et al., 2005). The GenomeVx online server was used for genome visualization (Conant and Wolfe, 2008).

2.4. Structural and organizational comparison of the chloroplast genome with related species

For the structural comparison of the complete genome of *Cenchrus clandestinus* with genomes of other Poaceae species, the plastomes of *C. americanus, C. purpureus*, *Setaria viridis,* and *Sorghum bicolor,* which belong specifically to the subfamily Panicoideae, were included (Table 1). The length of protein-coding sequences (CDS) and the gene order (synteny) were determined using the Artemis software. In addition, to identify the start and stop codon patterns, the annotation of plastomes deposited in the GenBank database was inspected in ARTEMIS and MEGA (Kumar et al., 2016).

2.5. Phylogenetic analysis

To determine the phylogenetic relationship between species of the subfamily Panicoideae, a Bayesian tree was constructed from 81 protein-coding genes found in the chloroplast using MrBayes (MB) v. 3.2 (Ronquist et al., 2012). For the analysis, trans-splicing copies of the rps12 gene were not included due to possible inconsistencies in the information. Two independent Markov chain Monte Carlo iterations were included within the parameters sampled every 1000 generations for 20 million generations, and 25% of the sampled generations were discarded as *burn-in*, keeping the other parameters by default. Convergence was based on a potential scale reduction factor (PSRF) equal to 1.0 and standard deviation of divided frequencies close to 0.0.

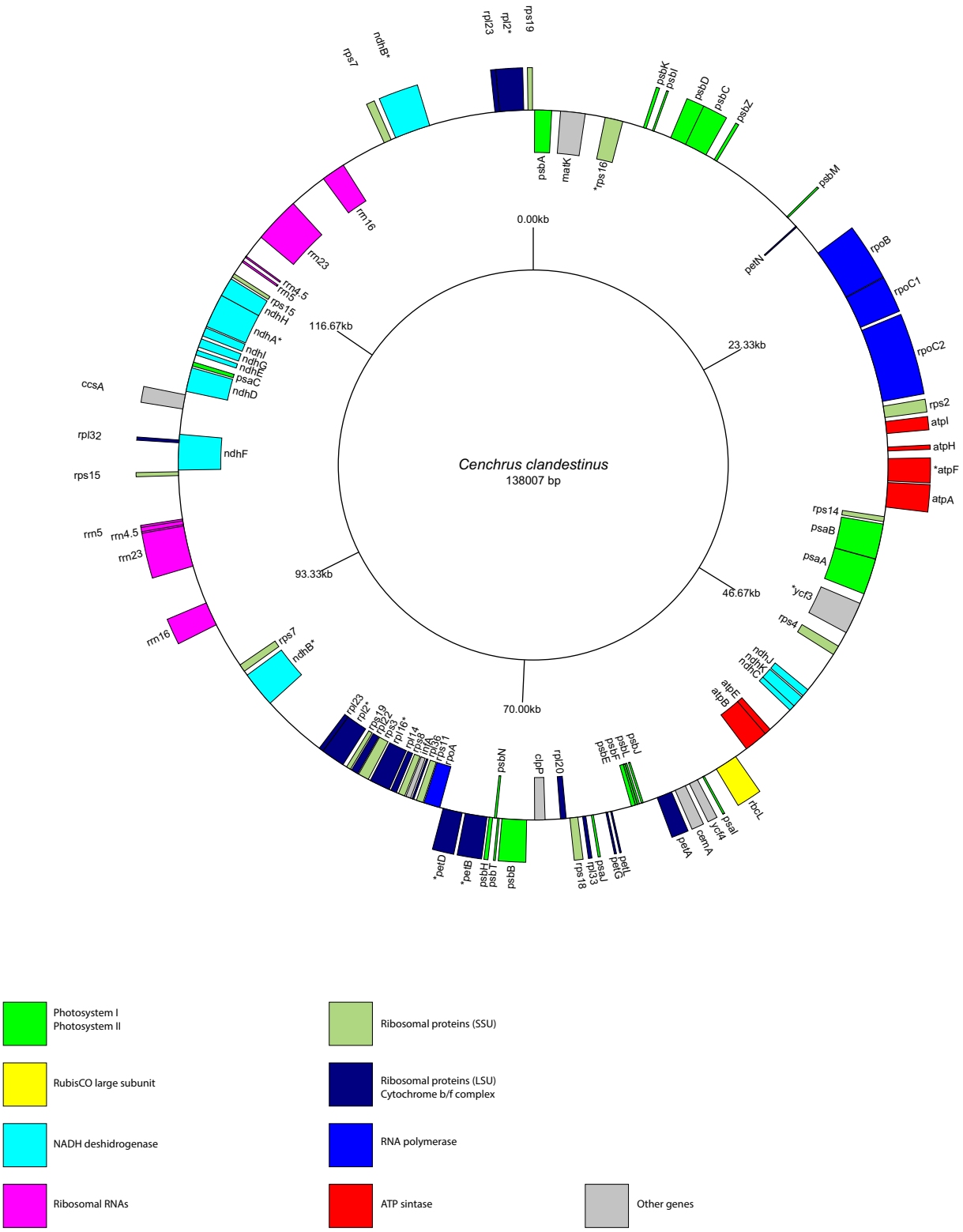


Fig. 1. Chloroplast genome map of *Cenchrus clandestinus* genes. The genes (protein-coding sequences and ribosomal RNA genes) are displayed on the outside of the large circle are transcribed clockwise, whereas those shown on the inside are transcribed counterclockwise. The asterisk represents genes with introns. The colors represent the functional group of the genes. The rps12 gene (transplicing) and transfer RNA genes are not shown in the figure.

Table 2
Group of genes coding plastome proteins of *Cenchrus clandestinus*.

Group of genes	Coding sequences
Photosystem I	*psaA, psaB, psaC,psaI, psaJ, ycf3**, ycf4*
Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI,psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
RuBisCO large subunit	*rbcL*
NADH dehydrogenase	*ndhA*, ndhB**(x2), *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI*
Cytochrome b/f complex	*petA, petB*, petD*, petG, petL,petN*
Small subunit of Ribosome	*rps2, rps3, rps4, rps7*(x2), *rps8, rps11, transplicing: rps12*** (×2), *rps14, rps15*(x2), *rps16*, rps18, rps19* (x2)
Large subunit of ribosome	*rpl2**(x2), *rpl14, rpl16*, rpl20, rpl22, rpl23*(x2), *rpl32, rpl33, rpl36*
RNA polymerase subunits	*rpoA, rpoB, rpoC1, rpoC2.*
Other genes	*matK, ccsA, cemA, clpP, infA*

Genes with two copies are indicated by (×2) after the name. One and two asterisks indicate genes with one or two introns, respectively.

Table 3
Position and anticodons of the tRNAs of the *Cenchrus clandestinus* chloroplast genome.

tRNA #	tRNA Begin	tRNA End	tRNA Type	Anti-codon
1	12,219	12,289	Gly	GCC
2	15,197	15,268	Thr	GGT
3	15,782	15,854	Glu	TTC
4	15,916	15,999	Tyr	GTA
5	16,347	16,420	Asp	GTC
6	45,909	45,995	Ser	GGA
7	48,927	48,999	Phe	GAA
8	52,763	52,835	Met	CAT
9	81,303	81,376	His	GTG
10	92,956	93,027	Val	GAC
11	100,842	100,919	Arg	ACG
12	106,829	106,908	Leu	TAG
13	117,610	117,681	Asn	GTT
14	132,070	132,150	Leu	CAA
15	135,389	135,461	Met	CAT
16	137,551	137,478	His	GTG
17	125,898	125,827	Val	GAC
18	118,012	117,935	Arg	ACG
19	101,244	101,173	Asn	GTT
20	86,784	86,704	Leu	CAA
21	83,465	83,393	Met	CAT
22	64,877	64,804	Pro	TGG
23	64,681	64,608	Trp	CCA
24	47,248	47,176	Thr	TGT
25	37,320	37,249	Arg	TCT
26	19,277	19,207	Cys	GCA
27	12,631	12,558	Met	CAT
28	11,404	11,317	Ser	TGA
29	7835	7748	Ser	GCT
30	6559	6488	Gln	TTG

3. Results

3.1. Annotation and composition of the C. clandestinus chloroplast genome

The annotation results indicated that the chloroplast genome of *C. clandestinus* (GenBank accession MT646354) is a molecule with a length of 138,007 bp that presents an average GC content of 38.63%. It is constituted by 121 predicted genes, which comprise 83 genes coding for proteins, 8 ribosomal RNA genes, and 30 transfer RNA genes (Fig. 1).

3.2. Protein-encoding genes

A number of plastome coding regions corresponded to a group of genes important for gene expression and is necessary for protein transcription or synthesis. These genes include *rpoA, rpoB, rpoC1,* and *rpoC2,* which encode α-, β, β'-, and β"-subunits of RNA polymerase (PEP, plastid-encoded plastid RNA polymerase), rpl genes encoding ribosomal proteins of the 50S large subunit (rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, and rpl36), and rps genes for the 30S small subunit (such *as rps2,rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps16,rps15, rps18,* and *rps19*). Two copies of the *rps19, rpl23, rps7,* and *rps15* genes were present in the genome (Table 2).

A large group of genes involved in the photosynthesis process was detected. These regions encode 11 subunits of NADH dehydrogenase complexes (*ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ,* and *ndhK genes*), ATP synthase (*atpA, atpB, atpE, atpF, atpH,* and *atpI* coding sequences), cytochrome *b/f* (petA, petB, petD, petG, petL, and petN), and the *rbcL* gene of the RuBisCO large subunit. A total of 15 *psb* genes encoding photosystem II subunits and the *psaA, psaB, psaC, psaI,* and *psaJ* coding sequences of photosystem I were also annotated. In addition, genes related to other functions were identified, such as *matK, clpP, cemA,* and *ccsA*.

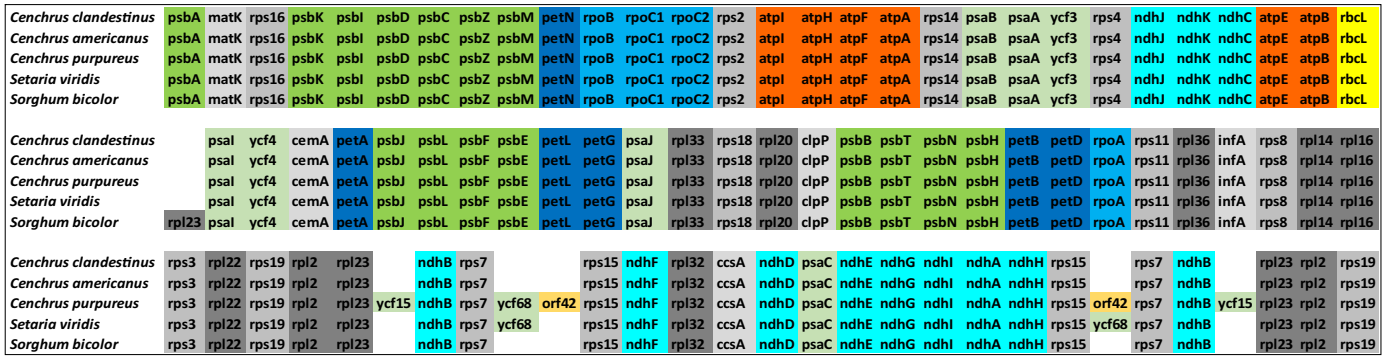


Fig. 2. Gene order in five species related to *Cenchrus clandestinus* within the subfamily Panicoideae. Colors represent groups of genes classified by function: photosystem I (*psa and ycf*); photosystem II (*psb*); cytochrome (*pet*) complex; subunits of RNA polymerase (*rpo*); small ribosome subunit (*rps*); large ribosome subunit (*rpl*); ATP synthase (*atp*); NADH dehydrogenase (*ndh*); RuBisCO (*rbcL*); and other genes (*matK, ccsA, cemA, clpP,* and *infA).*

Table 4
Length of the protein-coding sequences (CDS) of plastomes in five species of the subfamily Panicoideae. Differences in gene length between species with respect to kikuyu grass are highlighted in bold and underlined.

Gen	psbA	matK	rps16*	psbK	psbI	psbD	psbC	psbZ	psbM	petN	rpoB	rpoC1	rpoC2	rps2
Cenchrus clandestinus	1062	1542	**1118**	186	111	1062	1422	189	105	90	3228	2052	4587	711
Cenchrus americanus	1062	**1536**	**1105**	186	111	1062	1422	189	105	90	3228	2052	**4407**	711
Cenchrus purpureus	1062	**1536**	**1124**	186	111	1062	1422	189	105	**96**	3228	2052	**4407**	711
Setaria viridis	1062	1542	**1095**	186	111	1062	1422	189	105	90	3228	2052	**4614**	711
Sorghum bicolor	1062	**1548**	**1098**	186	111	1062	1422	189	105	90	3228	2052	**4563**	711

	atpB	rbcL	psaI	ycf4	cemA	petA	psbJ	psbL	psbF	psbE	petL	petG	psaJ	rpl33
Cenchrus clandestinus	1497	1428	111	558	693	963	123	117	120	252	96	114	129	201
Cenchrus americanus	1497	1428	111	558	693	963	123	117	120	252	96	114	129	201
Cenchrus purpureus	1497	1428	111	558	693	963	123	117	120	252	96	114	129	201
Setaria viridis	1497	1428	111	558	693	963	123	117	120	252	96	114	129	201
Sorghum bicolor	1497	**1431**	111	558	693	963	123	117	120	252	96	114	129	201

Gen	atpI	atpH	atpF*	atpA	rps14	psaB	psaA	ycf3*	rps4	ndhJ	ndhK	ndhC	atpE	rps18
Cenchrus clandestinus	744	246	1390	1524	312	2205	2253	1999	606	480	750	363	414	492
Cenchrus americanus	744	246	**1373**	1524	312	2205	2253	**1984**	606	480	750	363	414	492
Cenchrus purpureus	744	246	**1374**	1524	312	2205	2253	**1984**	606	480	750	363	414	492
Setaria viridis	744	246	**1371**	1524	312	2205	2253	**1996**	606	480	**744**	363	414	492
Sorghum bicolor	744	246	**1391**	1524	312	2205	2253	**2007**	606	480	**684**	363	414	492

	rpl20	clpP	psbB	psbT	psbN	psbH	petB*	petD*	rpoA	rps11	rpl36	infA	rps8	rpl14
Cenchrus clandestinus	360	651	1527	102	132	222	**1372**	**1217**	1020	432	114	306	411	372
Cenchrus americanus	360	651	1527	102	132	222	**1400**	**1210**	1020	432	114	306	411	372
Cenchrus purpureus	360	651	1527	**111**	132	222	**1410**	**525****	1020	432	114	306	411	372
Setaria viridis	360	651	1527	102	132	222	**1407**	**1223**	1020	432	114	**324**	411	372
Sorghum bicolor	360	651	1527	102	132	222	**1406**	**1223**	1020	432	114	**324**	411	372

	rpl16*	rps3	rpl22	rps19	rpl2*	rpl23	ndhB*	rps7	rps15	ndhF	rpl32	ccsA	ndhD
Cenchrus clandestinus	**1298**	675	450	282	1487	282	2243	471	237	2217	180	891	1503
Cenchrus americanus	**1330**	675	450	282	**1485**	282	2243	471	237	2217	180	**981**	1503
Cenchrus purpureus	**363****	675	450	282	**1482**	282	2243	471	237	2217	180	**981**	1503
Setaria viridis	**1333**	675	450	282	**1485**	282	2243	471	237	2217	180	**969**	1503
Sorghum bicolor	**1482**	675	447	282	**1485**	282	**2237**	471	237	2217	180	**966**	1503

	psaC	ndhE	ndhG	ndhI	ndhA	ndhH	rps15	rps7	rpl23	rps19
Cenchrus clandestinus	246	306	531	543	2132	1182	237	471	282	282
Cenchrus americanus	246	306	531	543	2106	1182	237	471	282	282
Cenchrus purpureus	246	306	531	543	2100	1182	237	471	282	282
Setaria viridis	246	306	531	543	2093	1182	237	471	282	282
Sorghum bicolor	246	306	531	543	2111	1182	237	471	282	282

* Genes encoding proteins with intron.
** Intron loss in protein-coding genes.

Out of the 121 genes identified in the plastome, 12 contained introns, corresponding to genes that encode proteins: seven presented a single copy in the genome (*rps16, atpF, ycf3, petB, petD, rpl16,* and *ndhA*), and three genes had two copies (rps12, rpl2, and ndhB). In all these genes, a single intron was identified, except for the *ycf3* gene and *rps12* gene (transplicing), which had two introns (Table 2).

3.3. Ribosomal RNA genes and transfer RNA genes

Four genes were identified for ribosomal RNA, namely, *rrn5, rrn4.5, rrn23,* and *rrn16,* each present in two copies within the genome. On the other hand, using the tRNAscan-SE search server, 30 tRNAs were predicted in the chloroplast (Table 3).

3.4. Structure and comparison of the plastome's genomic organization

The results of this study confirm the synteny of *C. clandestinus'* plastome to that of *C. americanus* (Fig. 2). In other species, in addition to synteny, differences in the type and number of gene copies were observed, which shows that not all species are composed of the same 83 coding sequences. Specifically, the plastome of *C. purpureus* contains three additional coding sequences, namely, *ycf15* and *orf42* (present in two copies) and *ycf68* (present in a single copy). *S. viridis* and *S. bicolor* contain two additional sequences of the *ycf68* gene and an additional copy of the *rpl23* gene, respectively.

The lengths of the coding sequences were similar in most of the species examined, except for the *rbcL* and *ndhB* genes of *S. bicolor* and the *petN* and *psbT* sequences of *C. purpureus.* The *infA* and *ndhK* sequences were similar only for the genus *Cenchrus,* whereas *rps16, ycf3, rpoC2, petB, petD, rpl16,* and *ccsA* were different for all species. The *matK* gene

Table 5
Differences in terms of start and stop codons between the protein-coding sequences of plastomes in five *Poaceae* species of the subfamily Panicoideae.

Start codons	ndhA	petN	psbT	rpl2	rpl2	rpl32
Cenchrus clandestinus	ATG	ATG	ATG	ACG	ACG	**AGC**
Cenchrus americanus	ATG	ATG	ATG	ACG	ACG	ATG
Cenchrus purpureus	**ATA**	**ATT**	**ATC**	**GCG**	**GCG**	ATG
Setaria viridis	**ATA**	ATG	ATG	ACG	ACG	ATG
Sorghum bicolor	ATG	ATG	ATG	ACG	ACG	ATG

Stop codons	ccsA	ndhH	psbT	rpl32	rps2	ycf4
Cenchrus clandestinus	TGA	TGA	TGA	**TTC**	**TAA**	TGA
Cenchrus americanus	TGA	TGA	TGA	TAA	TGA	TGA
Cenchrus purpureus	TGA	TGA	**TAA**	TAA	TGA	TGA
Setaria viridis	TGA	**TAG**	TGA	TAA	TGA	TGA
Sorghum bicolor	**TAA**	TGA	TGA	TAA	TGA	**TAG**

Different start and stop codons for the evaluated genes are highlighted in bold.

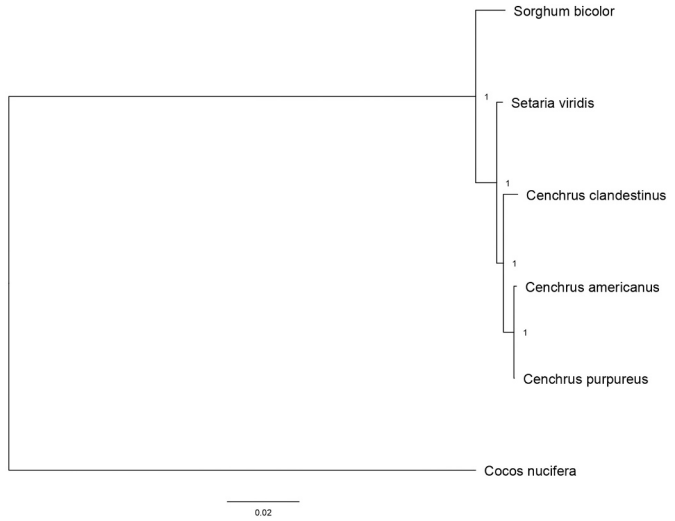


Fig. 3. Bayesian phylogenetic tree based on 81 chloroplast genes coding proteins within the Poaceae family members. *Cocos nucifera* was used as an outgroup.

sequences were only similar for *C. clandestinus* and *S. viridis*. Most sequences containing introns were different in length in all species (Table 4).

For all species, the most frequent start codon was ATG (76 CDS on average), followed by GTG (present in the *rps19* gene in all genomes), ACG (present in the *rpl2* gene in all genomes, except for *C. purpureus)*, AGC (present in *rpl32* only in *C. clandestinus)*, ATA (present in the *ndhA* gene of *C. purpureus* and *S. viridis)*, and finally ATC, ATT, and GCG (present in the psbT, petN, and rpl2 sequences of *C. purpureus*, respectively). The TAA and TGA stop codons were the most frequent, whereas TTC was only present in the *rpl32* gene of *C. clandestinus*. Table 5 summarizes the different start and stop codons analyzed for the Panicoideae family genes.

3.5. Phylogeny

The phylogenetic analysis was based on 81 genes coding for plastome proteins, and it revealed that species belonging to the family Poaceae were organized into three groups. As expected, one group was formed by species of the same genus, namely, kikuyu grass (*Cenchrus clandestinus*), pearl millet (*C. americanus*), and elephant grass (*C. purpureus*), whereas the second and third groups included the remaining two family members, namely, green bristlegrass (*Setaria viridis*) and sorghum (*Sorghum bicolor*), respectively. General topology is in line with the classification of these genera within the family Poaceae (Fig. 3).

4. Discussion

The present study characterized and annotated the complete plastome of kikuyu grass, *Cenchrus clandestinus,* as an informative contribution for future studies potentially investigating the evolution of plant genomes and, specifically, aiming to elucidate the phylogenetic relationships within the family Poaceae. The results would also be particularly useful as the delimitation between some genera of the subfamily *Panicoideae*, such as *Pennisetum* and *Cenchrus,* has been controversial, leading researchers to even propose the modification of the nomenclature of some species between these two genera (Donadío et al., 2009; Chemisquy et al., 2010). Here, to study the effect of mutational differences, phylogenetic relationships were explored in four species of the same subfamily based on 81 protein-coding genes present in the chloroplast.

The plastome is generally characterized by an integral conservation of its structure; however, depending on the species, its size can vary between 102 and 218 kb (Daniell et al., 2016). The chloroplast genome of *C. clandestinus* (138,007 bp) was slightly different from that reported in other species of the same genus, such as *C. americanus* (138,172 bp) and *C. purpureus* (138,199 bp). The genome of other groups of monocotyledonous plants belonging to the family Poaceae was different, such as in the case of *Cynodon dactylon* (134,297 bp) and *Pharus latifolius* (142,077pb). On the other hand, among dicotyledonous plants, species such as *Amborella trichopoda* (162,686pb) have a larger genome size (Huang et al., 2017; Raubeson and Jansen, 2005; Jansen et al., 2007; Saski et al., 2007). Generally, the family Poaceae has smaller genomes compared with the great majority of angiosperms, mainly due to the presence of the *ycf1* and *ycf2* genes, which have longer open reading frames. In addition, gene relocation events from the chloroplast to the nucleus determine a selective pressure on the size of the genome, as in the case of lost chloroplast genes (*accD*), which are already encoded by nuclear genes (ACCase) (Huang et al., 2017). The GC content detected in this study was similar to that found in *C. americanus* (39.1%) and *C. purpureus* (39.37%).

The arrangement of genes was very similar between *C. clandestinus* and *C. americanus,* whereas in the other genomes, they were compared with additional coding sequences, identified as *ycf15*, *orf42,* and *ycf68*. On the other hand, in *Sorghum bicolor,* an additional copy of the *rpl23* gene was detected. In most terrestrial plants, the organization of genomes is highly conserved; however, lineages in which the order and content of genes may vary have been identified (Raubeson and Jansen, 2005). These additional genes have also been reported in Bermuda grass (*Cynodon dactylon*) as pseudogenes, namely, *ycf15*, which had an internal stop codon; *cf68,* which degraded to residual fragments containing multiple internal stop codons; and *rpl23,* as a fragment without start or stop codons (Huang et al., 2017).

Most of the coding sequence lengths were similar, although differences were observed in most of the genes that contained introns. Introns in the chloroplast genomes of terrestrial plants are generally conserved, but the loss of protein-coding genes has been reported in some species (Jansen et al., 2007; Saski et al., 2007; Wu et al., 2009). The majority of these losses have been registered in different specific groups of plants, one of them being the species belonging to the monocotyledons (Poaceae); genes such as *clpP*, *atpF*, and *rpoC2*, which encode a protease, an ATP synthase, and an RNA polymerase, respectively, are some examples (Daniell et al., 2016; Jansen et al., 2007). On the other hand, some authors have reported different starting positions based on the alignment of several genes in the subfamily Panicoideae, for example, the init codon of the *infA* gene can change, appearing in 18 bp (access repetition point for mutation that can be lost) and generating two different positions due to an elimination event, which leads to the shortening of the *infA* gene in the species, maintaining the coding sequence (Burke et al.,

2016). This molecular event has been described in *C. americanus*, and our results indicate that it occurs also in *C. clandestinus*.

Most coding sequences generally share the start codon (usually ATG) and the stop codons (TAA, TGA), which are the most frequent in many other eukaryotic organisms (Sun et al., 2005; Chawla, 2002). On the other hand, in the present study, less-frequent codons were also present, and this normally occurs due to processes related to the editing system in highly specific chloroplast sites, which give rise to certain codon transitions. These editing processes may also involve the creation of new reading frames through the introduction of start codons (Aro and Bertil, 2001). Similarly, in this study, the tree based on 81 protein-coding genes identifying the phylogenetic relationship between the Poaceae family members exhibited consistency with the general topology of the usual classification of the evaluated genera. The results approached the approximations that several authors have made when exploring the phylogenetic relationships within this family and the subfamily Panicoideae (Chemisquy et al., 2010; Duvall et al., 2001; Burke et al., 2016; Huang et al., 2016; Wu and Zhou, 2019), identifying three clades belonging to the *Cenchrus* genus (*C. clandestinus, C. americanus,* and *C. purpureus*) and two others that corresponded to *Sorghum bicolor and Setaria viridis.*

This work presents the first report of the chloroplast genome of kikuyu grass (*Cenchrus clandestinus*) using an Illumina MiSeq platform. The kikuyu grass genome (138,007 bp) was analyzed and compared with four genomes of related species, showing in general a conserved organization of plastomes. The content, order, and structure of the genes were found to be similar, with some exceptions observed in *C. purpureus* and *Setaria viridis.* The chloroplast genome of *C. clandestinus* had 121 predicted genes, which comprised 83 protein-coding genes, 8 ribosomal RNA genes, and 30 transfer RNA genes. Phylogenetic analysis confirmed the position of kikuyu within the genus *Cenchrus* and congruent relationships among other species of the subfamily Panicoideae. This study provides important molecular information to complement genomic resources and improve the resolution of future studies investigating phylogenetic relationships within the family Poaceae.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

For financial support, we gratefully thank Universidad Nacional de Colombia Sede Medellín.

References

Aliscioni, S., Hester, L., Besnard, G., Pascal, C., Travis, J., Melvin, R., Edwards, J., Giussani, L., Hasenstab-lehman, K., Hilu, K., Hodkinson, T., Ingram, A., Kellogg, A., Morrone, M.S.O., Osborne, C., Salamin, N., Hanno, S., Spriggs, E., Smith, S., Zuloaga, F., 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C 4 origins. New Phytol. 193, 304–312. https://doi.org/10.1111/j.1469-8137.2011.03972.x.

Aro, E.M., Bertil, A., 2001. Advances in Photosynthesis and Respiration. Kluwer Academic Publishers.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Burke, S.V., Wysocki, W.P., Zuloaga, F.O., Craine, J.M., Pires, J.C., Edger, P.P., Mayfield-Jones, D., Clark, L.G., Kelchner, S.A., Duvall, M.R., 2016. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). BMC Plant Biol. 16, 1–11. https://doi.org/10.1186/s12870-016-0823-3.

Chawla, H.S., 2002. Introduction to Plant Biotechnology, Second. Science Publishers, Inc, Pantnagar,India.

Chemisquy, M.A., Giussani, L.M., Scataglini, M.A., Kellogg, E.A., Morrone, O., 2010. Phylogenetic studies favour the unification of *Pennisetum*, *Cenchrus* and *Odontelytrum* (Poaceae): a combined nuclear, plastid and morphological analysis, and nomenclatural combinations in Cenchrus. Ann. Bot. 106, 107–130. https://doi.org/10.1093/aob/mcq090.

Conant, G.C., Wolfe, K.H., 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics. 24, 861–862. https://doi.org/10.1093/bioinformatics/btm598.

Daniell, H., Lin, C.S., Yu, M., Chang, W.J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17, 1–29. https://doi.org/10.1186/s13059-016-1004-2.

Donadío, S., Giussani, L.M., Kellogg, E.A., Zuolaga, F.O., Morrone, O., 2009. A preliminary molecular phylogeny of *Pennisetum* and *Cenchrus* (Poaceae-Paniceae) based on the tmL-F, rpll6 chloroplast markers. Taxon. 58, 392–404. https://doi.org/10.1002/tax.582007.

Duvall, R., Noll, D., Minn, H., 2001. Phylogenetics of Paniceae (Poaceae). Am. J. Bot. 88, 1988–1992.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.

Fukumoto, G.K., Lee, C.N., 2003. Kikuyu grass for forage. https://www.ctahr.hawaii.edu/oc/freepubs/pdf/LM-5.pdf.

García, S., Islam, M., Clark, C.E., Martin, P., 2014. Kikuyu-based pasture for dairy production: a review. Crop. Pasture Sci. 65, 787–797. https://doi.org/10.1071/CP13414.

Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. PLoS One 6. https://doi.org/10.1371/journal.pone.0019254.

Huang, P., Shyu, C., Coelho, C.P., Cao, Y., Brutnell, T.P., 2016. *Setaria viridis* as a model system to advance millet genetics and genomics, front. Plant Sci. 7, 1–9. https://doi.org/10.3389/fpls.2016.01781.

Huang, Y.Y., Cho, S.T., Haryono, M., Kuo, C.H., 2017. Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. PLoS One 12. https://doi.org/10.1371/journal.pone.0179055.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. U. S. A. 104, 19369–19374. https://doi.org/10.1073/pnas.0709121104.

Katayama, H., Ogihara, Y., 1993. Structural alterations of the chloroplast genome found in grasses are not common in monocots. Curr. Genet. 23, 160–165. https://doi.org/10.1007/BF00352016.

Kellogg, E.A., 2015. Flowering Plants. Monocots: Poaceae, Thirteenth, New York.

Kugita, M., Kaneko, A., Yamamoto, Y., Takeya, Y., Matsumoto, T., Yoshinaga, K., 2003. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. Nucleic Acids Res. 31, 716–721. https://doi.org/10.1093/nar/gkg155.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054.

Marais, J., 2001. Factors affecting the nutritive value of kikuyu grass (*Pennisetum clandestinum*)-a review, Trop. Grasslands. 35, 65–84.

Margulies, M., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 437, 376–380. https://doi.org/10.1038/nature03959.

Miyasaka, S.C., Hansen, J.D., Fukumoto, G.K., 2007. Resistance to yellow sugarcane aphid: screening kikuyu and other grasses. Crop Prot. 26, 503–510. https://doi.org/10.1016/j.cropro.2006.04.022.

Palmer, J., 1991. Plastid chromosomes: Structure and evolution. In: Bogorad, L., Vasil, I. (Eds.), Mol. Biol. Plast., Cell Cult. Somat. Cell Genet. Plants, vol. 7A. Academic Press, San Diego,California, pp. 5–53.

Raubeson, L., Jansen, R., 2005. Chloroplast genomes of plants. In: Henry, R.J., Wallingford (Eds.), Plant Divers. Evol. Genotypic Phenotypic Var. High. Plants, CABI Publishing, pp. 45–68.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. https://doi.org/10.1093/sysbio/sys029.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation Kim, Bioinforma. Appl. Note. 16, 944–945.

Saski, C., Lee, S.B., Fjellheim, S., Guda, C., Jansen, R.K., Luo, H., Tomkins, J., Rognli, O. A., Daniell, H., Clarke, J.L., 2007. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. Theor. Appl. Genet. 115, 571–590. https://doi.org/10.1007/s00122-007-0567-4.

Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, 686–689. https://doi.org/10.1093/nar/gki366.

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 27, 863–864. https://doi.org/10.1093/bioinformatics/btr026.

Sun, J., Chen, M., Xu, J., Luo, J., 2005. Relationships among stop codon usage bias, its context, isochores, and gene expression level in various eukaryotes. J. Mol. Evol. 61, 437–444. https://doi.org/10.1007/s00239-004-0277-3.

Wilen, C.A., Holt, J.S., Ellstrand, N.C., Shaw, R.G., Science, W., Wilen, C.A., Holt, J.S., Ellstrand, N.C., Shaw, R.G., . Genotypic Diversity of Kikuyugrass (*Pennisetum clandestinum*) Populations in California Published by : Weed Science Society of America and Allen Press Stable. URL : http://www.jstor.org/stable/4045485. REFERENCES Linked references are available on JSTOR fo, Weed Sci, 43, pp. 209–214.

Wu, Y., Zhou, H., 2019. The complete chloroplast genome sequence of *Cenchrus purpureus*. Mitochond. DNA Part B Resour. 4, 51–52. https://doi.org/10.1080/23802359.2018.1536451.

Wu, F.H., Kan, D.P., Lee, S.B., Daniell, H., Lee, Y.W., Lin, C.C., Lin, N.S., Lin, C.S., 2009. Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. Tree Physiol. 29, 847–856. https://doi.org/10.1093/treephys/tpp015.