Plant Genetic Resources: Characterization and Utilization

cambridge.org/pgr

Molecular identification and phylogenetics of local pearl millet cultivars using internal-transcribed spacers of nuclear ribosomal DNA

Zainab M. Almutairi

Department of Biology, College of Science and Humanities, Prince Sattam bin Abdulaziz University, P.O. Box 83, Al-Kharj 11942, Saudi Arabia

Research Article

Cite this article: Almutairi ZM (2021). Molecular identification and phylogenetics of local pearl millet cultivars using internal-transcribed spacers of nuclear ribosomal DNA. *Plant Genetic Resources: Characterization and Utilization* **19**, 339–346. https://doi.org/10.1017/S1479262121000393

Received: 17 February 2021
Revised: 2 August 2021
Accepted: 3 August 2021
First published online: 1 September 2021

Key words:
5.8S rRNA; ITSs; *Pennisetum glaucum*; phylogenetics; secondary structure

Author for correspondence:
Zainab M. Almutairi,
E-mail: z.almutairi@psau.edu.sa

Abstract

Local cultivars of pearl millet in Saudi Arabia are known to tolerate extreme heat and drought stress. In the current study, the sequences of internal-transcribed spacers (ITSs) of six pearl millet cultivars were sequenced and analysed to investigate the genetic diversity among the local cultivars. The nucleotide polymorphism, secondary structures and phylogenetics were analysed for ITS sequences of the six local cultivars. The obtained sequences were 772–774 base pairs (bp) in length, including complete sequences of the ITS1–5.8S–ITS2 region and partial sequences of 18S and 26S rRNA. The nucleotide diversity among cultivars was higher in ITS2 sequences than ITS1 sequences. The ITS2 had four variable nucleotide sites in three native cultivars, whereas the ITS1 contained one base insertion. The secondary structures of the ITS1 and 5.8S region were highly conserved among the six cultivars and contained some motifs that are conserved across Viridiplantae. However, the ITS2 secondary structure for the two native cultivars, Sayah and Jazan, was distinct from the other cultivars, which confirms the applicability of the ITS2 sequence in distinguishing between genetically close taxa. Additionally, the phylogenetic analysis of the six investigated cultivars and 31 pearl millet accessions from the NCBI database showed close relationships between the local accessions and NCBI accessions from India and France. However, the local cultivar Sayah appeared to be distinct from the other cultivars in the phylogenetic trees. This study provides insights into the polymorphism within local pearl millet cultivars which is important for the identification and conservation of these cultivars.

Introduction

Pearl millet *Pennisetum glaucum* subsp. *monodii* (Maire) Brunken is an important crop widely cultivated in different regions, primarily in Asia and Africa, as a multipurpose cereal (ICRISAT, 1996; Havilah, 2017). In Saudi Arabia, pearl millet is grown in the western and southwestern regions during different seasons with various temperatures and different water needs. In the western region of Saudi Arabia, the local cultivars tolerate extreme heat where the day temperature ranges from 33 to 46°C in the summer season. In the southwestern region, where the temperature is more moderate (22–32°C), other pearl millet cultivars are grown under rainfed or limited irrigation. In these diverse environments, there are several local cultivars of pearl millet, but the number of cultivars remains undetermined (Bidinger and Hash, 2003). To date, limited efforts have been made to evaluate genetic diversity within local Saudi pearl millet cultivars. In one instance, 12 cultivars of pearl millet were collected from the Jazan region in southwestern Saudi Arabia and registered in the gene bank of the King Abdulaziz City for Science and Technology (Al-Turki *et al.*, 2019).

Previous studies have investigated the genetic diversity of millet cultivars from African and Asian countries using restriction fragment length polymorphism markers (Bhattacharjee *et al.*, 2002) and simple sequence repeat markers (McBenedict *et al.*, 2016). The efficiency of molecular markers in the classification of organisms has been reported in many studies (Kress *et al.*, 2005). The internal-transcribed spacer (ITS) regions of the nuclear ribosomal DNA (nrDNA) are useful markers to explore the evolutionary relationships among plants (Venkateswarlu and Nazar, 1991; Hemleben, 1993; Baldwin *et al.*, 1995). The ITS1 and ITS2 sequences are non-functional parts of the nrDNA located within the 18S, 5.8S and 26S rRNA genes; they are transcribed as a single-transcription unit pre-rRNA molecule and then removed and degraded during rRNA processing. The sequence and structural elements of ITSs are critical to produce mature rRNAs (Veldman *et al.*, 1981; Van Nues *et al.*, 1994). Moreover, the secondary structures of ITSs are more conserved than their nucleotide sequences, which makes them a useful tool to elucidate the evolutionary relationships among organisms (Coleman *et al.*, 1998; Schultz *et al.*, 2005).

Table 1. Name, collection location and registration numbers of the six pearl millet cultivars

Cultivars	Collection location (province, place)	Registration no.
Baydhan	Baydhan Valley, western region, Saudi Arabia	1316
Sayah	Sayah valley, western region, Saudi Arabia	1317
Jazan	Alasamlah, Jazan, southwestern region, Saudi Arabia	1318
Tihamah	Tihamah region, southwestern region, Saudi Arabia	1320
Yemeni	Mahayil, southwestern region, Saudi Arabia	1321
Indian	Qanuna Valley, Al Ardiyat, southwestern region, Saudi Arabia	1323

Yemeni and Indian cultivars were originally from Yemen and India, respectively, and currently cultivated by local farmers in Saudi Arabia. The other four cultivars are natives. The registration no. indicates the numbers of investigated cultivars in the Centre of Genetic Resource in the Ministry of Environment, Water and Agriculture in Saudi Arabia. The seeds of the investigated cultivars are available for researchers, farmers and breeders under International Treaty on Plant Genetic Resources for Food and Agriculture.

The nrDNA of higher plants is organized in blocks in one or more chromosomal regions in the genome (Buckler *et al.*, 1997). Within the family Poaceae, ITS sequences have been used to investigate the genetic relationships among 10 monocot species (Hsiao *et al.*, 1994). Moreover, phylogenetic relationships within the genus *Pennisetum* have been assessed using ITSs (Martel *et al.*, 2004).

Molecular characterization of local millets can help in the identification and conservation of local cultivars that have adapted to extreme environments. In the current study, we evaluated the genetic diversity among six local cultivars of the pearl millet (*P. glaucum* subsp. *monodii*) that are cultivated in different locations in Saudi Arabia based on ITS1 and ITS2 sequences of nrDNA. These cultivars are able to grow in various environments, from moderate temperature to extreme heat and from limited irrigation to rainfed conditions. Local cultivars constitute essential resources of genes for crop improvement programmes, and the evaluation of genetic diversity is important for their conservation.

Methods

Plant material, DNA extraction and primer design

Six local pearl millet cultivars grown in the southwestern and western regions of Saudi Arabia were collected from local farmers and named according to their origins (Table 1). Four cultivars were native pearl millets, and two were originally from Yemen and India. The collected seeds of the six cultivars were transferred to the Center of Genetic Resource in the Ministry of Environment, Water, and Agriculture in Saudi Arabia and given the registration numbers shown in Table 1. The seeds of the investigated cultivars are available for researchers, farmers and breeders under International Treaty on Plant Genetic Resources for Food and Agriculture. The rights of the farmers of the collected seeds are also guaranteed by this treaty. To isolate the nrDNA, genomic DNA was extracted from 5-day-old pearl millet seedlings leaves using Plant DNAzol reagent (Life Technologies). The isolated DNA samples were quantified by using a Qubit fluorometer and diluted to 100 ng/µl. The ITS-specific primers were designed based on the genomic region that includes the nrDNA. The *P. glaucum* subsp. *monodii* nrDNA sequences were retrieved from the NCBI GenBank. The *P. glaucum* ITS sequence (NCBI accession no. FJ766182.1) was used as the query sequence against pearl millet chromosome sequences to find similar genomic regions. Paralogous nrDNA sequences were found on chromosomes 3, 4 and 5; the highest similarity (90%) was observed in nrDNA on chromosome 5 (accession no. CM007986.2). Primer pairs were designed based on the genomic sequence from nucleotide 72922500 to 72924000 on chromosome 5, which spans from 18S to 26S rDNA.

DNA amplification, sequencing and sequence analysis

The ITS sequences were amplified using five pairs of primers to identify the most suitable primer pair to produce a single-polymerase chain reaction (PCR) product for all cultivars. Of all tested primers, the primer pair: forward: 5′-CCTGCCCTTTGTACACACCG-3′, reverse: 5′-ACGCCTCTCCAGACTAC-3′ successfully amplified the entire ITS region. PCR was performed using 2× Phusion Master Mix with a high-fidelity buffer (Thermo Fisher Scientific), 0.4 µM of each primer and 25 ng of DNA template in a total volume of PCR reaction of 50 µl. The Taq polymerase in 1× Phusion Master Mix is 1 unit in the total volume of the reaction. PCR conditions were as follows: a pre-denaturing step of 5 min at 95°C; 30 cycles of 1 min at 95°C, 30 s at 56°C and 1.5 min at 72°C; and a final extension for 7 min at 72°C. The PCR products obtained were sequenced by Sanger's method (Sanger *et al.*, 1977) in both directions, with a dye terminator DNA sequencing kit (Applied Biosystems) on an ABI genetic analyzer 3730xi (Applied Biosystems). Sequencing was conducted in the Central Laboratory for Science and Medical Studies at King Saud University. The sequence of the first amplicon was subjected to a BLAST search in NCBI to confirm that the amplified region included the whole ITS region. Consensus sequences were obtained from at least three replicates of each cultivar. Sequencing results were aligned and assembled by UniGene software (Sayers *et al.*, 2019), and the consensus sequences were subjected to further analysis.

Sequence variability and secondary structure analysis of ITSs

Sequence variability and GC content of the ITSs were analysed for the obtained nrDNA consensus sequences, including the ITS1–5.8S–ITS2 sequences and partial sequences of 18S and 26S rDNA, across the six cultivars. Sequence variation was investigated using DANSP (Rozas *et al.*, 2017), and the Oligo Calculator (http://biotools.nubic.northwestern.edu/OligoCalc.html) was used to calculate GC content (Kibbe, 2007). The secondary structures of the ITS1, 5.8S rRNA and ITS2 sequences of the six pearl millets were determined separately using the mfold webserver (http://mfold.rna.albany.edu/; Zuker, 2003). RNA motifs were predicted using the MEME webserver (http://meme-suite.org/tools/meme; Bailey and Elkan, 1994). The ITS1, 5.8S rRNA and ITS2 sequences of the six cultivars were imported to the MEME webserver to identify five motifs with a length of 8 bp in each alignment set of the ITS1, 5.8S rRNA and ITS2 sequences. Column alignment profile plots for ITS2 consensus secondary structures were generated using LocaRNA-P (http://rna.informatik.uni-freiburg.de/LocARNA/; Will *et al.*, 2012).

https://doi.org/10.1017/S1479262121000393 Published online by Cambridge University Press

Table 2. Accession numbers, GC content and variable nucleotide positions in the alignment of the obtained ITS sequences of the six *P. glaucum subsp. Monodii* cultivars

				Variable nucleotides position					
		ITS1, 5.8S and ITS2		ITS1	ITS2				26S
Pearl millet cultivars	NCBI accession no.	Length (bp)	GC (%)	217	515	609	631	648	725
Baydhan	MN781142	773	57	–	G	T	C	C	A
Sayah	MN781143	774	57	A	A	C	C	T	A
Jazan	MN781144	772	57	–	G	T	G	C	–
Tihamah	MN781145	773	57	–	G	T	C	C	G
Yemeni	MN781146	773	57	–	G	T	C	C	A
Indian	MN781147	773	57	–	G	T	C	C	A

Phylogenetic and polymorphism analyses

The consensus nrDNA sequences for the six pearl millet cultivars were aligned using MEGA X (Kumar *et al.*, 2018) and subjected to a BLAST search to retrieve the complete sequences of ITS, including the ITS1, 5.8S rRNA and ITS2 regions, of other millet cultivars in the NCBI database. In total, 31 ITS accessions from different *P. glaucum* subsp. *monodii* cultivars were selected for the phylogenetic and nucleotide polymorphism analyses. These accessions included ITS sequences of *Cenchrus americanus*, *Petalophyllum americanum* and the hybrid between *C. americanus* and *P. purpureum*, as *C. americanus* and *P. americanum* are synonyms of *P. glaucum* (Veldkamp, 2014). *P. purpureum* is the tetraploid relative of pearl millet ($2n = 4x = 28$; Jauhar and Hanna, 1998). The selected accessions of pearl millet cultivars and hybrids were grouped according to their countries of origin (China, India, Korea and France). The six ITS sequences obtained in this study were grouped together in a distinct group named 'local accessions'. Three ITS sequences from three *Panicum* species were used as an outgroup in the phylogenetic and polymorphism analyses. The accession numbers and origins of accessions imported from NCBI are listed in online Supplementary Table S1. The sequences of ITS1, 5.8S and ITS2 were determined by alignment with selected ITS sequences retrieved from NCBI. Then, the sequences of 18S and 26S rDNA from all accessions were removed from the phylogenetic analysis. Phylogenetic trees were constructed using two methods, maximum likelihood (ML; Felsenstein and Churchill, 1996) and maximum parsimony (MP; Felsenstein, 1985), with interior branch tests of 1000 replicates using the alignment of the complete sequences of ITS1, 5.8S and ITS2 on MEGA X software (Kumar *et al.*, 2018). DNASP version 6.12 (Rozas *et al.*, 2017) was used to determine the nucleotide diversity and the polymorphic and singleton sites in separate sequences of ITS1 and ITS2 among the 37 accessions.

Results

ITS sequence variability

The complete sequences of the ITS1–5.8S–ITS2 regions with partial sequences of 18S and 26S rRNA of six pearl millet cultivars were sequenced and submitted to NCBI (available under accession numbers from MN781142 to MN781147; Table 2). The obtained ITS sequences ranged from 772 to 774 bp in length. GC content was approximately 57% in all cultivars. The length of the ITS2 sequence was 219 bp in all six cultivars, whereas the length of ITS1 was 207 bp in all cultivars except the cultivar Sayah (208 bp). The partial sequence of 18S rRNA and complete sequence of 5.8S rRNA did not show any variation, with 100% similarity among all cultivars. However, the ITSs and the partial sequence of 26S rRNA showed nucleotide variation, and the ITS2 sequence showed the highest variation among the six pearl millet cultivars (Table 2). A single variable nucleotide site was observed in the 26S sequence (A_{725}→G substitution in Tihamah and deletion in Jazan cultivar), whereas ITS2 was contained of four variable nucleotide sites owing to base substitution (G→A, T→C and C→G) in the Sayah and Jazan cultivars. Additionally, thymine base insertion at nucleotide 217 was observed in the ITS1 sequence of Sayah cultivar. The cultivar Sayah exhibited the highest number of variable nucleotide sites (three sites), followed by Jazan and Tihamah, and the Baydhan, Yemeni and Indian cultivars showed 100% identity in the entire obtained sequences that included the ITS1–5.8S–ITS2 regions and partial sequences of 18S and 26S rRNAs.

Secondary structures of ITS1, 5.8S rRNA and ITS2

The secondary structure of the ITS1 sequence of the six cultivars was generated at minimum free energy (MFE) = 86.40 kcal/mol. As shown in Fig. 1a, the secondary structure contained an open external loop at 5' and 3' ends and two central multi-branched loops. The mismatch pair types A–G, A–A, G–A, C–A and U–U were found in the stem helix and helices II and III. In the ITS1 sequences, the motifs were predicted at *E*-value < 4.7×10^{-72} (online Supplementary Fig. S2A).

The secondary structure of the 5.8S sequence of the six cultivars was generated at MFE = 53.50 kcal/mol. The consensus secondary structure contained two central multi-branched loops, from which three helices emerged and two interior loops. The base mismatch type U–C was found in helix I, and the A–G and G–A mismatch pairs were found in helix IV (Fig. 1b). In the 5.8S sequences, the motifs were predicted at *E*-value < 1.7×10^{-77} (online Supplementary Fig. S2B).

The secondary structure of the ITS2 sequence of the Baydhan, Indian, Tihamah and Yemeni cultivars was generated at MFE = 105 kcal/mol (Fig. 2a). It contained four helices, of which helix III was the longest, with mismatch pair type U–G in helix II and types U–U and A–C in helix III. The secondary structure

https://doi.org/10.1017/S1479262121000393 Published online by Cambridge University Press

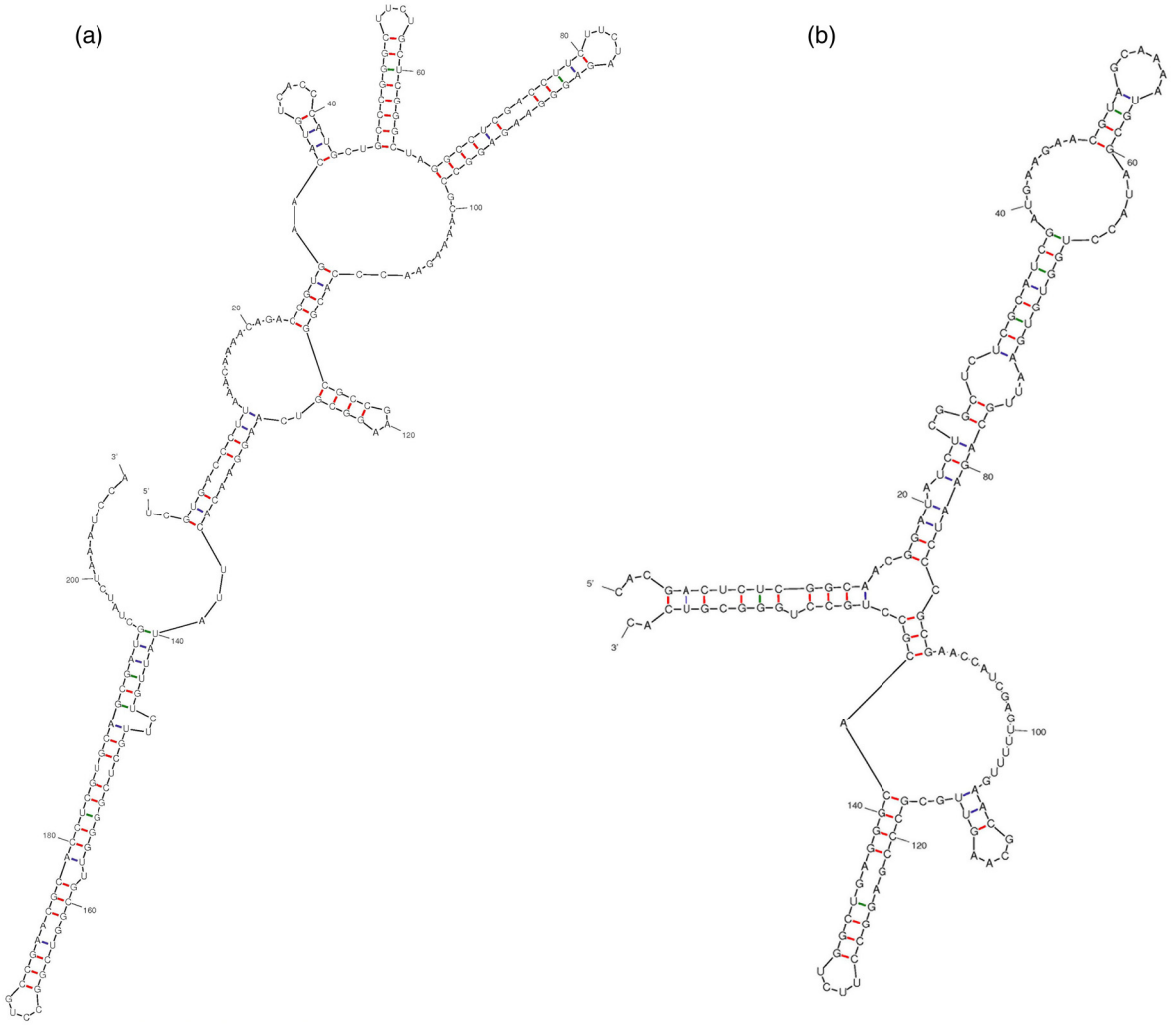

Fig. 1. Secondary structures of ITS1 and 5.8S sequences from local pearl millet cultivars predicted by the program mfold. Positions within the structures are numbered every 20 nucleotides. Helices are numbered I–V. The bonds between complementary AU pairs in the RNA helices are shown by blue lines, those between GC are shown by red lines and the stable pairs of GU are shown by green lines. (a) The secondary structure of ITS1 predicted at MFE value of 86.4 kcal/mol of all cultivars. (b) The secondary structure of the 5.8S sequence of all cultivars predicted at MFE value of 53.5 kcal/mol.

of the ITS2 sequence of the Jazan cultivar was generated at MFE = 102.6 kcal/mol. It contained five helices, of which helix III was the longest, one multi-branched loop and two interior loops (Fig. 2b). The mismatch pairs were U–G and U–C in helix II and U–U and A–C in helix III. The secondary structure of the ITS2 sequence of the Sayah cultivar was highly distinct from that of the other cultivars (Fig. 2c). It was generated at MFE = 101 kcal/mol and contained a central multi-branched loop from which three helices emerged and five interior loops. Helix II was the longest with mismatch pair types U–U and A–C. Additionally, U–U and C–C mismatch pairs were located in the stem helix. In ITS2 sequences, the motifs were predicted at *E*-value $< 3.3 \times 10^{-73}$ (online Supplementary Fig. S2C). The predicted motifs were distributed in the multi-branched loops or helices. The consensus secondary structure for the ITS2 of the six cultivars was predicted by LocARNA-P and compared with the structures predicted by mfold. As expected, the consensus secondary structure was identical to the secondary structure of the ITS2 of the Baydhan, Indian, Tihamah and Yemeni cultivars and different from that of Jazan and Sayah (online Supplementary Fig. S1).

The mfold secondary structures of the ITS2 sequences obtained in the current study were compared with the ITS2 secondary structure templates in the ITS2 database (http://its2.bioapps.biozentrum.uni-wuerzburg.de/; Ankenbrand *et al.*, 2015). The secondary structure of the ITS2 sequence of the Baydhan, Indian, Tihamah and Yemeni cultivars was predicted directly through the ITS2 database, and it was identical to the consensus ITS2 secondary structure predicted by mfold and LocARNA-P. However, the secondary structures of ITS2 sequences of Jazan and Sayah could not be directly predicted by the ITS2 database, but they could be modelled by homology with high-quality models from the ITS2 database. Five high-quality templates of the ITS2 secondary structure from *P. glaucum* subsp. *monodii* were found. The template number 49066367 was the best model with 78.3% similarity with the ITS2 of Jazan, and 77.4% similarity with Sayah, and 100% of helix transfers for helices I, III and IV, and 85.7% for helix II.

https://doi.org/10.1017/S1479262121000393 Published online by Cambridge University Press

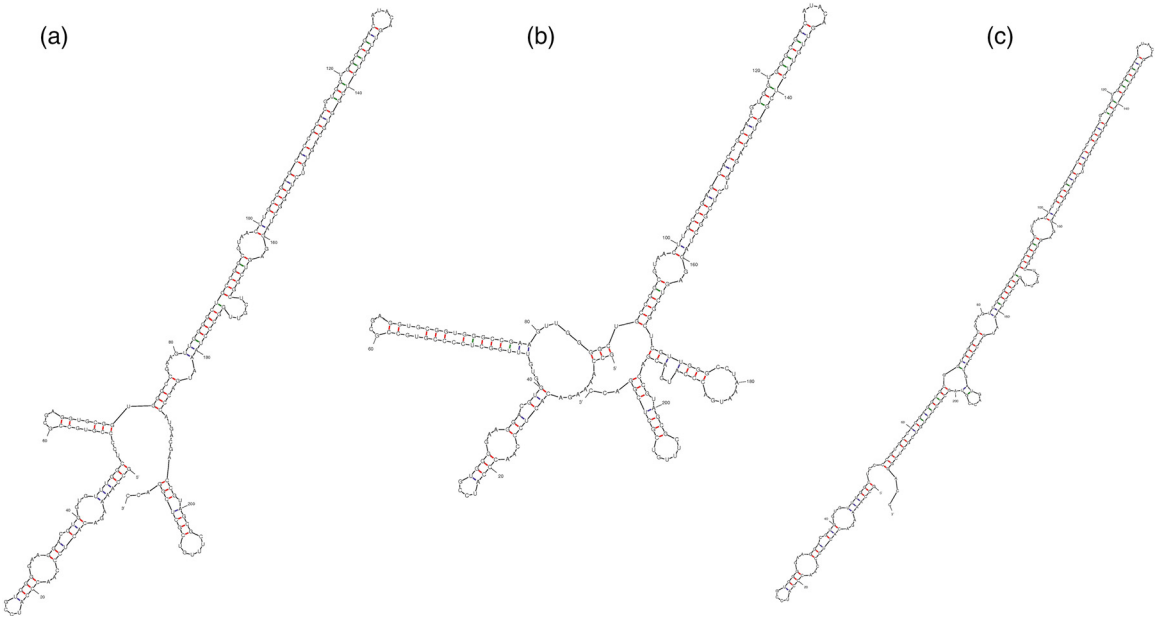

Fig. 2. Secondary structures of ITS2 sequences from local pearl millet cultivars predicted by the program mfold. Positions within the structures are numbered every 20 nucleotides. Helices are numbered IV. The bonds between complementary AU pairs in the RNA helices are shown by blue lines, those between GC are shown by red lines and the stable pairs of GU are shown by green lines. (a) The secondary structure of ITS2 sequence of all cultivars except for Sayah and Jazan predicted at 105 kcal/mol. (b) The secondary structure of ITS2 sequence of Jazan cultivar predicted at 102.6 kcal/mol. (c). The secondary structure of ITS2 sequence of the Sayah cultivar predicted at 101 kcal/mol.

Phylogenetic trees and polymorphism of cultivars

The constructed ML and MP trees showed similar clustering of the six sequences obtained in this study, which also showed close relationships with Indian and French accessions with certain differences in the positions in the Jazan and Baydhan cultivars (Fig. 3). The Korean accessions formed a separate cluster from all accessions with 100% bootstrap support in both ML and MP phylogenetic trees. The Chinese, Indian, French and the local cultivars clustered together in a large cluster in both trees. However, the Chinese accessions separated from other accessions in a separate subcluster with 100% bootstrap support in both trees. The second subcluster included the accession from the Indian, French and the local cultivars. In the ML tree (Fig. 3a), five local cultivars (Indian, Yemeni, Tihamah, Jazan and Baydhan) were separated as individual branches from the Indian accessions within the second subcluster with bootstrap values of 27–28%. In the MP tree (Fig. 3b), three cultivars (Indian, Yemeni and Tihamah) were separated as individual branches from the Indian and French accessions with a bootstrap value of 23% within the second subcluster. The local cultivars Jazan and Baydhan showed a different position in the MP tree from the ML tree. Jazan was clustered with the other group of Indian accession with 54% bootstrap support, and Baydhan was clustered with French accessions as an internal branch with a bootstrap value of 85%. The local cultivar Sayah formed a separate external branch from the second subcluster that included the six investigated cultivars with French and Indian accessions in both trees with bootstrap values of 44 and 98% in the ML and MP trees, respectively.

Nucleotide polymorphisms of ITS1 and ITS2 sequences were investigated using two multiple sequence alignment sets. The first alignment set included the six local cultivars. The second alignment set included 37 pearl millet cultivars, of which six accessions were the local cultivars and 31 accessions were imported from NBCI. ITS2 exhibited higher nucleotide polymorphism than ITS1 in both alignment sets. However, the degree of polymorphism among the six local cultivars was lower than that among the 37 accessions. As shown in Table 3, the number of variable sites and parsimony-informative sites and the nucleotide diversity value were all zero for the ITS1 sequences of the local cultivars due to the strict sequence conservation. However, the number of variable sites in the ITS2 sequences of local cultivars was four, and the nucleotide diversity value was 0.006.

Discussion

In the current study, we used molecular characterization of ITS regions to identify local pearl millet cultivars and to evaluate the polymorphism among these cultivars. Evaluation of genetic diversity among millet cultivars can help in the identification and conservation of these genetic resources. The obtained ITS1–5.8S–ITS2 sequences revealed that the local cultivar Sayah exhibited the highest number of variable nucleotide sites, followed by Jazan and Tihamah. These nucleotide variations between ITS sequences were due chiefly to point mutations and rarely to nucleotide insertions or deletions (Baldwin *et al.*, 1995).

The predicted motifs in ITS1–5.8S–ITS2 sequences in Saudi millet cultivars appeared to be conserved in plants. The motifs GAAGGCGU and AAGGAACA were previously identified as conserved motifs in ITS1 in flowering plants (Liu and Schardl, 1994). The secondary structure of ITS1 is a helpful tool in the classification of green algae at the species level (Coleman *et al.*, 1998). The conserved motifs within ITS1 may have a key role in the processing of rRNA genes (Liu and Schardl, 1994). The motifs AUGAAGAA and UUUUUGAA have been previously identified as conserved in 5.8S sequences among angiosperms (Liu and Schardl, 1994; Harpke and Peterson, 2008). The motifs AGAAUCCC and UGUGAAUU in the 5.8S sequence of local cultivars are well known to be conserved in flowering plants

https://doi.org/10.1017/S1479262121000393 Published online by Cambridge University Press

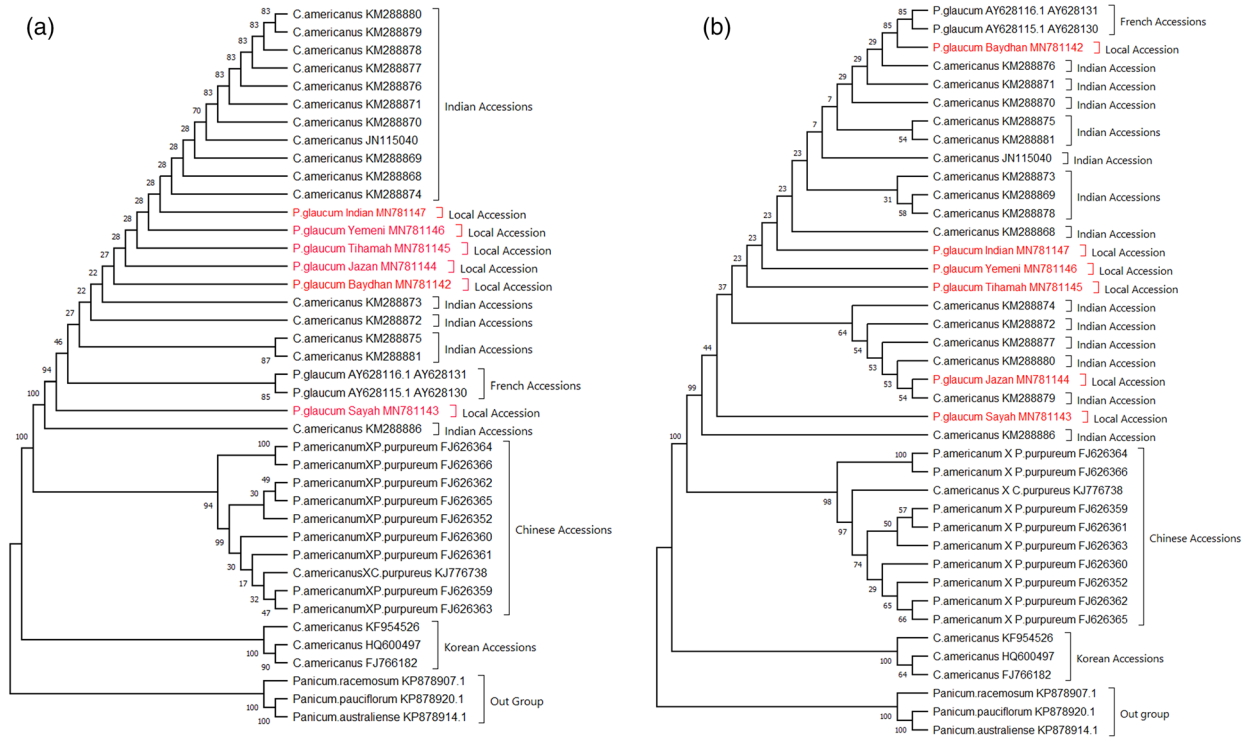

Fig. 3. Phylogenetic trees of pearl millet cultivars based on ITS1, 5.8S and ITS2 sequences. The local pearl millet accessions obtained in this study are shown in red. The other 31 millet accessions adopted from NCBI GenBank are grown in four different countries – China, India, Korea and France. The origins of the NCBI accessions are shown to the right of the clusters. Outgroup accessions included *Panicum* species. Species names appear as in the NCBI database (*C. americanus* is the synonym of *P. glaucum*, and *P. purpureum* is the tetraploid relative of pearl millet). The respective NCBI accession numbers are provided after each species name. Support for each node was evaluated using 1000 bootstrap resampling. Phylogenetic trees were constructed using (a) the ML method and (b) the MP method using MEGA X software.

Table 3. Nucleotide polymorphisms of ITS1 and ITS2 sequences within the six local cultivars and within the 37 pearl millet cultivars including the local cultivars

Polymorphism parameters	Local accessions		All 37 accessions	
	ITS1	ITS2	ITS1	ITS2
Number of sites	208	219	213	221
Sites with alignment gaps or missing data	1	0	19	7
Invariable (monomorphic) sites	207	215	148	160
Variable (polymorphic) sites	0	4	46	54
Total number of mutations	0	4	48	56
Singleton variable sites	0	4	14	8
Parsimony informative sites	0	0	32	46
Nucleotide diversity	0	0.006	0.043	0.049

(Jobes and Thien, 1997). The 5.8S secondary structure for pearl millets and the motifs identified in the current study are conserved among different plants and fungi (Freire *et al.*, 2012; Hodač *et al.*, 2014; Rampersad, 2014). The predicted motif ACGUGGUG included the UGGU motif, which was previously identified in eukaryotes near the 5′ end of helix III of ITS2 (Schultz *et al.*, 2005; Wolf *et al.*, 2005). This motif was located near the endonuclease cleaves site C2, where the initial step of the degradation process of ITS2 occurs (Geerlings *et al.*, 2000). The CCA single strand at 3'-end in ITS2 secondary structures appeared to be conserved in all angiosperms (Hershkovitz and Zimmer, 1996). In eukaryotes, the common ITS2 secondary structure is clover leaf-like with four helices, and it contains conserved motifs, such as the U/C–U mismatch in helix II, that are important for excision of the ITSs from pre-rRNA to obtain mature rRNAs (Mai and Coleman, 1997; Schultz *et al.*, 2005). The secondary structures of the ITS2 sequence of the Baydhan, Indian, Tihamah and Yemeni cultivars were constant with the clover leaf-like structure of eukaryotes. However, the Jazan genotype had an additional helix in the ITS2 structure, while the number of helices in the ITS2 sequences of Sayah cultivar was different from that in the commonly reported structure of eukaryotic ITS2. Additionally, the conserved U/C–U mismatch in helix II was only shown in Jazan cultivars. The variability in the ITS2 secondary structure among the different cultivars indicated the applicability of the ITS2 sequence in distinguishing between cultivars of the same species (Coleman, 2000, 2009).

The phylogenetic relationships within the species of *P. glaucum* subsp. *monodii*, based on ITS sequences, revealed close relationships between Indian and French pearl millet and all local cultivars, with the exception of the Sayah cultivar, which appeared to have distinct accession in the phylogenetic trees. This finding indicated that ITSs appeared to be sufficiently variable to facilitate differentiation between closely related organisms (Álvarez and

https://doi.org/10.1017/S1479262121000393 Published online by Cambridge University Press

Wendel, 2003). Nucleotide polymorphism analysis showed that the nucleotide diversity value and variable sites for the ITS2 sequences were higher than those in ITS1 sequences among the local cultivars due to the strict sequence conservation for ITS1 compared with ITS2. Previous studies have also reported consistent findings of the applicability of ITS2 sequences than ITS1 for differentiating between close cultivars (Gao *et al*., 2010; Han *et al*., 2013).

Our results showed that the native cultivar Baydhan appeared to share similar ITS sequences with Yemeni and Indian cultivars. However, high polymorphism in the ITS2 sequence was observed in the other three native cultivars. The Sayah cultivar was distinct from the other local cultivars as it showed an especially high-nucleotide polymorphism and a different secondary structure for ITS2. The uniqueness of the native cultivar, Sayah, emphasizes the importance of conservation of genetic diversity of the native cultivars. The findings of this study provide insights into the degree of variation among local cultivars and contribute to the conservation of native pearl millets. Furthermore, ITS2 sequencing is a valuable tool to further identify and evaluate the genetic diversity within close cultivars.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1479262121000393.

Acknowledgements. We are greatly indebted to the Central Laboratory for Science and Medical Studies at King Saud University for providing sequencing services. We also thank Maha Alrwais for her excellent technical assistance during RNA extraction. This study was financially supported under Project No. 2019/01/9512 from the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University.

References

Al-Turki TA, Al-Namazi AA and Masrahi YS (2019) Conservation of genetic resources for five traditional crops from Jazan, SW Saudi Arabia, at the KACST Gene-Bank. *Saudi Journal of Biological Sciences* **26**(7): 1626–1632.

Álvarez I and Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**(3): 417–434.

Ankenbrand MJ, Keller A, Wolf M, Schultz J and Förster F (2015) ITS2 database V: Twice as much. *Molecular Biology and Evolution* **32**(11): 3030–3032.

Bailey TL and Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. International Conference on Intelligent Systems for Molecular Biology* **2**: 28–36.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS and Donoghue MJ (1995) The ITS Region of Nuclear Ribosomal DNA A Valuable Source of Evidence on Angiosperm Phylogeny. *Annals of the Missouri Botanical Garden* **82**(2): 247–277.

Bhattacharjee R, Bramel P, Hash C, Kolesnikova-Allen M and Khairwal I (2002) Assessment of genetic diversity within and between pearl millet landraces. *Theoretical and Applied Genetics* **105**(5): 666–673.

Bidinger FR and Hash TC (2003) Pearl millet. In H T Nguyen and A Blum (Eds.), *Physiology and Biotechnology Integration for Plant Breeding* (pp. 225–270) New York: Marcel Dekker.

Buckler ES, Ippolito A and Holtsford TP (1997) The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications. *Genetics* **145**(3): 821–832.

Coleman A, Preparata R, Mehrotra B and Mai J (1998) Derivation of the Secondary Structure of the ITS-1 Transcript in Volvocales and its Taxonomic Correlations. *Protist* **149**: 135–146.

Coleman AW (2000) The Significance of a Coincidence between Evolutionary Landmarks Found in Mating Affinity and a DNA Sequence. *Protist* **151**(1): 1–9.

Coleman AW (2009) Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution* **50**(1): 197–203.

Felsenstein J (1985) Confidence Limits on Phylogenies: an Approach Using the Bootstrap. *Evolution* **39**(4): 783–791.

Felsenstein J and Churchill GA (1996) A hidden markov model approach evolution to variation among sites in rate of evolution, *Molecular Biology and Evolution* **13**(1): 93–104.

Freire MCM, Silva MR da, Zhang X, Almeida ÁMR, Stacey G and Oliveira LO de (2012) Nucleotide polymorphism in the 5.8S nrDNA gene and internal transcribed spacers in Phakopsora pachyrhizi viewed from structural models. *Fungal Genetics and Biology* **49**(2): 95–100.

Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, Panga X, Xu H and Chen, S (2010) Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of Ethnopharmacology* **130**(1): 116–121.

Geerlings TH, Vos JANC and Raue HA (2000) The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by 5′ → 3′ exonucleases. *RNA* **6**(12): 1698–1703.

Han J, Zhu Y, Chen X, Liao B, Yao H, Song J, Chen S and Meng F (2013) The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. *BioMed Research International* **2013** (741476): 1–7.

Harpke D and Peterson A (2008) 5.8S motifs for the identification of pseudogenic ITS regions. *Botany* **86**: 300–305.

Havilah E. J (2011) Forages and Pastures | Annual Forage and Pasture Crops – Species and Varieties. In J. W. Fuquay (Ed.), *Encyclopedia of Dairy Sciences* (2nd ed., pp. 552–562) San Diego: Academic Press.

Hemleben V (1993) Repetitive and highly repetitive DNA components as molecular markers for evolutionary studies and in plant breeding. *Current Topics in Molecular Genetics* (*Life Science Advances*) **1**:173–85

Hershkovitz MA and Zimmer EA (1996) Conservation patterns in angiosperm rDNA ITS2 sequences. *Nucleic Acids Research,* **24**(15): 2857–2867.

Hodač L, Scheben AP, Hojsgaard D, Paun O and Hörand E (2014) ITS polymorphisms shed light on hybrid evolution in apomictic plants: a case study on the *Ranunculus auricomus* complex. *PLoS ONE* **9**(7): e103003.

Hsiao C, Chatterton NJ, Asay KH and Jensen KB (1994) Phylogenetic relationships of 10 grass species: an assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. *Genome* **37**(1): 112–120.

ICRISAT (1996) Improving the unimprovable — succeeding with pearl millet. *Food from thought*. No. 3 Patancheru, India: ICRISAT.

Jauhar PP and Hanna WW (1998) Cytogenetics and Genetics of Pearl Millet. In D L Sparks (Ed.), *Advances in Agronomy* (Vol. 64, pp. 1–26) Academic Press. New York

Jobes DV and Thien LB (1997) A Conserved Motif in the 5.8S Ribosomal RNA (rRNA) Gene is a Useful Diagnostic Marker for Plant Internal Transcribed Spacer (ITS) Sequences. *Plant Molecular Biology Reporter* **15** (4): 326–334.

Kibbe WA (2007) OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Research* 35(Suppl_2): W43-W46.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA and Janzen D H (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* **102**(23): 8369–8374.

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**(6): 1547–1549.

Liu JS and Schardl CL (1994) A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. *Plant Molecular Biology* **26**(2): 775–778.

Mai JC and Coleman AW (1997) The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *Journal of Molecular Evolution* **44**(3): 258–271.

Martel E, Poncet V, Lamy F, Siljak-Yakovlev S, Lejeune B and Sarr A (2004) Chromosome evolution of Pennisetum species (Poaceae): Implications of ITS phylogeny. *Plant Systematics and Evolution* **249**(3–4): 139–149.

McBenedict B, Chimwamurombe P, Kwembeya E and Maggs-Kölling G (2016) Genetic diversity of Namibian *Pennisetum glaucum* (L) R. BR (Pearl Millet) landraces analyzed by SSR and morphological markers. *Scientific World Journal* **1439739**: 1–11.

Rampersad SN (2014) ITS1, 5.8S and ITS2 secondary structure modelling for intra-specific differentiation among species of the Colletotrichum gloeosporioides sensu lato species complex. *SpringerPlus* **3**(1): 1–10.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE and Sanchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34**(12): 3299–3302.

Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**(12): 5463–5467.

Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, Connor R, Fiorini N, Funk K, Hefferon T, Holmes JB, Kim S, Kimchi A, Kitts PA, Lathrop S, Lu Z, Madden TL, Marchler-Bauer A, Phan L, Schneider VA, Schoch CL, Pruitt KD and Ostell J (2019) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **47**(D1): D23–D28.

Schultz J, Maisel S, Gerlach D, Müller T and Wolf M (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* **2**: 361–364.

Van Nues RW, Rientjes JMJ, Van Der Sande CAFM, Zerp SF, Sluiter C, Venema J, Planta RJ and Raué HA (1994) Separate structural elements within internal transcribed spacer 1 of Saccharomyces cerevisiae precursor ribosomal RNA direct the formation of 17S and 26S rRNA *Nucleic Acids Research* **22**(6): 912–919.

Veldkamp JF (2014) A revision of Cenchrus incl. Pennisetum (Gramineae) in Malesia with some general nomenclatural notes. *Blumea: Journal of Plant Taxonomy and Plant Geography* **59**(1): 59–75.

Veldman GM, Klootwijk J, van Heerikhuizen H and Planta RJ (1981) The nucleotide sequence of the intergenic region between the 5.8S and 26S rRNA genes of the yeast ribosomal RNA operon. Possible implications for the interaction between 5.8S and 26S rRNA and the processing of the primary transcript. *Nucleic Acids Research* **9**(19): 4847–4862.

Venkateswarlu K and Nazar R (1991) A conserved core structure in the 18-25S rRNA intergenic region from tobacco, Nicotiana rustica. *Plant Molecular Biology* **17**(2): 189–194.

Will S, Joshi T, Hofacker IL, Stadler PF and Backofen R (2012) LocARNA-P: Accurate boundary prediction and improved detection of structural RNAs. *RNA* **18**(5): 900–914.

Wolf M, Schultz J, Dandekar T and Achtziger M (2005) Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. *RNA,* **11**(11): 1616–1623.

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**(13): 3406–3415.