RESEARCH ARTICLE

Genetic assessment of the internal transcribed spacer region (ITS1.2) in *Mangifera indica* L. landraces

Bahman Fazeli-Nasab¹ · R. Z. Sayyed² · Mohammad Farsi³ · Sahar Ansari⁴ · Hesham Ali El-Enshasy^{5,6}

Received: 17 September 2019 / Revised: 27 October 2019 / Accepted: 14 November 2019 / Published online: 11 December 2019
© Prof. H.S. Srivastava Foundation for Science and Society 2019

Abstract Mango (*Mangifera indica*) is one of the most important tropical fruits in the world. Twenty-two genotypes of native mangoes from different regions of southern Iran (Hormozgan and Kerman) were collected and analyzed for the ribosomal genes. GC content was found to be 55.5%. Fu and Li's D* test statistic (0.437), Fu and Li's F* test statistic (0.500) and Tajima's D (1.801) were positive and nonsignificant. A total of 769 positions were identified (319 with insertion or deletion including 250 polymorphic and 69 monomorphic loci; 450 loci without any insertion or deletion including 35 Singletons and 22 haplotypes). Nucleotide diversity of 0.309 and a high genetic differentiation including Chi square of 79.8; *P* value of 0.3605 and *df* value of 76 was observed among mango genotypes studied. The numerical value of the ratio dN/dS (0.45) indicated a pure selection in the examined gene and the

absence of any key changes. Cluster analysis differentiated the mango used in this research (*M. indica* L.) into two genotypes but could not differentiate their geographical locations. The results of this study indicated that a high genetic distance exists between HajiGholam (Manojan) and Arbabi (Rodan) genotypes and showed higher genetic diversity in mango of Rodan region. Results of present study suggested that for successful breeding, the genotypes of Rodan region mango especially Arbabi mango can be used as a gene donor and ITS can be a suitable tool for genetic evaluations of inter and intra species.

Keywords Mango · ITS region · dN/dS · GC content · Nucleotide diversity · Haplotype · Singleton · Ribosomal genes · Cluster analysis

✉ R. Z. Sayyed
sayyedrz@gmail.com

¹ Research Department of Agronomy and Plant Breeding, Agricultural Research Institute, University of Zabol, Zabol 9861335856, Iran

² Department of Microbiology, PSGVP Mandal's Arts, Science, and Commerce College, Shahada, Maharashtra 425 409, India

³ Department of Biotechnology and Plant Breeding, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

⁴ Department of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Zabol, Zabol, Iran

⁵ Institute of Bioproduct Development (IBD), School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru, Johor, Malaysia

⁶ City of Scientific Research and Technology Applications, New Burg Al Arab, Alexandria, Egypt

Introduction

The *Mangifera indica* L. commonly known as mango is popularly referred to as “Asia's King of Fruits” or a Royal fruit (Swaroop et al. 2018). It is a perennial and evergreen plant that belongs to *Mangifera* genus and *Anacardiaceae* family, and carries $2n = 2x = 40$ chromosome number (Litz 2009; Ravishankar et al. 2017). Mangoes are economically and nutritionally the most valuable fruits in the southern regions of Iran and in the Asian continent (Fazeli-Nasab and Fahmide 2018). Mangoes in Southeast Asia have been cultivated for about 4000 years (Lal et al. 2017; Swaroop et al. 2018; Emmami 2001). Determining the traits and the genetic diversity are necessary to improve their quantity and quality of plants (McKain et al. 2018; Yousef et al. 2018; Jani Pour et al. 2018).

Conventional mango breeding suffers from low heterozygosity, prolonged period of development (Singh and

Singh 2012 Tiwari et al. 2018). Molecular markers are powerful tools for investigating the diversity within and between the populations; as they are neutral in terms of phenotypes, remains unaffected by environmental conditions, and, despite the high degree of heritability, they show limited polymorphism (Naghavi et al. 2004; Sargazi et al. 2016). Internal Transcribed Sequence (ITS) marker have not been frequently used in genetic assessment of mango (Yonemori et al. 2002; Hidayat et al. 2013, Gitahi et al. 2016; Nashima et al. 2017). The ITS regions are located between the 18S and 28S rRNA genes and also have distinct advantages over other molecular markers. The advantages include increased sensitivity due to the existence of approximately 100 copies per genome (Henry et al. 2000) and sequence variation. These merits of ITS regions has led to their use in phylogenetic studies of many different organisms (Ansari et al. 2018). On the other hand among the regions of the ribosomal cistron, the ITS region has the highest probability of successful identification for the broadest range of varieties, with the most clearly defined barcode gap between inter- and intra- specific variation in wheat (Schoch et al. 2012). Although mango is the economically most important fruit, the classical breeding and lack of quality cultivars have been the main obstacle in the high production and development of this fruit. ITS marker offers several useful features

such as high copy number, small in size, possess highly conserved flanks, rapid concerted evolution and universality of primers (Hidayat et al. 2013). Hence the present study was aimed to use ITS1.2 regions to decipher the phylogenetic relationship and genetic diversity of Iranian *Mangifera indica* L. landraces.

Methods

Plant materials

The young, flushing leaves of *Mango landraces* (Table 1) were collected from Hormozgan and Kerman Province of south Iran totally with an area of 253,982 Km².

Preliminary identification of plant material

Identification of mango species was carried out by Department of Plant Biology, University of Zabol, Iran. The traits for identification were the whole plant (size, shape, etc.), its flowers (color, size, growing position, inflorescence, etc.), its stem (shape, node, outer character, bark pattern, etc.), its fruits (size, color, quality, etc.), and its leaves (shape, margin, pattern, texture, vein etc.).

Table 1 Names and place of collecting Mango genotypes

Row	Persian Name	Origin	Longitude	latitude	Average daily temperature (°C)	Average rainfall (mm)	Altitude (m)	Accession number
1	Helov	Jirof	57° 44'	28° 40'	25–26	190	685	KY433362
2	Nargil	Jirof	57° 44'	28° 40'	25–26	190	685	MH819725
3	Khelo	Hishin	57° 25'	28° 40'	24–25	230	1320	KY433364
4	Shahani1	Hishin	57° 25'	28° 40'	24–25	230	1320	KY433365
5	Shahani2	Hishin	57° 25'	28° 40'	24–25	230	1320	KY433375
6	Nabati1	Manojan	57° 44'	27° 24'	26	180	340	KY471734
7	Nabati2	Manojan	57° 44'	27° 24'	26	180	340	KY471735
8	Podney	Manojan	57° 30'	27° 24'	26	180	340	KY433358
9	Dogholoo	Kahnoj	57° 30'	27° 24'	26	180	340	MH847789
10	Hajigholam	Dehkahan	57° 17'	27° 19'	26–27	170	460	KY433282
11	Khiyar1	Dehkahan	57° 17'	27° 19'	26–27	170	460	KY457567
12	Mikhaki1	Rodan	57° 11'	27° 27'	27–28	250	195	MH847788
13	Charak	Rodan	57° 11'	27° 27'	27–28	250	195	KY471736
14	Kharvst	Rodan	57° 11'	27° 27'	27–28	250	195	KY435978
15	Zapak	Rodan	57° 11'	27° 27'	27–28	250	195	KY445941
16	Arbabi	Rodan	57° 11'	27° 27'	27–28	250	195	KY457412
17	Majlesi2	Fariab	57° 25'	27° 38'	26	250	500	KY471737
18	Zarak	Fariab	57° 25'	27° 38'	26	250	500	KY457566
19	Gol	Bolok	57° 40'	27° 52'	27	160	650	KY433359
20	Anonym2	Bolok	57° 40'	27° 52'	27	160	650	KY433361
21	Langra	Minab	57° 04'	27° 09'	55–12	253	16	KY449412
22	Senderi	Minab	57° 04'	27° 09'	55–12	253	16	KY449411

DNA extraction and polymerase chain reaction (PCR)

Genomic DNA was extracted as per Dellaporta et al. 1983. Qualitative and quantitative estimation was performed by using agarose gel (1%) and a UV–VIS spectrophotometer (Model, LUV-300, Labnika, USA), respectively. The ITS region was amplified by using Polymerase Chain Reaction (PCR) with a set of primers namely AB101 (5'-ACGAA TTCATGGTCCGGTGAAGTGTTTCG-3') as the forward primer and AB102 (5'-TAGAATTCCCCGGTTCGCT CGCCGTTAC-3') as the reverse primer (Robinson et al. 2001).

The PCR reaction mixture containing 20 ng DNA, 5 μ L master mix (Amplicon, Denmark), 0.7 μ L of each primer (10 pmol) and water to a final volume of 15 μ L was amplified on Eppendorf Thermal Cycler (22331 Hamburg) with the thermal profile: denaturation at 95 °C for 5 min, then followed by 35 cycles of 55 s at 94 °C, the optimized annealing temperature (60 °C) for 50 s and 72 °C for 55 s, finishing extension at 72 °C for 10 min. The amplified DNA fragment was visualized on 2% (w/v) agarose gel (Merck, Germany) in Tris–acetate–EDTA buffer (pH 7.8) with Gel Red staining using Gel document (E-Box-VX2-20 m, vilber-lourmat, France) (Timura et al. 2013).

Sequencing and data analysis

PCR fragments were sequenced and the sequences quality was analyzed using Chromas 2.3. Comparison of individual sequences with the sequences in the NCBI database was carried out by using BLAST online software. Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). The sequences were aligned and clustered and their similarity and genetic distances were defined using the ClustalW method, by DNASTAR software and MEGA6 software. DnaSP5 software was also used to perform and calculate additional analyses including dN and dS. Transitional and transversal nucleotide substitution was calculated on the basis of the Tamura–Nei model pattern (Tamura and Nei 1993; Tamura et al. 2011).

In order to evaluate the nucleotide diversity of all sequences, the number of polymorphic sites and Tajima's and Watterson's estimators of theta were calculated. Tajima's *D* was used to determine if the mango ITS gene family in these taxa evolved neutrally. Tajima's Neutrality Test (Tajima 1989) was conducted and the evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). Tajima's *D* and Fu and Li's *D*^{*} and *F*^{*} statistics uses the information from only intraspecific data, whereas Fu and Li's *D*^F and *F* statistics are based on the information from

the number of recent mutations; the latter, therefore, requires the presence of an outgroup to be computed (Fu and Li 1993; Simonsen et al. 1995). Negative *D* value indicates signature of population expansion. Negative Tajima's *D* reflects an excess of low frequency polymorphisms. It remains negative under selective sweeps (and population growth) and positive under balancing selection (or population structure with sampling from many populations). Fu & Li *D*^{*}'s negative value indicates an excess of singletons. Significant departure of these tests reflects an excess of new mutations as a result of evolutionary forces.

The numerical value of dN/dS was conducted based on the ratio of nucleotide variations that alter the amino acids (dN) and the nucleotide variations that have no impact on the resulting amino acid (dS). The dN/dS ratio is commonly used for the sequences sampled from a single population and their resulting differences reflect the segregating polymorphisms. The dN/dS measurements have vital significance in the interpretation of among population-genetic samples.

Codon usage was calculated by using GenScript rare codon analysis (<https://www.genscript.com/tools/rare-codon-analysis>). It has often been suggested that differential usage of codons recognized by rare tRNA species with helping rRNA (in ribosomal structure) (Sharp and Li 1986). The pattern of synonymous codon usage in regulatory genes reflects primarily the relaxation of natural selection. In particular, regulatory genes are reported to have an extraordinarily high frequency of rare codons.

Nucleotide diversity was estimated by Theta (Θ), the number of segregating sites (Watterson 1975), and its standard deviation ($S\Theta$). These parameters were estimated by Dnasp software version 6.10.01 (Rozas et al. 2003).

For examining the ability of ITS marker to segregate plants and landraces (at genus or species level) based on geographical region, at first it was conducted only for Iranian *Mangifera indica* L. landraces and then was used for one group from some sequences of species of *Anacardiaceae* family in GenBank along with all sequences of Iranian *Mangifera indica* L. landraces in this research.

Results

Sequencing of ITS1-5.8s rRNA-ITS2 regions

After sequencing, the homology rate of the examined sequences with the sequences of the reference species available in the NCBI was measured. Their overall similarity was in the range of 97–99%. Following the sequence alignment, an 1109 bp fragment was obtained in which 293 bp were preserved. All the sequences were registered

in the NCBI under the specified accession numbers (Table 1).

Using the sequences obtained by DnaSP software, a total of 769 positions (319 positions with insertion and deletion (250 polymorphic loci, 69 monomorphic loci) and 450 loci without deletion and addition (35 singletons were identified)). The proportion of 22 haplotypes (haplotype diversity index equal to 1) was identified. The numerical value of 0.45 was observed for dN/dS.

Transitional and transversional nucleotide substitution analysis revealed the maximum and the minimum transitional rates which belonged to purine (6.54%) and pyrimidine (6.49%), respectively and also the maximum (10.58%) and the minimum (7.41%) transversional rates belonged to C with G and A and the A with T and C respectively. The ratio of various nucleotides to the entire nucleotides was in the following order, the average ratios of 22.3% (thymine), 28.5% (cytosine), 22.2% (adenine), and 27% (guanine) (Table 2).

The rest of traits like Theta (per sequence) from Eta, Theta (per site) from Eta, Fu's Fs statistic and Strobeck's S statistic and traits like Fu and Li's D* test statistic (0.43767), Fu and Li's F* test statistic (0.50006) and Tajima's D (1.801413) indicated positive and none significant.

Codon usage of mango ITS gene family is shown in Table 3. Analysis of negative CIS elements and repeat sequences showed that both of them were 4 and zero (0), respectively. Result of rare codon analysis is shown in Table 4.

A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+G]). The estimated value of the shape parameter for the discrete Gamma Distribution is 1.2419. Mean evolutionary rates in these categories were 0.15, 0.42, 0.75, 1.23, 2.44 substitutions per site. For estimating ML values, a tree topology was automatically computed. The maximum Log likelihood for this computation was − 3241.848.

Comparing genetic data between paired regions

Genetic distance matrix between the *Mangifera indica* L. landraces indicated that the genetic distance among the samples was from 0.02551 (between HajiGholam (Manojan) and Gol (Jiroft)) to 1.19293 (between HajiGholam (Manojan) and Arbabi (Rodan) with average 0.47224 (Table 5). Landraces in mango were not clearly separated, but Manojan and Jiroft landraces were completely separated as different group (Fig. 1). Moreover, the highest rate of the regional diversity was observed in the Mango of Rodan region (Table 6), indicating the presence of a new gene and protein in the mango of Rodan.

Table 2 The base pair ratio of ITS sequences in all landrace of mango

Genotypes	T(U)	C	A	G
Nabati1	21.5	30.7	24.1	23.7
Nabati2	18.1	26.8	24.6	30.5
HajiGholam	15.9	28.4	24.3	31.4
Mikhaki1	24.6	29.5	23.0	23.0
Charak	21.2	26.4	22.3	30.1
Podnehei	19.1	25.0	23.8	32.0
Gol	19.9	25.5	26.8	27.8
Anonymus2	23.8	29.6	21.0	25.6
Majlesi2	23.3	27.4	21.9	27.4
Helov	23.7	30.1	20.0	26.2
Nargil	24.5	28.4	21.3	25.9
Khlo	23.5	29.2	20.3	27.1
Shahani1	24.5	30.5	18.8	26.1
Shahani2	25.2	29.4	21.5	23.9
Dogholoo	21.5	23.8	24.9	29.8
Kharvast	23.2	30.3	19.5	27.0
Zapak	23.7	30.9	20.3	25.1
Khiyar1	19.8	23.0	27.6	29.6
Zarak	26.1	29.8	19.9	24.2
Langra	24.8	29.1	21.7	24.4
Senderi	21.5	31.8	22.1	24.6
Arbabi	22.6	32.0	18.5	26.9
Average	22.3	28.5	22.2	27.0

Since ITS marker was unable to separate the landraces inside regions, the result of the sequences of some plants in other species of *Anacardiaceae* available at NCBI were used for further assessments of the ITS ability in assessing the genetic diversity of different crops. They were compared with the sequences of the present study. The results indicated that each species of the family was in a separate group (Fig. 2).

In silico analysis of *Mangifera indica* L. landraces

The consensus sequence of all samples, analyzed by using BioEdit Sequence Alignment Editor 7 software (Hall 1999), was entered in Nebcutter (<http://tools.neb.com/NEBcutter2/>) to identify the specific restriction enzymes that are able to recognize haplotypes. The results showed that different enzymes are capable of cutting in the amplified area, so three types of restrictions enzymes PsuG1 with the cut site (5'...BBCGD...3' and 3'...VVGCH...5') can cut and distinguish the amplified area in different samples.

Table 3 Relative synonymous codon usage is given in parentheses following the codon frequency

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	7.4	1.22	UCU(S)	2.3	0.94	UAU(Y)	2	1.25	UGU(C)	1.6	0.48
UUC(F)	4.7	0.78	UCC(S)	3.2	1.31	UAC(Y)	1.2	0.75	UGC(C)	5.2	1.52
UUA(L)	3.5	1.08	UCA(S)	1.5	0.59	UAA(*)	1.9	1.15	UGA(*)	1.4	0.85
UUG(L)	4.3	1.34	UCG(S)	3.5	1.4	UAG(*)	1.7	1.01	UGG(W)	3.5	1
CUU(L)	2.7	0.83	CCU(P)	5.2	0.86	CAU(H)	2	0.82	CGU(R)	3	0.89
CUC(L)	4	1.26	CCC(P)	10.5	1.74	CAC(H)	3	1.18	CGC(R)	3.3	0.95
CUA(L)	1.9	0.59	CCA(P)	3.9	0.65	CAA(Q)	4.5	1.37	CGA(R)	2.8	0.82
CUG(L)	2.9	0.89	CCG(P)	4.5	0.75	CAG(Q)	2.1	0.63	CGG(R)	5.9	1.7
AUU(I)	1.8	0.79	ACU(T)	1.6	0.54	AAU(N)	3.6	0.85	AGU(S)	2.6	1.07
AUC(I)	3.5	1.55	ACC(T)	2.9	0.97	AAC(N)	4.9	1.15	AGC(S)	1.7	0.68
AUA(I)	1.5	0.66	ACA(T)	3.3	1.13	AAA(K)	6.8	1.32	AGA(R)	2.2	0.65
AUG(M)	2.9	1	ACG(T)	4	1.36	AAG(K)	3.5	0.68	AGG(R)	3.4	0.99
GUU(V)	2.1	0.79	GCU(A)	2.5	0.68	GAU(D)	2.4	0.89	GGU(G)	3.3	0.72
GUC(V)	3.3	1.25	GCC(A)	5.3	1.41	GAC(D)	3	1.11	GGC(G)	3.5	0.77
GUA(V)	1.5	0.58	GCA(A)	2.8	0.75	GAA(E)	3.8	1.06	GGA(G)	2.8	0.61
GUG(V)	3.6	1.37	GCG(A)	4.3	1.16	GAG(E)	3.4	0.94	GGG(G)	8.8	1.91

All frequencies are averages over all taxa

Table 4 Rare codon analysis

S.O.V	Actual value	Ideal value	Terms
CAI	0.64 ↓	0.8–1.0	A CAI of 1.0 is considered ideal The lower the number, the higher the chance that your gene will be expressed poorly GenScript's OptimumGene™ codon optimization tool can typically improve your sequence
GC Content	55.5%	30–70%	The ideal percentage range of GC content is between 30 and 70%
CFD	13%	< 30%	This un-optimized gene employs tandem rare codons that can reduce the efficiency of translation or even disengage the translational machinery GenScript's OptimumGene™ can give you the option to solve this problem

CAI codon Adaptation Index

Discussion

In this study, the value of dN/dS was 0.45 which was less than one indicating that the pure selection has occurred on the desired gene without any key changes and also it was similar to previous research (Forni et al. 2013; Liu et al. 2014). The number of haplotypes was 22 that was similar to previous study (Goka et al. 2009) though, the number of haplotypes varies depending on the species and herbaceous species, in a study of four haplotypes in the meadow grasshopper *Chorthippus parallelus* (Parkin & Butlin 2004), six haplotypes in *Undaria pinnatifida* (Uwai et al. 2006), 26 haplotypes in *Amphibian chytridiomycosis* (Goka et al. 2009), 74 in *Symbiodinium* (Rodriguez-Lanetty 2003). Up to 135 haplotypes have been reported in

Trichoderma and *Hypocrea* (Druzhinina et al. 2005). We reported the GC content value of 55.5% vis-à-vis. The ideal percentage range of GC content is between 30 and 70% (Fullerton et al. 2001).

The cluster analysis based on the ITS marker could not distinguish between different populations of Mango based on geographic regions. These results were similar to the results of Azimzadeh et al. (Azimzade et al. 2014) in the ITS study of *Bunium persicum* ecotypes. The placement of samples collected from different geographic regions in the same clusters could be a sign of genetic similarity or the physical exchange of seed among different regions, but the interpretation based on physical interactions is more justifiable than genetic similarities (Bahari et al. 2015).

Table 5 Distance Matrix of ITS region sequence variation among genotypes of *M. indica* L. landraces

Genotypes	1	2	3	4	5	6	7	8	9	10
2	0.640									
3	0.811	0.149								
4	0.357	0.685	0.650							
5	0.641	0.090	0.141	0.681						
6	0.686	0.104	0.123	0.714	0.108					
7	0.759	0.123	0.026	0.643	0.115	0.097				
8	0.176	0.753	0.871	0.257	0.752	0.806	0.850			
9	0.243	0.784	0.974	0.430	0.740	0.860	0.919	0.234		
10	0.176	0.832	1.073	0.371	0.832	0.934	0.988	0.097	0.197	
11	0.180	0.796	0.911	0.317	0.804	0.861	0.881	0.108	0.243	0.097
12	0.176	0.803	1.002	0.342	0.843	0.924	0.958	0.104	0.230	0.059
13	0.180	0.816	0.979	0.331	0.771	0.874	0.924	0.093	0.230	0.083
14	0.156	0.753	0.900	0.307	0.751	0.805	0.850	0.086	0.252	0.101
15	0.636	0.363	0.421	0.743	0.348	0.374	0.410	0.845	0.692	0.827
16	0.213	0.828	0.923	0.311	0.865	0.887	0.903	0.141	0.312	0.126
17	0.160	0.696	0.761	0.226	0.694	0.720	0.745	0.097	0.274	0.145
18	0.695	0.201	0.122	0.621	0.188	0.180	0.137	0.790	0.838	0.926
19	0.265	0.823	0.878	0.256	0.793	0.849	0.857	0.164	0.321	0.168
20	0.168	0.831	1.022	0.336	0.831	0.911	0.964	0.097	0.205	0.052
21	0.152	0.694	0.852	0.275	0.711	0.735	0.806	0.079	0.274	0.111
22	0.226	0.926	1.193	0.410	0.864	0.982	1.079	0.192	0.262	0.134

Genotypes	11	12	13	14	15	16	17	18	19	20	21
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12	0.087										
13	0.086	0.076									
14	0.079	0.097	0.101								
15	0.816	0.867	0.781	0.824							
16	0.123	0.119	0.119	0.126	0.868						
17	0.115	0.134	0.123	0.111	0.771	0.104					
18	0.834	0.907	0.847	0.843	0.260	0.867	0.722				
19	0.161	0.146	0.153	0.185	0.776	0.209	0.189	0.785			
20	0.055	0.052	0.083	0.072	0.845	0.108	0.122	0.904	0.153		
21	0.108	0.111	0.093	0.119	0.764	0.130	0.104	0.739	0.176	0.111	
22	0.177	0.157	0.145	0.188	0.832	0.193	0.213	1.013	0.257	0.145	0.180

Some mango samples collected from different geographical regions being in one group (cluster) indicated that the genetic similarity among them, may be because of

the physical interchange of seeds between different regions, whose interpretation based on physical interactions is more justifiable than genetic similarities (Ninou et al. 2017).

Fig. 1 Cluster analysis of the genotypes of Mango based on ITS sequences. The sequences were aligned and clustered and their similarity and genetic distances were defined using the using Clustal W and NJ methods by using DNASTAR software and MEGA6 software

Table 6 Distance matrix of ITS region sequence variation among genotypes of Mango based on cities that were collected

Province	Genetic distance
Manojan	0.371
Rodan	0.456
Jiroft	0.324
Fariab	0.321
Kahnoj	0.121
Minab	0.111

ITS marker could not distinguish populations of different mangoes based on geographic region at species level but it could segregate all plants (the sequences of Iranian *Mangifera indica* L. landraces and some sequences of other species of *Anacardiaceae* family in GenBank) at genus level (or interspecies) (Fig. 2). Thus, ITS marker is an appropriate tool for evaluations of interspecies. These results are in accordance with the observations reported in other

Fig. 2 Cluster analysis of the genotypes of Mango with the sequences of Mango species of *Anacardiaceae* ITS available at NCBI were compared following the Muscle and UPGMA methods by using MEGA6 software

studies (Varela et al. 2004; Wu et al. 2013; Haidari et al. 2014). The inter and intra species genetic diversity is an important source of typology and is one of the important indices of biodiversity. Therefore, it is important to study biodiversity in such a way that the genetic diversity in inter and intra species among plants.

One of the reasons for the non-segregation of Mango genotypes in this research is the excessive conservation of the sequence of ribosomal genes. This reason has been confirmed by Tajima's D, Chi square because genetic differentiation by Chi square estimates and it is conducted

that *P* value of Chi square was 0.3605 and non-significant and also nucleotide diversity, given that average number of nucleotide differences ($k = 98.8$) but nucleotide diversity (per site) ($\pi = 0.30$). Low nucleotide diversity could suggest that several kind of polymorphisms is out of conserved sequence among different Mango species because in earlier reports, analysis of ITS region of nrDNA to the phylogenetic relationships among 14 *Mangifera* L. species of Thailand and then parsimony and neighbor-joining (NJ) analyses revealed that several species have nucleotide additive (two different nucleotides at the same locus) at

several sites in the ITS region. Also, *M. indica* had several polymorphisms among cultivars. This finding may suggest a possibility of the hybrid origin of *Mangifera* species, although *Mangifera* species are all assumed to be diploid having chromosome number of $2n = 2x = 40$ (Yonemori et al. 2002). Genomic variability has been observed among different mango species using different markers and as a result some reports identified low polymorphism and others observed high polymorphism between different mango genotypes studied (Gitahi et al. 2016). For example, Analysis of molecular variance (AMOVA), for RFLP, ITS, SSRs, AFLP, direct sequence data, and phylogenetic trees indicated high variation (97%) between groups, but low (3%) within groups. Karihaloo et al. (2003) reported that 94.7% of the genetic diversity in Mango existed within regions. Kumar et al. (2001) reported a moderate degree of genetic diversity among the mango cultivars. Other studies reported that similarity values ranged between 2 and 70% with an average of 34% (Damodaran et al. 2012).

Since Iran is considered as one of the richest countries in terms of natural resources due to its high climate diversity and herbaceous genetic resources, the study of genetic diversity in breeding programs plays an important role in plant improvement and will greatly help to advance the research programs (Dehdashtian et al. 2011). Result of present study indicated that a high genetic distance exists between HajiGholam (Manojan) and Arbabi (Rodan) genotypes and showed higher genetic diversity in mango of Rodan region. It suggested that for programs of successful breeding in mango, the genotypes of Rodan region especially the genotype of Arbabi can be used as a genetic treasury and as a gene donor in the breeding programs.

The codon usage in mango ITS genes is predominantly determined by mutation bias based on the Tajima's D and the FU and LI tests (Fu and Li 1993). If these values are large and positive, it has long branches close to the root and also there is an excess of old mutations or a reduction of young mutations. In this research, Tajima's D was positive and non-significant with a value of more than one (1.801413) (Table 4) and phylogenetic tree of mango ITS gene family has long branches close to the root (Fig. 2).

Given that the ribosomal genes have a high percentage of the conserved sequence, it is expected that the least mutation will occur in these parts, so based on the result of Tajima's D that it was positive and non-significant with a value of more than one, it suggests that new mutations do not occur (Fu and Li 1993), therefore, it proves that the sequences of ribosomal gene family are conserved.

The results of this study showed high polymorphism among Iranian Mango genotypes and it reports highest rate of the regional diversity among the Mango of Rodan region. These are in line with the previous reports (Fazeli-Nasab and Fahmide 2018). According to Vavilov's theory,

the origin or diversity center of plants belong to the centers that have the most variety and diversity (Adams et al. 1991). Therefore, the region of Rodan can be identified as the diversity center of Mango in Iran.

Conclusion

The results of this study indicated the existence of high genetic distance between HajiGholam (Manojan) and Arbabi (Rodan) genotypes. It further suggests that for a successful breeding in mango, the genotypes of Rodan region especially the genotype of Arbabi can be used as a genetic treasure and as a gene donor and that ITS is the best suitable tool for evaluations of inter and intra species genetic diversity.

Acknowledgements HE would like to thank the MOE and UTM-RMC (Malaysia) for support through the HICoE grant No. RJ130000.7846.4J262. RZS would like to extend his sincere appreciation to Dr. Rajeev K Varshney, Director, Center of Excellence in Genomics and Systems Biology, ICRISAT, Patancheru - 502324, Hyderabad, India for reviewing and editing the paper.

Compliance with ethical standards

Human and animal rights statements This article does not contain any studies with human or animal subjects.

References

- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252(5013):1651–1656
- Ansari S, Solouki M, Fakheri B, Fazeli-Nasab B, Mahdinezhad N (2018) Assessment of molecular diversity of internal transcribed spacer region in some lines and landrace of Persian Clover (*Trifolium resupinatum* L.). *Potravin Slovak J Food Sci* 12(1):657–666
- Azimzade M, Amir R, Osare MH, Bihanta MR, Frotan M (2014) Evaluation of genetic diversity of Iranian Cumin (*Bunium persicum* Boiss) of nuclear ribosomal DNA using ITS. *J Gen Res Plant Breed Past For* 22(1):1–10
- Bahari Z, Shojaeiyan A, Rashidi Monfared S, Mirshekari A, Nasiri K, Amiryan M (2015) Investigation of genetic diversity among some Iranian Dill (*Anethum graveolens* L.) landraces, using ISSR markers. *Plant Gen Resour-C J* 2(1):11–22
- Damodaran T, Kannan R, Ahmed I, Srivastava R, Rai R, Umamaheshwari S (2012) Assessing genetic relationships among mango (*Mangifera indica* L) accessions of Andaman Islands using inter simple sequence repeat markers. *N Z J Crop Hortic* 40(4):229–240
- Dehdashtian Z, Mr Wahabi, Fazilati M, Ghaedi K, Salamian A (2011) Analysis of the genetic diversity of *Astragalus gossypinus* population in Isfahan. *Genet 3rd Milenn* 9(3):2474–2480
- Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA miniprep: version II. *Plant Mol Biol Rep* 1(4):19–21
- Druzhinina IS, Kopchinskiy AG, Komoń M, Bissett J, Szakacs G, Kubicek CP (2005) An oligonucleotide barcode for species

- identification in *Trichoderma* and *Hypocrea*. Fungal Genet Biol 42(10):813–828
- Emmami S (2001) Culture and growing of Mango. The Production Office of Promotion Programs and Technical Press, p 353
- Fazeli-Nasab B, Fahmide L (2018) Evaluation of antioxidant properties and phenolic compounds of different mango (*Mangifera indica* L.) genotypes in Kerman and Hormozgan. J Crop Sci Res Arid Regions (in press)
- Forni D, Cagliani R, Pozzoli U, Colleoni M, Riva S, Biasin M, Filippi G, De Gioia L, Gnudi F, Comi GP (2013) A 175 million year history of T cell regulatory molecules reveals widespread selection, with adaptive evolution of disease alleles. Immunity 38(6):1129–1141
- Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. Genetics 133(3):693–709
- Fullerton SM, Bernardo Carvalho A, Clark AG (2001) Local rates of recombination are positively correlated with GC content in the human genome. Mol Biol Evol 18(6):1139–1142
- Gitahi R, Kasili R, Kyallo M, Kehlenbeck K (2016) Diversity of threatened local mango landraces on smallholder farms in Eastern Kenya. For Trees Livelihoods 25(4):239–254
- Goka K, Yokoyama J, Une Y, Kuroki T, Suzuki K, Nakahara M, Kobayashi A, Inaba S, Mizutani T, Hyatt AD (2009) Amphibian chytridiomycosis in Japan: distribution, haplotypes and possible route of entry into Japan. Mol Ecol 18(23):4757–4774
- Haidari P, Mehrabi AA, Nasrollah Nejad Ghomi AA (2014) Genetic diversity of Balm (*Melissa officinalis* L.) landraces and genetic relationship within and between them using ITS markers. J Crop Breed 6(13):29–39
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: Nucleic acids symposium series. Information Retrieval Ltd., London, c1979–c2000, pp 95–98
- Henry T, Iwen PC, Hinrichs SH (2000) Identification of *Aspergillus* species using internal transcribed spacer regions 1 and 2. J Clin Microbiol 38(4):1510–1515
- Hidayat T, Arif SM, Samad AA (2013) Molecular biodiversity of selected mango cultivars based on DNA sequences of internal transcribed spacer region. Pak J Biol Sci 16(19):1072–1075
- Jani Pour L, Fahmideh I, Fazeli-Nasab B (2018) Genetic assessment of some populations of the medicinal plant Caraway (*Carum carvi*) using RAPD and ISSR markers. J Iran Plant Ecophysiol Res 12(48):78–91
- Karihaloo J, Dwivedi Y, Archak S, Gaikwad AB (2003) Analysis of genetic diversity of Indian mango cultivars using RAPD markers. J Hortic Sci Biotechnol 78(3):285–289
- Kumar H, Narayanaswamy P, Prasad T, Mukunda G, Sondur S (2001) Estimation of genetic diversity of commercial mango (*Mangifera indica* L.) cultivars using RAPD markers. J Hortic Sci Biotechnol 76(5):529–533
- Lal S, Singh AK, Singh SK, Srivastav M, Singh BP, Sharma N, Singh NK (2017) Association analysis for pomological traits in mango (*Mangifera indica* L.) by genic-SSR markers. Trees 31(5):1391–1409
- Litz RE (2009) The mango: botany, production and uses. CABI, Wallingford
- Liu F, Hu Y, Wang Q, Li HM, Gao GF, Liu CH, Zhu B (2014) Comparative genomic analysis of *Mycobacterium tuberculosis* clinical isolates. BMC Genom 15(1):469
- McKain MR, Johnson MG, Uribe Convers S, Eaton D, Yang Y (2018) Practical considerations for plant phylogenomics. Appl Plant Sci 6(3):e1038
- Naghavi MR, Mardi M, Ramshini HA, Fazeli-Nasab B (2004) Comparative analyses of the genetic diversity among bread wheat genotypes based on RAPD and SSR markers. Iran J Biotechnol 2(3):195–202
- Nashima K, Terakami S, Kuniyama M, Nishitani C, Shoda M, Matsumura M, Onoue-Makishi Y, Urasaki N, Tarora K, Ogata T (2017) Retrotransposon-based insertion polymorphism markers in mango. Tree Genet Genomes 13(5):110
- Ninou EG, Mylonas IG, Tsivelikas AL, Ralli PE (2017) Phenotypic diversity of Greek dill (*Anethum graveolens* L.) landraces. Acta Agric Scand B-Soil Plant 67(4):318–325
- Parkin EJ, Butlin RK (2004) Within-and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. Mol Biol Evol 21(8):1595–1601
- Ravishanker K, Padmakar B, Lavanya B, Mani B, Dinesh M (2017) Development and characterization of microsatellite loci from mango (*Mangifera indica* L.). Indian J Biotechnol 6(2):250–253
- Robinson J, Harris S, Juniper B (2001) Taxonomy of the genus *Malus* Mill. (*Rosaceae*) with emphasis on the cultivated apple, *Malus domestica* Borkh. Plant Syst Evol 226(1):35–58
- Rodriguez-Lanetty M (2003) Evolving lineages of *Symbiodinium*-like dinoflagellates based on ITS1 rDNA. Mol Phylogenet Evol 28(1):152–168
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19(18):2496–2497
- Sargazi A, Fakheri B, Soloki M, Fazeli-Nasab B (2016) Genetic diversity of some population of medicinal Ajowan (*Carum copticum*) using RAPD marker. J Med Plants Biotechnol 2(3):22–36
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci 109(16):6241–6246
- Sharp PM, Li W-H (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucl Acids Res 14(19):7737–7749
- Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141(1):413–429
- Singh Z, Singh SP (2012) Mango. Crop Post-harvest: Sci Technol 108–142
- Swaroop A, Stohs SJ, Bagchi M, Moriyama H, Bagchi D (2018) Mango (*Mangifera indica* Linn) and anti-inflammatory benefits: versatile roles in mitochondrial bio-energetics and exercise physiology. Funct Food Health Dis 8(5):267–279
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585–595
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10(3):512–526
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12):2725–2729
- Tiwari DK, Patel V, Pandey A (2018) Floral induction in mango: physiological, biochemical and molecular basis. IJCS 6(1):252–259
- Uwai S, Yotsukura N, Serisawa Y, Muraoka D, Hiraoka M, Kogame K (2006) Intraspecific genetic diversity of *Undaria pinnatifida* in Japan, based on the mitochondrial *cox3* gene and the ITS1 of nrDNA. Hydrobiologia 553(1):345–356
- Varela ES, Lima JP, Galdino AS, Pinto LdS, Bezerra WM, Nunes EP, Alves MA, Grangeiro TB (2004) Relationships in subtribe Diocleinae (*Leguminosae*; *Papilionoideae*) inferred from

- internal transcribed spacer sequences from nuclear ribosomal DNA. *Phytochemistry* 65(1):59–69
- Watterson G (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7(2):256–276
- Wu C-T, Hsieh C-C, Lin W-C, Tang C-Y, Yang C-H, Huang Y-C, Ko Y-J (2013) Internal transcribed spacer sequence-based identification and phylogenetic relationship of I-Tiao-Gung originating from *Flemingia* and *Glycine* (*Leguminosae*) in Taiwan. *J Food Drug Anal* 21(4):356–362
- Yonemori K, Honsho C, Kanzaki S, Eiadthong W, Sugiura A (2002) Phylogenetic relationships of *Mangifera* species revealed by ITS sequences of nuclear ribosomal DNA and a possibility of their hybrid origin. *J Syst Evol* 231(1–4):59–75
- Yousef EA, Mueller T, Börner A, Schmid KJ (2018) Comparative analysis of genetic diversity and differentiation of cauliflower (*Brassica oleracea* var. botrytis) accessions from two ex situ genebanks. *PLoS ONE* 13(2):0192062

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.