Indian Journal of Agricultural Sciences **93** (7): 795–800, July 2023/Short Communication
https://doi.org/10.56093/ijas.v93i7.130701

Highly informative simple sequences repeat (SSR) markers reveal the large genetic diversity of mango (*Mangifera indica*) germplasm in China

YUJUAN TANG[1], SHIXING LUO[1], YU ZHANG[1], LIMEI GUO[1], YING ZHAO[1], RIWANG LI[1], ENLIANG SONG[1] and GUODI HUANG[1]*

Guangxi Subtropical Crops Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China 530 001.

Received: 25 November 2022; Accepted: 22 June 2023

Keywords: SSR markers, Genetic relationship, Geographical origins, Polymorphism, PIC

Mango (*Mangifera indica* L.) is a tall, evergreen tree belonging to genus *Mangifera* in family Anacardiaceae. Due to its extensive cultivation range, high nutritional value, attractive appearance and distinctive flavour, it is regarded as the king of tropical fruits (Lamba *et al.* 2021) and is adored by consumers (Tharanathan *et al.* 2006). Mango is cultivated extensively in tropical and subtropical climates across the globe, as well as in a variety of marginal locations. India, China, Thailand, Mexico, Indonesia and Pakistan are the five largest producers of mango (Mitra 2014). There are 69 mango species in the globe, and simple sequence repeats (SSRs) have been used in some studies on mango regional genetic diversity, including Iranian genotypes (Shamili 2012), Nigerian genotypes (Ajayi *et al.* 2019), Indian genotypes (Ravishankar *et al.* 2015) and others. Although practical SSR markers have been developed, information on mango germplasm resources in China and around the world is still scarce.

The mango varieties cultivated in Chinese production belong to *M. indica* while *M. persiciformis* acts as a traditional Chinese medicine that has antitussive, antiasthmatic and expectorant effects (Baughman 2022). Phylogenetic analysis of *Mangifera* species has been a popular research topic, and some genetic relationships have only been deduced from the whole chloroplast genome sequences using a limited number of genotypes (Niu *et al.* 2021). Therefore, larger-scale sampling is required to better understand the diversity of *Mangifera*. To gain information for cultivar identification and variety in mango genetic resources, we investigated the genetic diversity and relatedness of 188 mango accessions that represent nearly all mango collections in China and other countries, utilizing 40 polymorphic SSR markers. Phylogeographic relationships were discussed in greater detail compared to prior research.

In this study, leaves of 188 mangoes came from 20 countries or regions planted in Nanning city (22°8′N, 108°3′E), Guangxi province, China, including 136 *M. indica*, 48 *M. persiciformis*, 3 *M. sylvatica* and 1 *M. hiemalis* were collected in 2021 (Supplementary Table 1). Total genomic DNA was extracted from young leaves as described by Murray and Thompson (1980). A total of 40 previously published SSR markers were used for PCR in this analysis (Duval *et al.* 2005, Schnell *et al.* 2005). The allele frequency of each accession and the total number of alleles for each SSR locus were determined. The genetic data was evaluated using the following parameters: number of alleles per locus (A), effective number of alleles ($Ne = 1/1-He$), observed heterozygosity (Ho, direct count), expected heterozygosity ($He = 1-\Sigma p^2_i$, where p_i is the frequency of the ith allele), Wright's fixation index (F = 1−Ho/He), and the probability of identity ($PI = 1-\Sigma p^4 i + \Sigma\Sigma(2p_ip_j)^2$, where p_i and p_j are the frequencies of the ith and jth alleles, respectively). ARLEQUIN version 3.01 was utilized to compute A, Ho and He. The program POPGENE 1.32 was used to calculate Ne and F. Analyses of polymorphism information content (PIC) values were performed according to Dillon *et al.* (2013). Shannon's index (I) was computed using Identity 1.0. (Shamili *et al.* 2012). An unrooted dendrogram was created using the neighbor-joining (NJ) approach. Distance calculations, tree construction and bootstrapping were performed in PowerMarker V3.0. For structure analysis, the most likely number of clusters (K) was estimated considering the plateau criterion using the nonparametric Wilcoxon test and ΔK method.

Across 188 mango genotypes, a total of 303 alleles were discovered at the loci of 40 microsatellite markers. The number of alleles per locus produced by each marker ranged from 3–18 alleles per locus, with an average number of 7.58 (Table 1), which was similar to the 6.0 alleles per locus reported by Yamanaka *et al.* (2019). This implies that

[1]Guangxi Subtropical Crops Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China. *Corresponding author email: gxhgd@126.com

Table 1 Genetic parameters and differentiation coefficient generated by 40 SSR markers on 188 mango genotypes

Marker	Na	Ne	I	Ho	He	PIC	Fis	Fit	Fst	Nm
M101-1	17	6.202	2.156	0.27	0.839	0.823	0.459	0.574	0.213	0.925
M104-1	8	3.677	1.518	0.756	0.728	0.689	-0.143	-0.01	0.113	1.957
M105-1	13	6.464	2.1	0.844	0.845	0.829	-0.136	0.01	0.128	1.7
M108-1	8	5.503	1.851	0.569	0.818	0.795	0.024	0.206	0.186	1.096
M109-1	10	3.575	1.551	0.47	0.72	0.68	0.123	0.284	0.183	1.116
M118-1	8	2.489	1.282	0.27	0.598	0.566	0.484	0.56	0.147	1.445
M129-1	11	5.044	1.906	0.669	0.802	0.781	0.093	0.233	0.154	1.374
M130-1	5	1.65	0.78	0.089	0.394	0.366	0.797	0.858	0.299	0.586
M132-1	6	2.343	1.018	0.191	0.573	0.503	0.479	0.549	0.135	1.607
M133-1	18	4.557	1.957	0.818	0.781	0.759	-0.139	0.002	0.124	1.767
M136-1	11	3.215	1.61	0.63	0.689	0.667	-0.034	0.152	0.18	1.14
M138-1	8	4.436	1.635	0.691	0.775	0.74	-0.054	0.091	0.138	1.567
M141-1	9	3.607	1.599	0.676	0.723	0.694	-0.12	0.012	0.118	1.86
M142-1	11	4.995	1.83	0.777	0.8	0.775	-0.085	0.116	0.185	1.098
M144-1	6	2.722	1.197	0.58	0.633	0.563	-0.089	-0.002	0.08	2.871
M148-1	5	3.074	1.265	0.652	0.675	0.623	-0.161	-0.048	0.097	2.32
M152-1	6	2.043	0.986	0.53	0.511	0.466	-0.135	0.026	0.141	1.519
M155-1	6	1.936	1.055	0.503	0.483	0.462	-0.298	-0.158	0.108	2.068
M18-1	5	2.858	1.19	0.642	0.65	0.586	-0.072	-0.001	0.066	3.538
M53-1	3	2.77	1.058	0.983	0.639	0.565	-0.709	-0.569	0.082	2.783
M57-1	6	5.189	1.709	0.464	0.807	0.779	0.279	0.39	0.154	1.862
M72-1	17	5.684	2.095	0.874	0.824	0.807	-0.193	-0.052	0.118	3.514
M83-1	4	1.424	0.538	0.309	0.298	0.264	-0.109	-0.035	0.066	0.868
M90-1	6	3.435	1.425	0.213	0.709	0.665	0.588	0.68	0.224	1.367
M92-1	1	4.739	1.934	0.403	0.789	0.768	0.499	0.577	0.155	1.205
M93-1	9	4.11	1.592	0.613	0.757	0.721	0.049	0.212	0.172	0.979
M95-1	13	5.317	1.978	0.746	0.812	0.791	0.003	0.206	0.203	1.695
mangoES1	4	1.448	0.597	0.293	0.309	0.281	0.184	0.289	0.129	2.14
mangoES126	10	3.04	1.501	0.47	0.671	0.642	-0.215	-0.088	0.105	2.17
mangoES170	5	3.048	1.259	0.724	0.672	0.621	0.123	0.214	0.103	2.17
mangoES35	4	1.225	0.409	0.166	0.184	0.176	0.011	0.129	0.12	1.842
MCR11	3	1.531	0.556	0.309	0.347	0.291	0.014	0.088	0.075	3.093
MCR169	5	2.518	1.03	0.337	0.603	0.52	0.308	0.473	0.238	0.802
MCR22	4	2.934	1.113	0.691	0.659	0.587	-0.272	-0.078	0.152	1.395
MCR220	4	2.482	0.989	0.611	0.597	0.512	-0.253	-0.102	0.121	1.82
MCR303	4	2.036	0.744	0.489	0.509	0.388	0.1	0.162	0.069	3.35
MCR360	3	1.394	0.537	0.177	0.283	0.26	0.218	0.321	0.132	1.65
MCR380	3	1.289	0.408	0.156	0.224	0.201	0.054	0.478	0.449	0.307
MCR39	5	3.029	1.251	0.689	0.67	0.62	-0.246	-0.068	0.143	1.494
MCR55	6	2.18	0.965	0.497	0.541	0.485	0	0.193	0.193	1.045
Average	7.575	3.28	1.304	0.52	0.624	0.583	0.036	0.172	0.15	1.708

Note: Na, number of alleles; Ne, effective number of alleles; I, Shannon's index; Ho, observed heterozygosity; He, expected heterozygosity; PIC, polymorphic information content; Fis, Wright's fixation index; Fst, F-statistics; Fit, inbreeding coefficient; Nm, the migration number per generation.

there was enough allelic variation, which was required for assessing genetic diversity. The average number of alleles found in this study was much more than the 4.37, 2.7 and 2.5 alleles per locus found by Dillon *et al.* (2014), Azmat *et al.* (2016) and Ajayi *et al.* (2019), respectively. Variability in the number of alleles discovered per locus might be attributable to the utilization of varied genotypes. The number of effective alleles per locus ranged from 1.225–6.464 with a mean of 3.280 (Table 1), which was higher than the 1.51 previously reported by Shamili *et al.* (2012). The mean observed heterozygosity (Ho) of the genotypes was 0.52. A high level of Ho has also been reported in another study on mango (Shamili *et al.* 2012), and this could be attributed to its allogamous mode of reproduction. The average Ho was below the average expected heterozygosity (He) in the present study, which was consistent with Dillon's report. This indicated a tendency toward inbreeding, most likely due to population isolation (Dillon *et al.* 2014).

PIC is the proportion of heterozygosity within and between genotypes based on variations in allele frequencies. Over 60% of the primers examined in this study were highly polymorphic, with a mean PIC value of 0.583, suggesting that the chosen microsatellites were quite informative in separating the tested genotypes. Kumar *et al.* (2013) noted a comparable PIC value of 0.552 among mango genotypes. Furthermore, the present study reported a much higher mean PIC value than the 0.03 reported by Ajayi *et al.* (2019). The low level of genetic diversity was ascribed to the frequent use of a few parents in breeding and the narrow genetic base among selected mango varieties in Oyo State. Therefore, the differences in PIC values were closely linked to the selection of different markers and the diversity of test genotypes.

Shannon's index (I) represents variety differentiation among the collections of germplasm. The genetic diversity increases as the indices rise. Correspondingly, I was 1.304 (Table 1) in our study, which was higher than the findings of Jena and Chand (2021), who reported an I of 0.45 in a microsatellite-based study that involved 70 Indian mango accessions. The high genetic diversity observed in this study could be attributed to the worldwide origins of the germplasm collection.

Wright's F-statistics, including the fixation index (Fis), Fit and Fst, calculated for all mango populations were 0.036, 0.172 and 0.151, respectively. The mean value of the Fis of all the loci was low, revealing a deficit in heterozygosity in the overall population. Nineteen loci showed excess heterozygosity with negative Fis values (Table 1). The estimated F-statistic, Fst, which represents genetic variation among populations, varied from 0.066 at M18-1 to 0.449 at MCR380. The mean Fst value of 0.151 suggested that there was high genetic diversity among the populations. The number per migrant generation (Nm) was used to measure the gene flow between subpopulations. In the present study, Nm = 1.708 > 1, which indicated that there was a high level of gene flow between genotypes.

PCoA based on a genetic distance matrix with data normalization revealed a significant distinction between *M. indica* and *M. persiciformis*, despite the presence of

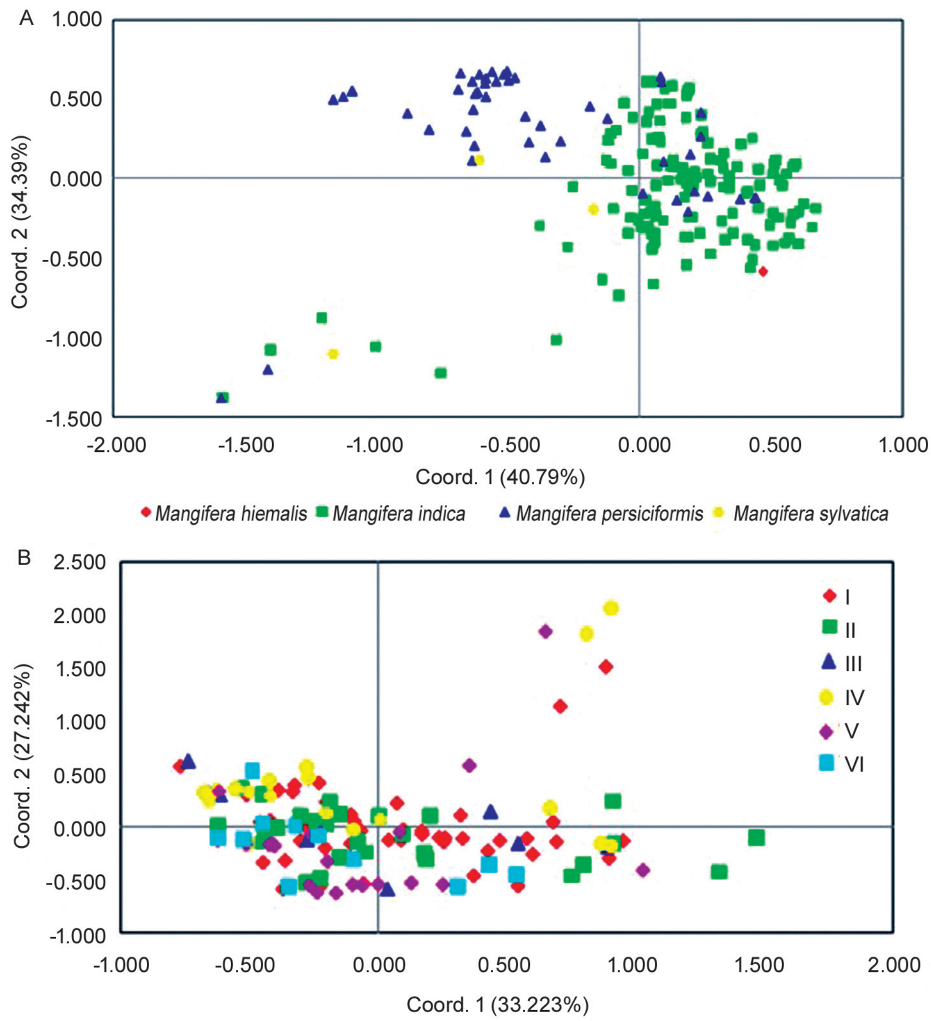

Fig 1 Principal component analysis of the 188 mango genotypes based on (A), four species and of 139 genotypes of *M. indica* from different regions; (B), Group I to VI were divided as follows: I, Yunnan and Guangxi; II, Malaysia and India; III, Hainan and Taiwan; IV, Thailand and Vietnam; V, Australia and the USA; and VI, the other regions.

certain overlapping zones (Fig 1A). The first axis explained 40.79% of the overall variance for SSRs, and the second axis explained 34.39%. We further employed 136 genotypes of *M. indica* and the genotypes were divided into six groups, mainly according to geolocation (Fig 1B). The six groups were as follows: group I, Yunnan and Guangxi; group II, Malaysia and India; group III, Hainan and Taiwan; group IV, Thailand and Vietnam; group V, Australia and the USA; and group VI, the other regions. Overall, the scatter plot showed that all 136 accessions were dispersed evenly, indicating that genetic resources in China have a certain amount of genetic variation in terms of SSR markers.

The SSR dataset was utilized for the structure implementation of the model-based Bayesian clustering approach. The value increased with increasing K value, but there was no sharp fluctuation (Fig 2A). Delta K had a maximum value of 95 at K = 3, which was much larger than the other K values (Fig 2B). Taken together, K = 3 was the most suitable group, and the 188 genotypes were grouped into three subpopulations (Fig 2C). Thirty-one mango genotypes, representing 16.5% of the population, were assigned to subpopulation 1 (Pop 1), and 27.7% and 55.8% of the population was grouped into subpopulation 2 (Pop 2) and subpopulation 3 (Pop 3), respectively. The results showed that Pop 1 contained *M. persiciformis* genotypes. Pop 2 mainly consisted of *M. indica* genotypes from GX and YN, while Pop 3 mainly contained *M. indica* genotypes from the USA and Southeast Asian countries.

The UPGMA dendrogram (Fig 3A), generated from the shared allele matrix, indicated that the studied germplasm was divided into two major clusters. The first cluster consisted of most genotypes of *M. persiciformis*, while the remaining genotypes of *M. indica* were grouped in the second cluster. There were several subclusters based on the values of genetic distance matrices in the second cluster. Nevertheless, the accessions were mixed. To better evaluate the relationship between genotypes from different regions, we constructed a phenogram of the 139 accessions of *M. indica* based on their origins (Fig 3B). The genotypes originating from GX, India, SC, Mya and YN were clustered

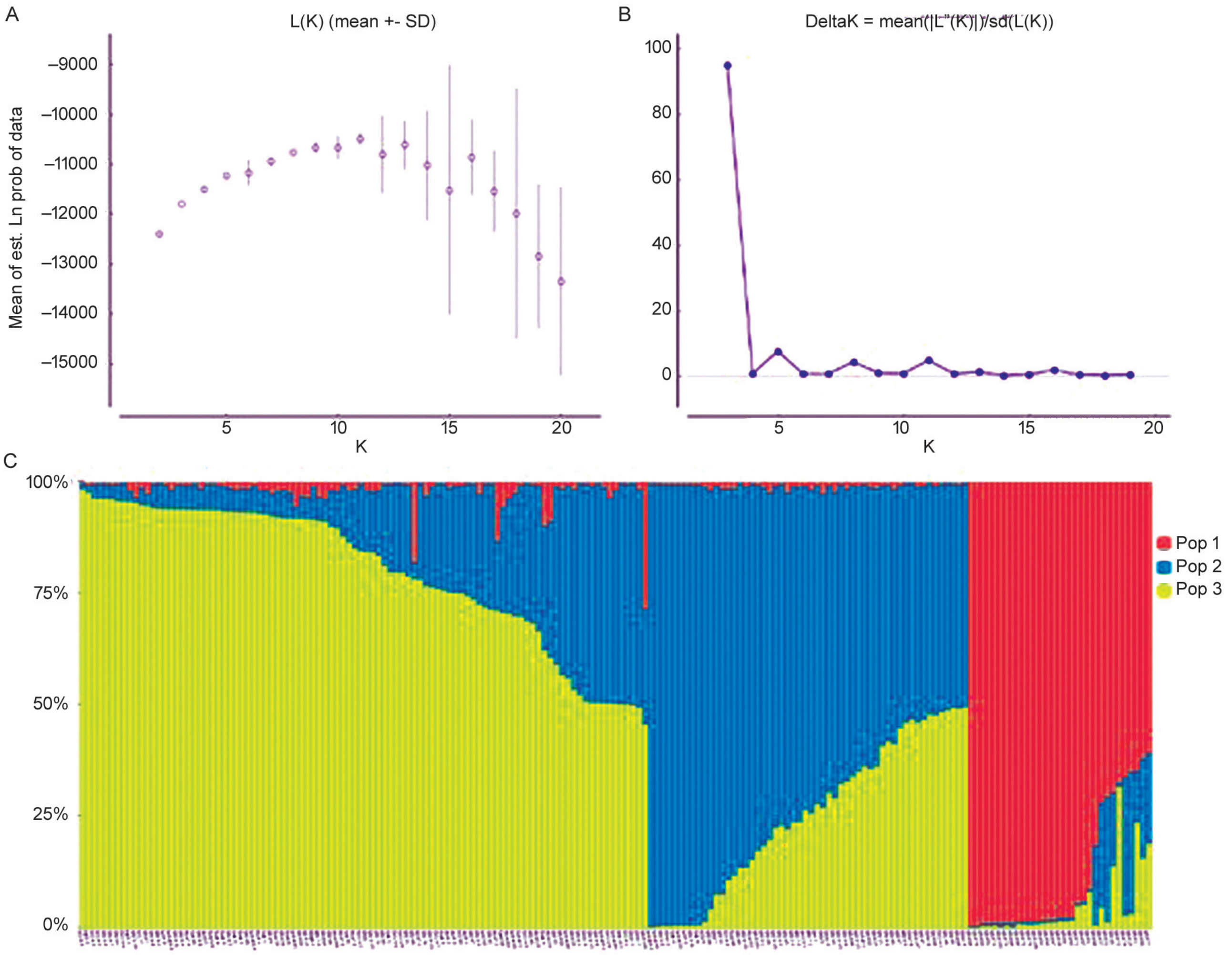

Fig 2 Results of Bayesian model-based clustering structure analysis of individuals of mango. (A), The probability of the data LnP(D) (±SD) against the number of K clusters; (B), Delta K value from the mean log-likelihood probabilities from structure runs where inferred clusters (K) ranged from 1–20 and; (C), Estimated genetic clustering (K=3) obtained with the population and black vertical line in the bar chart is population identifier.

together, which corresponded to Groups I and II in the PCoA (Fig 3B). The genotypes originating from Ind and HN and those from Vie and Tha clustered together, corresponded to Groups III and IV, respectively. Yamanaka *et al.* (2019) reported that accessions from Thailand and Vietnam grouped together, were consistent with our observation. Then, *M. persiciformis* from Guangxi (GXP) and Guizhou (GZP) clustered with the above mentioned *M. indica* genotypes, indicated a close relationship between *M. persiciformis* and *M. indica* from Asia. This may be due to the low degree of sequence variation in the chloroplast genome observed between *M. persiciformis* and *M. indica*, indicating that the two species of *Mangifera* were highly conserved (Niu *et al.* 2021). The phenogram results reflected the geographic distance between the different districts where the samples were collected.

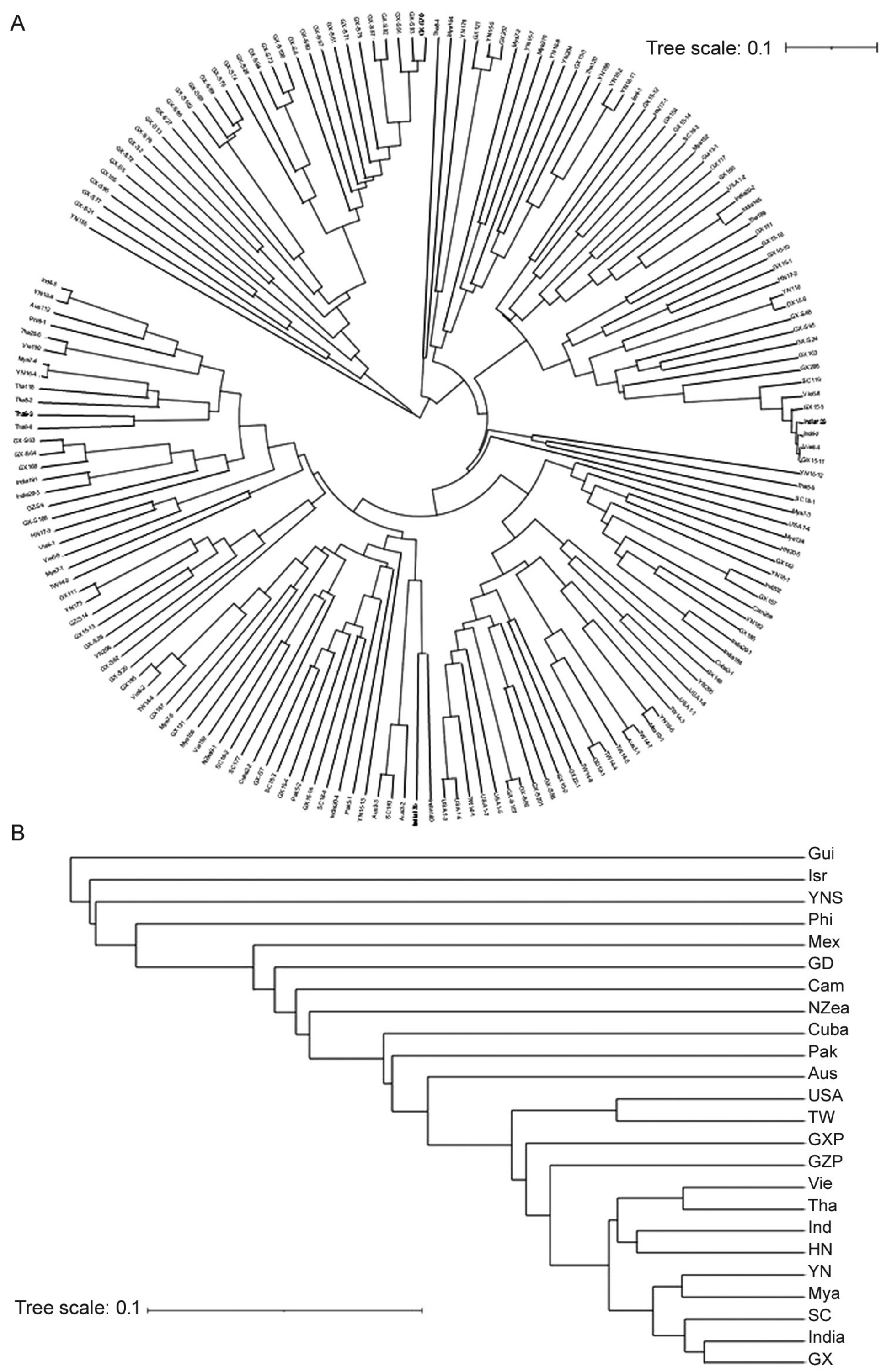

Fig 3 Genetic relationship among 188 mango genotypes basing on genotype (A) or region (B) using 40 SSR markers. The origin abbreviations are the origins shown in Supplementary Table 1.

SUMMARY

The genetic diversity of mango (*Mangifera indica* L.) was determined among 188 mango accessions using 40 SSR markers. A total of 303 alleles were discovered, with a mean value of 7.58 and an average PIC of 0.583, showing that the SSR markers utilized in this investigation was quite informative. High Shannon's index (1.304) and He (0.624) reflected the high genetic diversity of Chinese mango genetic resources. PCoA analysis and phenogram analysis divided the accessions broadly into groups representing their geographical origins and suggested a clear separation between *M. indica* and *M. persiciformis*. This expanded awareness of the genetic diversity of mango germplasm would aid breeders in choosing better parents, hence accelerating the delivery of improved cultivars to industry in order to satisfy consumer demand.

REFERENCES

Ajayi I, Olawuyi O, Ayodele A and Faneye A. 2019. Molecular relationship among *Mangifera indica* L. (Mango) varieties using simple sequence repeat (SSR) marker. *Journal of Advances in Biology and Biotechnology* **22**: 1–16.

Azmat M A, Khan A A, Khan I A, Rajwana I A, Cheema H M N and Khan A S. 2016. Morphological characterization and SSR based DNA fingerprinting of elite commercial mango cultivars. *Pakistan Journal of Agricultural Sciences* **53**.

Baughman M. 2022. 'Isolation and characterization of an immunomodulatory compound from a plant used in traditional chinese medicine'. PhD Thesis, Middle Tennessee State University, United States.

Dillon N L, Bally I S, Wright C L, Hucks L, Innes D J and Dietzgen R G. 2013. Genetic diversity of the Australian national mango genebank. *Scientia Horticulturae* **150**: 213–26.

Dillon N L, Innes D J, Bally I S, Wright C L, Devitt L C and Dietzgen R G. 2014. Expressed sequence tag-simple sequence repeat (EST-SSR) marker resources for diversity analysis of mango (*Mangifera indica* L.). *Diversity* **6**: 72–87.

Duval M F, Bunel J, Sitbon C and Risterucci A M. 2005. Development of microsatellite markers for mango (*Mangifera indica* L.). *Molecular Ecology Notes* **5**: 824–26.

Jena R C and Chand P K. 2021. Multiple DNA marker-assisted diversity analysis of Indian mango (*Mangifera indica* L.) populations. *Scientific Reports* **11**: 1–15.

Kumar M, Ponnuswami V, Nagarajan P, Jeyakumar P and Senthil N. 2013. Molecular characterization of ten mango cultivars using simple sequences repeat (SSR) markers. *African Journal of Biotechnology* **12**: 6568–73.

Lamba A, Kumar S and Prakash S. 2021. Pre-harvest application of chemicals and pesticide on mango and its effects on skin colour. *Journal of Pharmacognosy and Phytochemistry* **10**: 816–18.

Mitra S. 2014. *Mango Production in the World–Present Situation and Future Prospect*, pp. 287–96. XXIX International Horticultural Congress on Horticulture: Sustaining Lives, Livelihoods and Landscapes (IHC2014): IV 1111.

Murray M and Thompson W. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* **8**: 4321–26.

Niu Y, Gao C and Liu J. 2021. Comparative analysis of the complete plastid genomes of *Mangifera* species and gene transfer between plastid and mitochondrial genomes. *PeerJ* **9**: e10774.

Ravishankar K, Bommisetty P, Bajpai A, Srivastava N, Mani B, Vasugi C, Rajan S and Dinesh M. 2015. Genetic diversity and population structure analysis of mango (*Mangifera indica*) cultivars assessed by microsatellite markers. *Trees* **29**: 775–83.

Schnell R, Olano C, Quintanilla W and Meerow A. 2005. Isolation and characterization of 15 microsatellite loci from mango (*Mangifera indica* L.) and cross-species amplification in closely related taxa. *Molecular Ecology Notes* **5**: 625–27.

Shamili M, Fatahi R and Hormaza J. 2012. Characterization and evaluation of genetic diversity of Iranian mango (*Mangifera indica* L., Anacardiaceae) genotypes using microsatellites. *Scientia Horticulturae* **148**: 230–34.

Tharanathan R, Yashoda H and Prabha T. 2006. Mango (*Mangifera indica* L.), The king of fruits–An overview. *Food Reviews International* **22**: 95–123.

Yamanaka S, Hosaka F, Matsumura M, Onoue-Makishi Y, Nashima K, Urasaki N, Ogata T, Shoda M and Yamamoto T. 2019. Genetic diversity and relatedness of mango cultivars assessed by SSR markers. *Breeding Science* **69**: 332–44.