



# Genome-wide distribution of simple sequence repeats in pomegranate and their application to the analysis of genetic diversity

Chunyan Liu<sup>1,2</sup> · Jiyu Li<sup>1,2</sup> · Gaihua Qin<sup>1,2</sup>

Received: 9 September 2019 / Revised: 21 January 2020 / Accepted: 2 March 2020 / Published online: 23 March 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

Pomegranate (*Punica granatum* L.) germplasm collections are increasingly appreciated as a repository for the genetic improvement of species, and their genetic analysis is an essential prerequisite for their utilization in pomegranate breeding. Sufficient markers are a prerequisite for genetic analysis. In this study, we identified 36,792 simple sequence repeat (SSR) markers in the pomegranate genome. Dinucleotide repeats are the most abundant SSRs, and of the mono-, di-, tri-, and tetranucleotide repeats, A/T, AT/AT, AAT/ATT, and AAAT/ATTT were the most dominant motif type. There were only 1.80% SSR motifs located in the coding regions, while most (98.20%) SSR motifs were located in noncoding regions. We confirmed the reliability of 50 SSR markers by PCR amplification, from which 11 highly polymorphic SSR markers were selected to study the genetic diversity of 218 pomegranate accessions. A total of 63 alleles was detected for the 11 SSR primers, ranging from 2 (PG152) to 9 (PG093, PG080) across the pomegranate accessions. The polymorphism information content values of each SSR primer pair ranged from 0.22 to 0.70, with a mean value of 0.45. Based on population structure analysis, these accessions were clustered into three main populations and 42 accessions were chosen as the core set. The availability of these SSR markers and the clustering result for the pomegranate germplasm provide new important information for pomegranate genetics and breeding programs.

**Keywords** Pomegranate · Genome · Simple sequence repeat (SSR) · Genetic diversity · Molecular markers

## Introduction

Pomegranate (*Punica granatum* L.), a deciduous shrub, belongs to the Lythraceae family (Yuan et al. 2018). Pomegranate is diploid with  $n = 9$  (Qin et al. 2017). Pomegranate is considered to be highly valuable due to its

health-promoting traits in edible and nonedible parts of the fruit that can be used for a wide range of human disease. (Da Silva et al. 2013). Pomegranate was introduced into China 2000 years ago via the Silk Road (McColl 1991; Zhou et al. 1995). China is a main producer of pomegranates worldwide in six main production areas, including Huaiyuan of Anhui Province, Zaozhuang of Shandong Province, Lintong of Shanxi Province, Mengzi of Yunnan Province, Kaifeng of Henan Province, and Yecheng of the Xinjiang Uygur Autonomous Region. As a result of natural selection, introduction, domestication, and breeding of pomegranates, excellent local cultivars have been developed (Zhao et al. 2013), although the actual genealogy of various cultivars remains unknown.

Pomegranate breeding programs are based on the exploitation of genetic variability, so information on germplasm genetics, especially with respect to the genetic relationships among accessions and cultivars, is necessary for the rapid development of improved pomegranate cultivars and the achievement of breeding progress. The insensitivity of DNA markers to environmental changes makes them valuable for

---

Communicated by G. G. Vendramin

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11295-020-1428-4>) contains supplementary material, which is available to authorized users.

✉ Gaihua Qin  
qghahstu@163.com

<sup>1</sup> Key Laboratory of Genetic Improvement and Ecophysiology of Horticultural Crops, Anhui Province, Horticultural Research Institute, Anhui Academy of Agricultural Sciences, Hefei 230001, China

<sup>2</sup> Key Laboratory of Fruit Quality and Development Biology, Anhui Academy of Agricultural Sciences, Hefei 230001, China

evaluating germplasm diversity (Roy et al. 2006). Since 2007, various DNA markers, including amplified fragment length polymorphisms (AFLP) (Yuan et al. 2007), random amplified polymorphic DNA (RAPD) (Narzary et al. 2009), restriction fragment length polymorphisms (RFLP) (Melgarejo et al. 2009), and simple sequence repeats (SSRs) (Parvaresh et al. 2012), have been developed and applied to pomegranate genetic studies.

SSRs or microsatellites consist of tandemly arranged repeats of short DNA motifs (1–6 bp in length) with desirable genetic attributes including hypervariability, codominant inheritance, widespread abundance, and economic feasibility of large-scale use (Kantety et al. 2002). SSRs may play crucial roles in affecting chromatin organization, gene activity, recombination, and DNA replication more than just being “neutral” genetic markers (Li et al. 2002). SSR markers have been applied to cultivar identification, genetic diversity assessment, genetic mapping, and molecular evolution in plants. In spite of the recent progress in the development of single nucleotide polymorphisms (SNPs) for pomegranate (Ophir et al. 2014; Qin et al. 2017), SSR markers remain useful for practical breeding applications due to their reliability, high polymorphism, and ease of use. Up to now, the SSRs of pomegranate were mainly developed from enriched genomic libraries (Curro et al. 2010; Soriano et al. 2011; Parvaresh et al. 2012), a time-consuming and laborious process. The small number of SSR markers reported for pomegranate has limited their use in pomegranate genetic diversity study as well as marker-assisted selection. Identification of SSRs from the genome sequence has proved to be a robust, rapid, and widely applicable strategy (Zhang et al. 2012; Liu et al. 2016; Cui et al. 2017). A pomegranate “Dabenzi” genome sequence has been reported by Qin et al. (2017), which provides an opportunity to screen for SSRs across the genome more efficiently.

In the present study, we identified SSR motifs in a genome-wide manner and also quantified and characterized these SSR motifs. Specific primers flanking these SSR motifs were designed and applied to a genetic diversity study on a local pomegranate germplasm collection.

## Materials and methods

### Plant materials

A local collection of 218 pomegranate accessions, grown at the Gangji Eco-Agricultural Demonstration site, Anhui Academy of Agricultural Sciences, was involved in SSR evaluation and genetic diversity study (Supplementary Table S1). The 218 accessions were gathered from six provinces from amateurs, regional repositories, and wild populations. Latitude, longitude, altitude, and main climate conditions of each location were indicated in Supplementary Table S2.

### SSR motif identification and primer design

We searched for SSRs using the default parameters of the MISA (<http://pgrc.ipk-gatersleben.de/misa>) microsatellite identification tool (Thiel et al. 2003) from the whole genome sequence of pomegranate “Dabenzi” (DNA Data Bank of Japan/European Nucleotide Archive DDBJ/ENA/GenBank under the accession number MTKT00000000). The search criteria were 20 repeat units for mononucleotide repeats, 10 repeat units for dinucleotide repeats, 7 repeat units for trinucleotide repeats, 5 repeat units for tetranucleotide repeats, 4 repeat units for both pentanucleotide repeats, and hexanucleotide repeats. The flanking sequences of SSRs were used as targets for primer design according to Cui et al. (2017).

### Genomic DNA extraction

Genomic DNA was isolated from young leaves using a modified CTAB method (Doyle and Doyle 1990). Quality and concentration of the DNA was confirmed using 1% agarose gel electrophoresis and NanoDrop 8000 (Thermo Fisher Scientific, Waltham, MA, USA).

### Validation of new pomegranate SSR primer pairs by PCR amplification

Fifty new SSR primer pairs were randomly selected to prime the amplification of SSR fragments from the genomic DNA of six pomegranate accessions, one from each province. Primer pairs with clear bands, stable amplification, and high polymorphism were further selected to genotype the 218 pomegranate accessions. Forward primers of the primer pairs were 5' labeled with the fluorescence dye (FAM or HEX). PCR was carried out in a 10- $\mu$ l reaction mixture containing 50 ng of DNA, 0.2  $\mu$ M of each primer, 2 mM of MgCl<sub>2</sub>, 200  $\mu$ M of each dNTP, 0.5 U TaqDNA Polymerase (TIANGEN Biotech (Beijing) Co., Ltd., Beijing, China), and ddH<sub>2</sub>O to the total volume. Touchdown PCR was carried out on GeneAmp PCR System 9700 thermal cycler (Applied Biosystems, Foster City, USA) under the following conditions: 5 min at 95 °C followed by 11 cycles each of 30 s at 95 °C, 30 s at 65 °C (decreasing at 0.8 °C per cycle), 50 s at 72 °C, followed by 22 cycles of 30 s at 95 °C, 30 s at 55 °C, 50 s at 72 °C, and a final extension of 8 min at 72 °C.

The PCR products were separated on an Applied Biosystems 3730 capillary sequencer (Foster City, CA, USA) with GeneScan 500 LIZ as an internal size standard by the Sangon Biotech Company (Shanghai, China) and processed using GeneMarker (version 2.2.0; SoftGenetics, PA, USA). SSR alleles were manually checked for accuracy according to Pan et al. (2003). On the basis of the electrophoresis

results, DNA fragments were scored as present (1) or absent (0) and on the basis of allele size.

## Data analysis

Polymorphism information content (PIC), the allele number, the effective number of alleles ( $N_e$ ), Shannon's information index ( $I$ ), expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ), and Wright's fixation index (inbreeding coefficient,  $F_{is}$ ) were calculated using POPGENE software (<http://www.ualberta.ca/~fyeh>) (Yeh et al. 1997). The estimated frequency of null alleles ( $F(\text{null})$ ) was calculated using MICRO-CHECKER (<http://www.microchecker.hull.ac.uk/>) (Van Oosterhout et al. 2006). To evaluate the phylogenetic relationships among the accessions, the neighbor-joining algorithm was used to build a dendrogram based on genetic distances, using MEGA6.0 (Tamura et al. 2011). In total, 1000 replicates were used for all bootstrap tests. Structure 2.2 (<http://www.stats.ox.ac.uk/~pritch/home.html>) (Pritchard et al. 2000) was used to identify population groups by estimating the posterior probability of the data for a given  $K$ ,  $Pr(X/K)$ . The statistic  $K$  was determined by an admixture model with correlated alleles and  $K = 2-10$ . Ten independent runs of 100,000 Markov chain Monte Carlo replicas and 100,000 generations of burn-in were used to estimate each  $K$  value. The optimal  $K$  value was determined by the ad hoc statistic  $\Delta K$  (Evanno et al. 2005).

In addition, PowerCore 1.0 (<http://genebank.rda.go.kr/powercore.do>) was used to develop the core set with an effort to maximize the number of observed alleles at the SSR loci (Kim et al. 2007). Comparison at significant level  $p < 0.05$  in variability parameters (allele no.,  $I$ ,  $H_o$ , and  $H_e$ ) between the entire accessions and core collection was carried out with SPSS v.15.0 (SPSS, Chicago, IL, USA).

## Results

### SSR motif content in the pomegranate genome

A total of 36,792 SSRs were identified with an average frequency of 112 SSRs per megabase of the pomegranate genomic sequence (Supplementary Table S3). Of these SSRs, the dinucleotides were the most abundant (22,819, 62.02% of the total), followed by trinucleotides (4493, 12.21%), tetranucleotides (2999, 8.15%), pentanucleotides (2700, 7.34%), hexanucleotides (2483, 6.75%), and mononucleotide (1298, 3.53%) (Fig. 1). The proportion of different SSR motifs was not uniform. Briefly, with regard to the mononucleotide repeats, A/T were dominant, accounting for 60.7% of the total (Fig. 2a). Of the dinucleotide repeats, AT/TA were the most frequent (68.76%) followed by AG/CT (27.91%), AC/GT (3.31%), and CG/GC (0.02%) (Fig. 2b). Among the trinucleotide repeats, AAT/ATT (52.47%) were the most abundant,

followed by AAG/CTT (26.74%) and AGG/CCT (8.27%) (Fig. 2c). Of the tetra-, penta-, and hexanucleotide repeats, AAAT/TTTA, AACTC/GAGTT, and AAAATC/GATTTT were the most abundant repeats in each class (Fig. 2d-f).

### Characterization of SSR motifs in the pomegranate genome

The positions of the SSR motifs were determined by comparing with the genome annotations. There were only 662 SSR motifs (1.80% of the total) located in the coding regions, and 36,130 (98.20%) SSR motifs were located in noncoding regions. Among those SSRs in the coding regions, trinucleotide repeats were the most abundant SSRs (62.53%), followed by hexanucleotide repeats (21.76%).

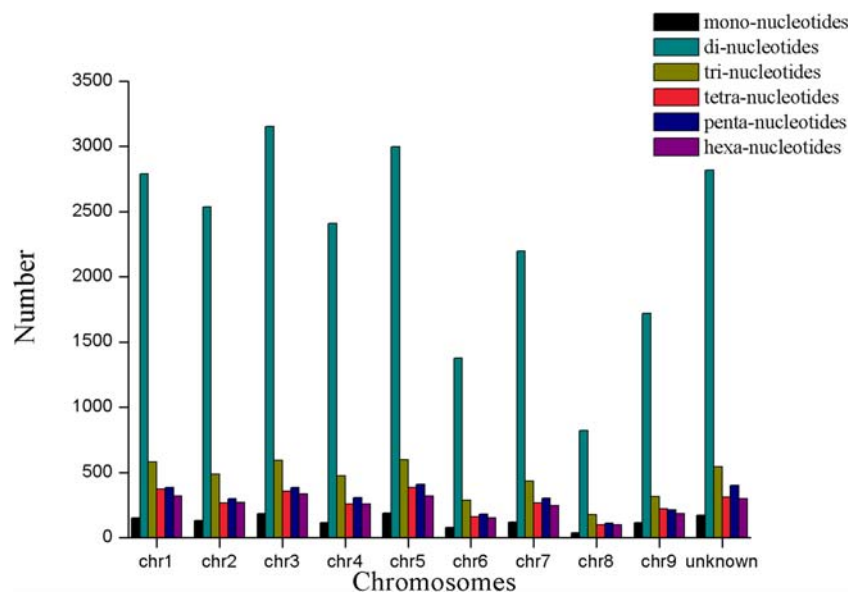
A total of 32,249 SSRs were distributed on the nine chromosomes, while 4543 SSRs were unanchored scaffolds. A chromosome-biased distribution of SSRs was found in the genome, the highest density being 159.66/Mb on chromosome 9, followed by 150.08/Mb on chromosome 7. The lowest density (125.90/Mb) of SSRs occurred on chromosome 1 (Supplementary Table S4).

### Analysis of genetic diversity in pomegranate

Out of the 50 primer pairs tested, 11 primer pairs displayed polymorphism among the six accessions. These 11 primer pairs were then used for genetic diversity analysis of the 218 pomegranate accessions. All primer pairs successfully amplified multiple DNA fragments (alleles), ranging from 2 (PG152) to 9 (PG093, PG080) across the pomegranate accessions. A total of 63 alleles were detected with a mean number of 5.72 alleles per primer pair (Table 1). Polymorphism information content (PIC) provides information on how informative a marker is or provides an estimate of the discriminatory power of the locus, as shown in Table 1. The PIC values of each SSR primer pair ranged from 0.22 to 0.70, with a mean value of 0.45. The results showed that PG093 and PG077 were the most informative SSR primer pairs as they had the highest PIC values. The Shannon information index ( $I$ ) varied from 0.45 for PG140 to 1.56 for PG093, with an average of 1.01; the observed heterozygosity ( $H_o$ ) varied from 0.16 for PG130 to 0.67 for PG077, with an average of 0.41. The expected heterozygosity ( $H_e$ ) varied from 0.25 for PG130 to 0.76 for PG093, with an average of 0.55. Wright's fixation index ( $F_{is}$ ) varied from 0.01 of PG077 to 0.45 of PG070, with an average of 0.23, implying that inbreeding increased among populations.

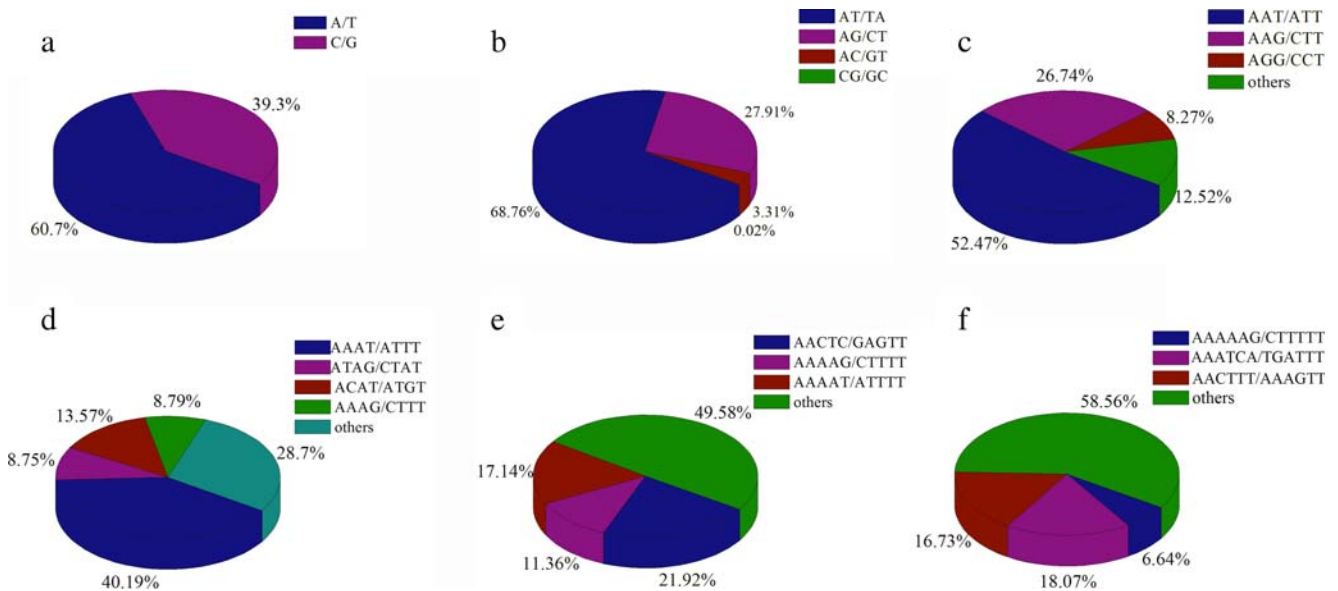
Structure software was used to infer pomegranate population structure (for  $K = 2, 3, \dots, 10$ ). The grouping of individuals using the model-based method revealed that the most likely value of  $K$  chosen with Evanno's method and log likelihood parameter was 3, indicating that the total panel was split into

**Fig. 1** Number of simple sequence repeat (SSR) motifs across nine chromosomes of the pomegranate “Dabenzi” reference genome



three main populations labeled P1, P2, and P3, consisting of 77, 81, and 60 samples. The P1 contained most accessions from Shandong, Shanxi, and Henan, along with all accessions from Tibet Autonomous Region and Xinjiang Uygur Autonomous Region, while the accessions from Anhui Province were divided into two populations (P2 and P3). The P2 contained most accessions from Huaibei City, Anhui Province, along with accessions from Shandong. The P3 contained all accessions from Huaiyuan County, Anhui Province, and several accessions from Huaibei City, Anhui Province. Intermix accessions, showing complicated genetic components, were displayed in all three populations, and there are more intermix accessions in P3 population.

Figure 3 shows the proportion of each population that contributed to each of the three clusters. The neighbor-joining analysis revealed three major clusters. Cluster I included 24 accessions from Anhui (11 accessions), Shandong (7 accessions), Shanxi (3 accessions), Xinjiang Uygur Autonomous Region (1 accession), Tibet Autonomous Region (1 accession), and Henan (1 accession). Cluster II contained 56 accessions, including 22 accessions from Shandong, 15 accessions from Anhui, 10 accessions from Henan, 6 accessions from Shanxi, and 3 accessions from Xinjiang Uygur Autonomous Region. Cluster III contained 138 accessions, which were clustered into five subclusters. Subcluster I contained 26 accessions mainly from Anhui (11 accessions), Shanxi (7



**Fig. 2** The frequency (percentage) of classified simple sequence repeat (SSR) motifs in the pomegranate genome. **a** Mononucleotides. **b** Dinucleotides. **c** Trinucleotide. **d** Tetranucleotides. **e** Pentanucleotides. **f** Hexanucleotides

**Table 1** Eleven SSR primer pairs used for analysis of genetic diversity in pomegranate

Markers	SSR motif	Chromosomal location	Primer sequences (5'–3')	Allele no.	$N_e$	$H_o$	$H_e$	$I$	$F(null)$	$Fis$	$PIC$
PG080	CA	4	F:ctgactgttcagagagtaggctg R:aggagggtgaaacaacgaatagctg	9/9	2.91/4.73	0.46/0.49	0.66/0.78	1.37/1.77	0.07	0.30	0.47
PG130	AG	2	F:ctcatatggcgattctctctctt R:aagttcgataaattgcactgtgtg	6/6	1.33/1.60	0.16/0.31	0.25/0.37	0.48/0.74	0.07	0.37	0.22
PG139	CT	3	F:gtttcctccctcaacccaaaa R:agtgggattttaccaagtgaaca	4/4	3.42/3.84	0.59/0.55	0.71/0.74	1.30/1.36	0.00	0.13	0.65
PG152	CT	8	F:catcagaatctcccttctgtg R:cagagagaagaagagagaccgagc	2/2	1.97/1.98	0.48/0.45	0.50/0.50	0.69/0.69	-0.01	0.08	0.37
PG153	AGA	8	F:gtgtttgatgctccattctctt R:gcctcaacgggtcttctctct	4/4	1.66/1.51	0.33/0.33	0.40/0.34	0.68/0.64	0.03	0.26	0.33
PG140	TG	3	F:caagaaagtgtgtagcgattgat R:ccaaacctaccctctctctc	3/3	1.35/1.53	0.24/0.33	0.26/0.35	0.45/0.59	0.03	0.08	0.23
PG098	AT	7	F:tgcttcttaaggactcaccaac R:ctaacctcatgcaactgtcatcca	7/7	2.80/3.37	0.34/0.37	0.64/0.70	1.25/1.42	0.15	0.43	0.56
PG070	TA	2	F:cacctctgctcagcaaaacaata R:caactcaacacaatatccaacca	5/5	2.36/2.67	0.33/0.37	0.58/0.63	0.99/1.08	0.11	0.45	0.35
PG077	CT	3	F:gtagtctcctctcttcaatgg R:agacgaagcacctgagaaggaat	6/6	3.27/3.66	0.67/0.83	0.70/0.73	1.31/1.40	0.00	0.01	0.65
PG090	AT	6	F:attctttataactcaaaatttgcca R:atgtcatgagaggaccacaaa	8/8	2.32/2.97	0.47/0.52	0.57/0.66	1.04/1.33	0.04	0.12	0.47
PG093	TA	6	F:cgtcaatagagctccctgagata R:gatgacgtggcagagtaagagagc	9/9	4.10/4.56	0.66/0.65	0.76/0.78	1.56/1.70	0.04	0.10	0.70
Mean	-	-	-	5.72/5.72	2.47/2.96	0.41/0.47	0.55/0.59	1.01/1.16*	0.04	0.23	0.45
total	-	-	-	63/63	-	-	-	-	-	-	-

*Allele no.*, observed number of alleles; *PIC*, polymorphism information content; *I*, Shannon's information index;  $H_o$ , observed heterozygosity;  $H_e$ , expected heterozygosity;  $N_e$ , effective number of alleles;  $F(null)$ , estimated frequency of null alleles; *Fis*, inbreeding coefficient. The first number in each pair represents values for the entire germplasm collections; the second number represents values for core collection.

\* $p < 0.05$ . Significant differences between the entire accessions and core collection as determined by *t* test are noted

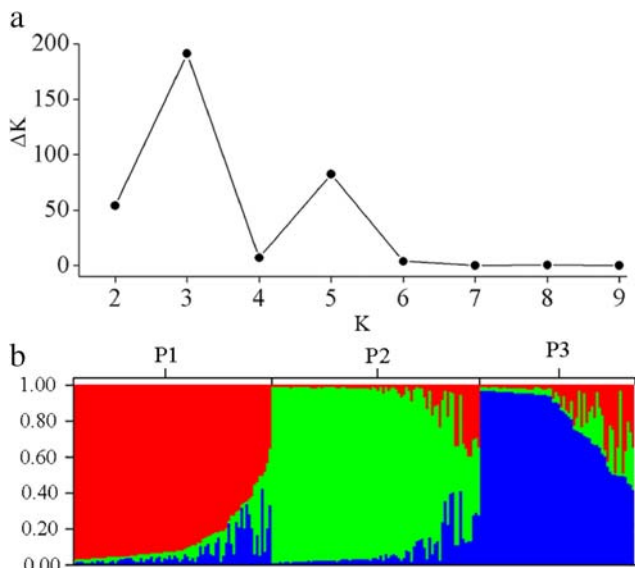
accessions), Shandong (4 accessions), Henan (2 accessions), and Xinjiang Uygur Autonomous Region (2 accessions). Subcluster II contained 18 accessions mainly from Anhui (10 accessions), Shandong (4 accessions), Henan (3 accessions), and Shanxi (1 accession). Subcluster III contained 15 accessions mainly from Anhui (12 accessions). Subcluster IV contained 23 accessions including 12 accessions from Anhui, 7 accessions from Shandong, 2 accessions from Henan, 1 accession from Xinjiang Uygur Autonomous Region, and 1 accession from Tibet Autonomous Region. Subgroup V contained the most accessions (56), namely 42 accessions from Anhui, 7 accessions from Shandong, 3 accessions from Henan, 2 accessions from Tibet Autonomous Region, 1 accession from Xinjiang Uygur Autonomous Region, and 1 accession from Shanxi (Fig. 4).

The 63 alleles detected at the eleven SSR loci were used to develop a core set of pomegranate germplasm collection using the program PowerCore 1.0. The core collection was selected with an effort to maximize the number of observed alleles with minimum redundancy. Forty-two accessions were selected as the entries in the core set germplasm collection (Table 2),

accounting for about 19.3% of the entire accession collection and 100% coverage of the alleles with minimum redundancy. The core collection includes 4 accessions from cluster I, 7 accessions from cluster II, and 30 accessions from cluster III. The average value of the observed heterozygosity ( $H_o$ ) was 0.47, the average value of the expected heterozygosity ( $H_e$ ) was 0.59, and the Shannon's information index (*I*) was 1.16. The parameters of allele no.,  $H_o$ , and  $H_e$  were detected by *t* value, showing that no significant difference occurred between the 218 pomegranate germplasm and 42 core collection.

## Discussion

Prior to this study, the SSR markers available for pomegranate were mainly developed from enriched genomic libraries, a process which is time-consuming and laborious (Hasnaoui et al. 2010; Soriano et al. 2011). In this study, we identified a total of 36,792 SSRs across the genome sequence of pomegranate cultivar "Dabenzi." Given the reference genome assembly size of 328.38 Mb, the average marker density was



**Fig. 3** **a** Exploration of  $K$  value for structure analysis of pomegranate germplasm by estimates of the rate of change of the slope of the log likelihood curve ( $\Delta K$ ) calculated according to Evanno et al. (2005) plotted against  $K$ . **b** Population structure of the studied accessions examined by genetic admixture analysis using the program structure with  $K=3$ . Each individual is presented by thin vertical bar and individual codes, which can be partitioned into  $K$ -colored segments that represent the individual estimated membership of the  $K$  cluster. Each vertical bar with more than one color indicates an admixed genetic constitution

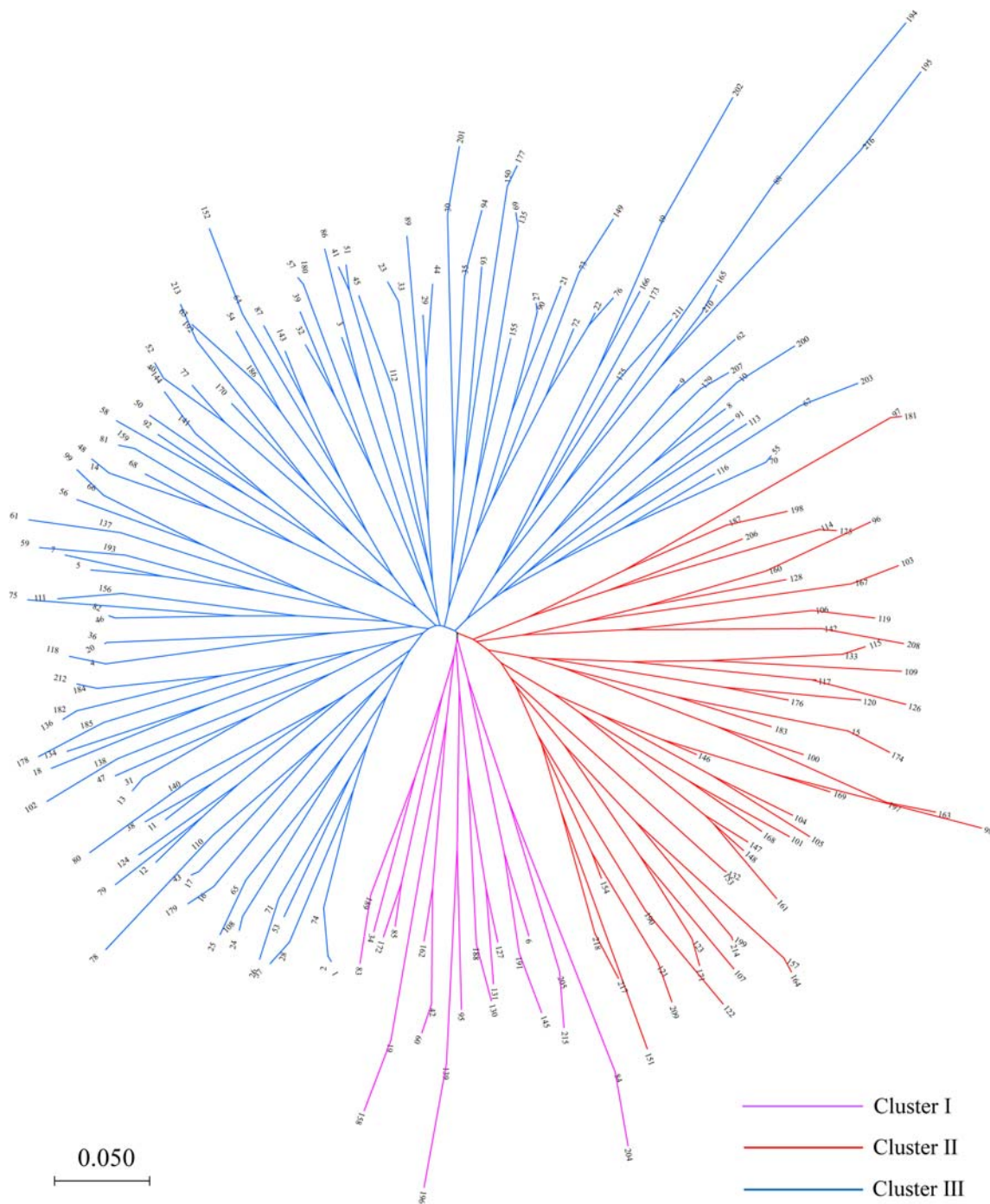
one SSR per 8.9 kb. Compared with one SSR per 24.53 kb in apple (Zhang et al. 2012), or one SSR per 12.5 kb in kiwifruit (Liu et al. 2016), the SSR density in pomegranate appears to be relatively high. This supports the idea that SSR density is negatively correlated with genome size in plants (Morgante et al. 2002). The dinucleotide repeats dominated. In pomegranate, when the number of motif repeats (except for mononucleotide repeats) increased, their occurrence frequency decreased, a finding consistent with previous observations in apple and kiwifruit (Zhang et al. 2012; Liu et al. 2016).

The 11 highly polymorphic SSR primer pairs amplified 63 SSR fragments from the whole collection of 218 accessions. The average PIC value was 0.45, which is higher than those previously reported for pomegranate (Pirseyedi et al. 2010; Soriano et al. 2011), with four markers (PG077, PG093, PG098, and PG139) each having a  $PIC \geq 0.5$ , at which point markers are considered to be highly polymorphic (Botstein et al. 1980). The polymorphic level of markers has been positively correlated with the genetic diversity within the test population and the number of accessions tested (Wu et al. 2014). Higher level of polymorphism could be detected by analyzing larger collections of accessions from various geographical origins, or wild populations (Ferrara et al. 2014). The 218 accessions used in the current study were collected from a broad array of geographic regions and might have contributed to the high PIC values of the 11 SSR primer pairs.

The distribution of different types of SSR motifs is not uniform and the most frequent motif type often differs between various plant species. For example, for dinucleotide repeats, AG/CT repeats dominated in *Brachypodium* and *Oryza sativa*, while AT/TA were more frequent in *Populus* and *Medicago*. For trinucleotide repeats, the motifs AGC/CGT, AGG/CCT, and CCG/CGG were observed more frequently among monocot species, whereas A/T-rich repeats were more frequent in dicots (Sonah et al. 2011). In pomegranate, SSRs proved to be strongly biased toward AT-rich repeat motifs. Among mono-, di-, tri-, and tetranucleotide repeats in the pomegranate genome, AT-rich repeats predominated, while GC-rich repeats were scarce. The distribution of SSRs varies in different regions across a genome; the noncoding regions generally contain greater frequencies of SSRs than the coding regions do (Hancock 1996). A similar characteristic was observed in this study on pomegranate, which might be due to negative selection against frameshift mutations in coding regions (Metzgar et al. 2000).

Pomegranate is an ancient fruit crop grown widely around the world, which has been cultivated widely in China. Due to restricted germplasm collection and limited molecular marker availability, few studies on the genetic diversity of pomegranate have been undertaken in China. Until recently, the most detailed characterization of genetic diversity in pomegranate in China was conducted on 85 accessions using fluorescent-AFLP markers (Yuan et al. 2007). More recently, we assembled a collection of 218 accessions from around China for use in this study, including wild types from Tibet Autonomous Region. Pomegranate accessions from China in this study showed higher levels of genetic diversity ( $H_o = 0.41$ ) than that reported in China with 46 pomegranate cultivars ( $H_o = 0.2126$ ) (Zhao et al. 2013) and 85 cultivars ( $H_o = 0.1539$ ) (Yuan et al. 2007). The high level of genetic diversity of pomegranate in China in this study might be because of the larger number of materials from different geographic origin and especially wild types in Tibet Autonomous Region. The set of SSR loci used in this study illustrated that the genetic base of the China pomegranate collection is broad enough to ensure future progress in breeding programs and will facilitate mining novel alleles to improve horticulturally important traits in the fruit.

Of the three genetic clusters obtained with STRUCTURE, P1 contained most accessions from Shandong, Shanxi, and Henan, along with 7 accessions from Xinjiang Uygur Autonomous Regional, and 5 wild accessions from Tibet Autonomous Region, indicating a common genetic background for these accessions. Pomegranate was spread from Central Asia to China along the Silk Road, first planted in Yecheng of Xinjiang, then spreading in Shanxi, Henan, Shandong, and Anhui Provinces (Zhou et al. 1995). Yuan et al. (2007) also suggested that Henan, Xinjiang, and Shanxi may be the earliest production areas. The accessions



**Fig. 4** Phylogenetic relationships identified based on simple sequence repeat (SSR) data using the distance-based neighbor-joining method

from Anhui Province were divided into two populations (P2 and P3), which showed two gene pools. The P3 contained all accessions from Huaiyuan County, Anhui Province, and several accessions from Huaibei City, Anhui Province, while P2 contained most accessions from Huaibei City, Anhui Province. It appeared that several pomegranate accessions from Huaibei City might be introduced from Huaiyuan County, as Huaiyuan County is the main production area of Anhui.

In cluster analysis, the accessions were not closely grouped geographically. Geographic origin was not always documented and a typical so-called local cultivar could be attributed to an accession that was finally shown to be a duplicate of a well-known foreign cultivar, which was common phenomenon in fruit trees (Liang et al. 2015; Lassois et al. 2016). Due to the lack of documentation about geographic origin and historical widespread exchanges of pomegranate cultivars over geographic regions, it was extremely difficult to know the real

**Table 2** The 42 accessions selected as entries in the core set

Accession	Province	Cluster	Accession	Province	Cluster
AHHB13	Anhui	Cluster III	HN06	Henan	Cluster III
AHHB24	Anhui	Cluster III	HN10	Henan	Cluster III
AHHB27	Anhui	Cluster III	HN12	Henan	Cluster III
AHHB43	Anhui	Cluster III	HN13	Henan	Cluster III
AHHB44	Anhui	Cluster III	HN4	Henan	Cluster III
AHHB45	Anhui	Cluster III	SXXA10	Shanxi	Cluster III
AHHB48	Anhui	Cluster III	SXXA14	Shanxi	Cluster II
AHHB57	Anhui	Cluster III	SXXA16	Shanxi	Cluster II
AHHB62	Anhui	Cluster III	SXXA21	Shanxi	Cluster III
AHHB63	Anhui	Cluster III	SXXA25	Shanxi	Cluster I
AHHB64	Anhui	Cluster III	SXXA26	Shanxi	Cluster I
HS03	Anhui	Cluster III	SXXA3	Shanxi	Cluster II
HY02	Anhui	Cluster II	SXXA5	Shanxi	Cluster II
HY04	Anhui	Cluster III	SXXA9	Shanxi	Cluster III
HY21	Anhui	Cluster II	SD14	Shandong	Cluster III
HY25	Anhui	Cluster III	SD16	Shandong	Cluster II
CY01	Xizang	Cluster III	SD18	Shandong	Cluster III
SD39	Shandong	Cluster III	SD20	Shandong	Cluster I
SD57	Shandong	Cluster II	SD27	Shandong	Cluster II
SD58	Shandong	Cluster I	SD30	Shandong	Cluster III
SD59	Shandong	Cluster III	SD33	Shandong	Cluster II

region of origin of many accessions studied. It could also explain that no clear relationship between the geographical origin and the genetic structure was found in the studied China germplasm, a finding which is in agreement with previous studies (Yuan et al. 2007; Jbir et al. 2014). The local pomegranate accessions have evolved as a result of long-term natural and artificial selection of genetic variations adaptive to specific environments.

The core collection represents 100% of the total diversity of all pomegranate germplasm, covering the major pomegranate production areas in China. The parameters of  $H_o$ , and  $H_e$  show that no significant difference occurred between the 218 pomegranate germplasm and 42 core collections. The results indicated that the core collection is representative of all of the genetic polymorphism. The genetic information of these core set accessions are very useful for pomegranate germplasm conservation and genetic utilization.

## Conclusion

We identified 36,792 SSR motifs from the whole genome sequence of pomegranate “Dabenzi.” Of the mono-, di-, tri-, and tetranucleotide repeats, the dominant types were A/T, AT/AT, AAT/ATT, and AAAT/ATTT. These SSR markers were also marked with the relative position information of both the

physical location and the linkage groups, and the adjacent genes of the SSR loci, which can provide better help for related research work and the developed SSR markers can be utilized by researchers and breeders for future pomegranate improvement. Fifty pairs of SSR primers were validated by PCR amplification of genomic DNA from 6 accessions, and 11 pairs of highly polymorphic SSR primers were selected to assess the genetic diversity. The germplasm is diverse with an expected heterozygosity  $H_e$  of 0.55 and observed heterozygosity  $H_o$  of 0.41. Analysis of the distribution of 63 amplified SSR alleles across 218 pomegranate accessions resulted in three distinct populations, namely, P1, P2, and P3. The population structure and genetic information will be useful in designing association mapping studies, which will further facilitate mining novel alleles to improve horticulturally important traits in pomegranate fruit. Forty-two accessions were selected as the entries in a core set of pomegranate germplasm collection for future pomegranate genetics and breeding programs.

**Acknowledgments** We are thankful for Professor Pan Yongbao (Agricultural Research Services, United States Department of Agriculture) for the help in polishing the language.

**Funding information** This research was funded by the Project of the Key Laboratory of Anhui Academy of Agricultural Sciences (2019YL013) and the Natural Science Foundation of Anhui Province (1708085MC85).

**Data archiving statement** This research contains no data that requires submission to a public database. The 36,792 SSR motifs found in this study are listed in Supplementary Table S3.

## References

- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331
- Cui J, Cheng J, Nong D, Peng J, Hu Y, He W, Zhou Q, Dhillon NPS, Hu K (2017) Genome-wide analysis of simple sequence repeats in bitter melon (*Momordica charantia*). *Front Plant Sci* 8:1103
- Curro S, Caruso M, Distefano G, Gentile A, La Malfa S (2010) New microsatellite loci for pomegranate, *Punica granatum* (Lythraceae). *Am J Bot* 97:e58–e60
- Da Silva JA, Rana TS, Narzary D, Verma N, Meshram DT, Ranade SA (2013) Pomegranate biology and biotechnology: a review. *Sci Hortic* 160:85–107
- Doyle J, Doyle J (1990) Isolation of DNA from small amounts of plant tissues. *BRL Focus* 12:13–15
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Ferrara G, Giancaspro A, Mazzeo A, Giove SL, Matarrese AMS, Pacucci C, Punzi R, Trani A, Gambacorta G, Bianco A (2014) Characterization of pomegranate (*Punica granatum* L.) genotypes collected in Puglia region, Southeastern Italy. *Sci Hortic* 178:70–78
- Hancock JM (1996) Simple sequences and the expanding genome. *Bioessays* 18:421–425



- Hasnaoui N, Buonamici A, Sebastiani F, Mars M, Trifi M, Vendramin G (2010) Development and characterization of SSR markers for pomegranate (*Punica granatum* L.) using an enriched library. *Conserv Genet Resour* 2:283–285
- Jbir R, Melgarejo P, Hernández F, Haddioui A, Hannachi AS (2014) Efficiency of Inter Simple Sequence Repeat (ISSR) markers for the assessment of genetic diversity of Moroccan pomegranate (*Punica granatum* L.) cultivars. *Biochem Syst Ecol* 56:24–31
- Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* 48:501–510
- Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, Kim TS, Cho EG, Park YJ (2007) PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* 23:2155–2162
- Lassois L, Denancé C, Ravon E, Guyader A, Guisnel R, Hibrand-Saint-Oyant L, Poncet C, Lasserre-Zuber P, Feugey L (2016) Genetic diversity, population structure, parentage analysis, and construction of core collections in the French apple germplasm based on SSR markers. *Plant Mol Biol Report* 34(4):827–844
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* 11:2453–2465
- Liang W, Dondini L, De Franceschi P, Paris R, Sansaviniet S, Tartarini S (2015) Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. *Plant Mol Biol Report* 33(3):458–473
- Liu CY, Zhang Q, Yao XH, Zhong CH, Yan CL, Huang HW (2016) Characterization of genome-wide simple sequence repeats and application in interspecific genetic map integration in kiwifruit. *Tree Genet Genomes* 12:21
- McColl R (1991) China's silk roads: a modern journey to China's Western regions. *Focus Geogr* 41:1
- Melgarejo P, Martínez J, Hernández F, Martínez R, Legua P, Oncina R, Martínez-Murcia A (2009) Cultivar identification using 18S–28S rDNA intergenic spacer-RFLP in pomegranate (*Punica granatum* L.). *Sci Hortic* 120:500–503
- Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res* 10:72–80
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* 30:194
- Narzary D, Mahar KS, Rana TS, Ranade SA (2009) Analysis of genetic diversity among wild pomegranates in Western Himalayas, using PCR methods. *Sci Hortic* 121:237–242
- Ophir R, Sherman A, Rubinstein M, Eshed R, Schwager MS, Harel-Beja R, Bar-Ya'akov I, Holland D (2014) Single-nucleotide polymorphism markers from de-novo assembly of the pomegranate transcriptome reveal germplasm genetic diversity. *PLoS One* 9(2): e88998
- Pan Y-B, Cordeiro GM, Richard E, Henry R (2003) Molecular genotyping of sugarcane clones with microsatellite DNA markers. *Maydica* 48:319–329
- Parvaresh M, Talebi M, Sayed-Tabatabaei B-E (2012) Molecular diversity and genetic relationship of pomegranate (*Punica granatum* L.) genotypes using microsatellite markers. *Sci Hortic* 138:244–252
- Pirseyedi SM, Valizadehghan S, Mardi M, Ghaffari MR, Mahmoodi P, Zahravi M, Zeinalabedini M, Nekoui SMK (2010) Isolation and characterization of novel microsatellite markers in pomegranate (*Punica granatum* L.). *International Journal of Molecular Sciences* 11
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Qin G, Xu C, Ming R, Tang H, Guyot R, Kramer EM, Hu Y, Yi X, Qi Y, Xu X (2017) The pomegranate (*Punica granatum* L.) genome and the genomics of punicalagin biosynthesis. *Plant J* 91:1108–1128
- Roy A, Bandyopadhyay A, Mahapatra AK, Ghosh SK, Singh NK, Bansal KC, Koundal KR, Mohapatra T (2006) Evaluation of genetic diversity in jute (*Corchorus species*) using STMS, ISSR, and RAPD markers [J]. *Plant breed* 125(3):292–297
- Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK, Gacche RN, Rana JC, Singh NK, Sharma TR (2011) Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. *PLoS One* 6:e21298
- Soriano JM, Zuriaga E, Rubio P, Llácer G, Infante R, Badenes ML (2011) Development and characterization of microsatellite markers in pomegranate (*Punica granatum* L.). *Mol Breed* 27:119–128
- 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* 28:2731–2739
- Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422
- Van Oosterhout C, Weetman D, Hutchinson WF (2006) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Mol Ecol Notes* 6(1):255–256
- Wu H-B, Gong H, Liu P, He X-L, Luo S-B, Zheng X-M, Zhang C-Y, He X-M, Luo J (2014) Large-scale development of EST-SSR markers in sponge gourd via transcriptome sequencing. *Mol Breed* 34: 1903–1915
- Yeh FC, Yang R, Boyle TB, Ye Z, Mao JX (1997) POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada 10:295–301
- Yuan Z, Yin Y, Qu J, Zhu L, Li Y (2007) Population genetic diversity in Chinese pomegranate (*Punica granatum* L.) cultivars revealed by fluorescent-AFLP markers. *Journal of Genetics and Genomics* 34: 1061–1071
- Yuan ZH, Fang YM, Zhang TK, Fei Z, Han F, Liu C, Liu M, Xiao W, Zhang W, Wu S, Zhang M, Ju Y, Xu H, Dai H, Liu Y, Chen Y, Wang L, Zhou J, Guan D, Yan M, Xia Y, Huang X, Liu D, Wei H, Zheng H (2018) The pomegranate (*Punica granatum* L.) genome provides insights into fruit quality and ovule developmental biology. *Plant Biotechnol J* 16(7):1363–1374
- Zhang Q, Ma B, Li H, Chang Y, Han Y, Li J, Wei G, Zhao S, Khan MA, Zhou Y (2012) Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. *BMC Genomics* 13:537
- Zhao L, Li M, Cai G, Pan T, Shan C (2013) Assessment of the genetic diversity and genetic relationships of pomegranate (*Punica granatum* L.) in China using RAMP markers. *Sci Hortic* 151:63–67
- Zhou GJ, Yuan YY, Zeng FZ, Xian KZ, Chen ZG (1995) Status and prospect of pomegranate production in China. *Southwest China Journal of Agricultural Sciences* 8:111–116 (in Chinese with an English abstract)

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