The Genotypes of Polymorphic Simple Sequence Repeat Loci Revealed by Whole-genome Resequencing Data of 30 *Pyrus* Accessions

Shuang Jiang¹, Haishan An¹, Xiaoqing Wang, Chunhui Shi, and Jun Luo²

Forestry and Pomology Research Institute, Shanghai Key Lab of Protected Horticultural Technology, Shanghai Academy of Agricultural Sciences, Shanghai, 201403, China

Yuanwen Teng

Department of Horticulture, The State Agricultural Ministry Key Laboratory of Horticultural Plant Growth, Development and Quality Improvement, Zhejiang University, Hangzhou, Zhejiang, 310058, China

ADDITIONAL INDEX WORDS. bioinformatics, breeding, marker selection, pear

ABSTRACT. Simple sequence repeats (SSRs) are widely used in cultivar identification, genetic relationship analysis, and quantitative trait locus mapping. Currently, the selection of hybrid progeny plants in molecular marker-assisted breeding mostly relies on SSR markers because of their ease of operation. In *Pyrus*, a large number of SSR markers have been developed previously. The method to identify polymorphic SSRs quickly is still lacking in cultivated as well as wild pear species. We present a large number of polymorphic SSRs identified using a quick in silico approach applied across 30 cultivated and wild accessions from *Pyrus* species. A total of 49,147 SSR loci were identified in *Pyrus*, and their genotypes were evaluated by whole-genome resequencing data of 30 *Pyrus* accessions. The results show that most SSR loci were dinucleotide repeat motifs located in intergenic regions. The genotypes of all SSR loci were revealed in all accessions. A total of 23,209 loci were detected, with more than one genotype in all *Pyrus* accessions. We selected 702 highly polymorphic SSR loci to characterize the pear accessions with an average polymorphism information content value of 0.67, suggesting that these SSR loci were highly polymorphic. The genetic relationship of *Pyrus* species in the neighbor-joining (NJ) tree and population structure showed a clear division between the oriental and occidental accessions. The population structure split all oriental pears into two groups: cultivars and wild accessions. These new findings of the polymorphic SSR loci in this study are valuable for selecting appropriate markers in molecular marker-assisted breeding in *Pyrus*.

The genus *Pyrus* (pears) consists of important fruit trees, and ≈20 primary species are generally accepted by most taxonomists (Challice and Westwood, 1973). Based on their geographic distribution, *Pyrus* species are divided into oriental and occidental species (Bailey, 1917; Rubstov, 1944). Oriental pears, which include 12 to 15 species, are mostly native to China (Teng and Tanabe, 2004). Occidental pears comprise more than 10 species and are distributed in Europe, northern Africa, and central Asia (Rubstov, 1944). Most cultivated pears are assigned to four species: *P. pyrifolia*, *P. ussuriensis*,

P. ×sinkiangensis, and *P. communis* (Teng et al., 2002). The former three species are oriental pears; the last one is an occidental pear. Pear breeding is based on intraspecific and interspecific hybridization and has traditionally relied on the evaluation of morphological characteristics, mainly fruit weight, sugar content, and taste. Recently, with the development of molecular breeding technology, the breeding process can be accelerated by molecular marker-assisted selection (Collard et al., 2005).

SSRs, also known as microsatellites, are one of the most efficient genetic markers. An SSR refers to a DNA sequence 1 to 6 bp in length that is repeated a variable number of times (Zietkiewicz et al., 1994). SSR markers are characterized by a multiallelic nature, codominant inheritance, and good genome coverage (Powell et al., 1996). SSR markers have been largely applied in cultivar identification (Kimura et al., 2002), genetic diversity studies (Bao et al., 2007), and quantitative trait locus mapping (Perchev et al., 2015; Yamamoto et al., 2014). Compared with other molecular markers, SSR markers provide a number of advantages, such as polymerase chain reaction (PCR) screening, relative abundance, and low cost (Naghavi et al., 2007; Palombi and Damiano, 2002). For a selected locus, the genotype is identified by the variable sequence length determined after separation of PCR products amplified with the SSR primers. SSR markers with a high degree of transferability among species have been useful for comparative genetics and cultivar identification. For example,

Received for publication 8 Apr. 2019. Accepted for publication 4 June 2019. This work was financed by a grant from Shanghai Science and Technology Committee Rising-Star Program (19QB1403500) rewarded to JS, a grant for Discipline Construction in Shanghai Academy of Agricultural Sciences (2018-31701886) rewarded to JS, and a specialized research fund for Major Science and Technique of Zhejiang Province of China (no. 2016C02052-4) rewarded to TU. Shuang Jiang and Haishan An performed the experiments and wrote the manuscript. Xiaoqing Wang, Chunhui Shi and Fangjie Xu helped to analyze the data and revise the manuscript. Jun Luo and Yuanwen Teng were involved in designing the research and revised the manuscript. All authors read and approved the manuscript.

The chromosome data of ‘Dangshansuli’ were downloaded from the Center of Pear Engineering Technology Research of Nanjing Agricultural University, China (<http://peargenome.njau.edu.cn/>). The whole-genome resequencing data of the 30 *Pyrus* accessions were downloaded from BIG Data Center and National Center for Biotechnology Information (Table 1). The remaining data supporting our findings are contained within the article and supplementary files.

¹These authors contributed equally to this work.

²Corresponding author. E-mail: pear_luo@126.com.

Table 1. Mapped reads to 49,147 simple sequence repeat sequences in the resequencing data of 30 *Pyrus* accessions.

Accession number	Sample	Reads (no.)	Sequencing depth	Mapped reads to SSR database (no.)	Avg mapped reads for each SSR locus (no.)
SRR609906, SRR609907, SRR609912, SRR609915	<i>P. pyrifolia</i> Chinese White Pear ‘Suli’	214,779,326	80×	9,765,299	199
CRR019693	<i>P. pyrifolia</i> ‘Zaoshengxinshui’	25,059,608	14×	2,055,856	42
CRR019694	<i>P. pyrifolia</i> ‘Qiushui’	22,768,452	12.7×	1,916,030	39
CRR046438	<i>P. pyrifolia</i> ‘Hosui’	29,762,912	16.6×	1,373,956	28
CRR046439	<i>P. pyrifolia</i> ‘Cuiguan’	30,476,109	17×	1,415,216	29
CRR046440	<i>P. ussuriensis</i> ‘Nanguo’	25,541,135	14.3×	1,202,316	24
CRR046441	<i>P. ussuriensis</i> ‘Huagai’	30,224,078	16.9×	1,839,742	37
CRR019695	<i>P. pashia</i>	28,683,576	16×	2,493,872	51
CRR019696	<i>P. betulifolia</i>	25,877,886	14.5×	2,526,692	51
CRR019697	<i>P. calleryana</i>	24,759,507	13.8×	2,267,829	46
CRR019698	<i>P. communis</i> ‘Comice’	26,322,987	14.7×	1,725,282	35
CRR046442	<i>P. communis</i> ‘Abate Fetel’	29,462,529	16.5×	1,126,394	23
SRR7135495	<i>P. cossonii1</i>	25,428,148	14.2×	1,876,184	38
SRR7135502	<i>P. cossonii2</i>	25,919,967	14.5×	1,908,813	39
SRR7135501	<i>P. elaeagnifolia1</i>	33,008,338	18.4×	2,375,952	48
SRR7135519	<i>P. elaeagnifolia2</i>	30,545,132	17.1×	2,321,576	47
SRR7135505	<i>P. spinosa1</i>	35,509,116	19.8×	2,602,538	53
SRR7135506	<i>P. spinosa2</i>	24,586,516	13.7×	1,900,720	39
SRR7135511	<i>P. syriaca1</i>	27,238,434	15.2×	2,049,611	42
SRR7135512	<i>P. syriaca2</i>	36,956,089	20.7×	2,587,862	53
SRR7135517	<i>P. fauriei1</i>	25,263,650	14.1×	2,792,369	57
SRR7135520	<i>P. fauriei2</i>	26,049,767	14.6×	2,903,227	59
SRR7135521	<i>P. mamorensis1</i>	35,724,135	20×	2,731,496	56
SRR7135522	<i>P. mamorensis2</i>	36,323,356	20.3×	2,934,594	60
SRR7135534	<i>P. regelii1</i>	41,706,444	23.3×	3,505,812	71
SRR7135535	<i>P. regelii2</i>	34,901,196	19.5×	3,084,778	63
SRR7135518	<i>P. glabra</i>	48,153,439	26.9×	3,559,823	72
SRR7135527	<i>P. nivalis</i>	25,943,547	14.5×	1,911,041	39
SRR7135541	<i>P. salicifolia</i>	24,116,379	13.5×	1,771,072	36
SRR7135542	<i>P. sachokiana</i>	22,018,673	12.3×	1,726,697	35
Avg		31,077,351	18.2×	2,475,088	50

some SSR markers isolated from *Malus* could be used in *Pyrus* (Yamamoto et al., 2001).

In previous research, SSR markers were widely used in genetic diversity analyses in *Pyrus* (Bao et al., 2007; Cao et al., 2012; Yue et al., 2018). A large number of polymorphic SSR markers were needed. With the development of next-generation sequencing technology, whole-genome data of *Pyrus* were published, and the development of SSR markers based on whole-genome data was possible (Wu et al., 2013). The availability of sequences generated as a result of advances in next-generation sequencing present an opportunity to identify a large number of polymorphic SSRs in pear. A total of 1756 SSR markers, including 1341 newly designed SSRs based on whole-genome sequencing of *Pyrus*, were first evaluated for polymorphism to construct a genetic map in pear (Chen et al., 2015). Xue et al. (2018) designed a total of 101,694 pairs of SSR primers, and only 332 primer pairs were tested and selected as clear, stable, and polymorphic SSR markers. Although these two groups of researchers developed a large number of SSR markers in *Pyrus*, the polymorphism of a small number of markers was tested because of the high cost of screening SSR markers by *silico* PCR. Illumina (San Diego, CA) HiSeq

provides short read sequences, which are 83 and 150 bp in length in HiSeq 2000 and HiSeq 2500+, respectively. The length of the core sequences of SSRs mostly ranges from 12 to 100+ bp, which makes some reads cover the repeat sequence (Zhao et al., 2015). It is possible to use reads to obtain the various lengths of the sequences of SSR loci in *Pyrus* accessions.

In this study, a genome-wide analysis of microsatellites from the draft chromosome data of the pear ‘Dangshansuli’ (*P. pyrifolia* Chinese White Pear Group) was performed. The genotype of isolated microsatellites was detected by whole-genome resequencing data to enhance our understanding of genetic diversity in pear. Furthermore, the polymorphism information of the isolated markers will be helpful to breeders.

Materials and Methods

DNA EXTRACTION AND PLANT MATERIALS. In this study, we sequenced five pear accessions (*P. pyrifolia* ‘Hosui’, *P. pyrifolia* ‘Cuiguan’, *P. ussuriensis* ‘Nanguo’, *P. ussuriensis* ‘Huagai’, and *P. communis* ‘Abate Fetel’). Genomic DNA was

extracted from the young leaves of each specimen using the CTAB protocol according to Doyle and Doyle (1987). The genome resequencing data of 25 pear accessions in public databases were also analyzed in this study. Six of them were from our previous research (Jiang et al., 2019) and 19 of them were from the other researcher's report (Wu et al., 2018). The data were deposited in BIG Data Center (BIG Data Center Members, 2018) and the National Center for Biotechnology Information [NCBI (Sharma et al., 2018)]. Their accession numbers are shown in Table 1.

WHOLE-GENOME RESEQUENCING DATA. High-quality genomic DNA was interrupted randomly by ultrasound. DNA fragments ranging from 150 to 800 bp were isolated by electrophoresis. T4 DNA Polymerase, Klenow DNA Polymerase, and T4 PNK were used to make blunt ends from cohesive ends of double-stranded DNA. The DNA fragment was ligated to adaptors. All fragments were recovered by electrophoresis and then paired-end-sequenced using an Illumina HiSeq. 4000. The adaptors and low-quality reads with more than 20% bases of quality value ≤ 10 in the raw sequence data in FASTQ format were filtered by Trimmomatic (Bolger et al., 2014). Only clean reads were used in the subsequent analysis. The clean read data were deposited in the Genome Sequence Archive in the BIG Data Center, and the accession numbers are CRR046438 through CRR046442. The chromosome data of 'Dangshansuli' (version V121121) were downloaded from the Center of Pear Engineering Technology Research of Nanjing Agricultural University, China (Wu et al., 2018).

IDENTIFICATION OF SSR LOCI. The workflow diagram for this study is shown in Fig. 1. The SSR loci in the pear chromosome data were identified using a microsatellite identification tool [MISA (Thiel et al., 2003)], which is based on the Perl language. The SSR loci containing repeat units of two to six nucleotides were identified, and the minimum SSR length

criteria were defined as six iterations for dinucleotide repeats and five iterations for other repeat units. The sequence of all SSR loci and their left and right flanking regions of 120 bp were reserved as a database (SSRDB), which was used in the sequencing data mapping. Moreover, the conserved left and right flanking regions of 10 bp around SSR loci were also reserved as two databases, L10DB and R10DB, respectively.

THE ANALYSIS OF SSR LOCI BASED ON GENOME RESEQUENCING DATA. First, all sequencing data of the 30 pear accessions were mapped to the SSRDB by Magic-Blast (Grzegorz et al., 2018) with a parameter of "-score 80", and the mapped reads for each SSR loci in every accession were identified. Second, an in-house Perl script (Supplemental Data 1) was used to scan the mapped reads of each SSR locus against the two corresponding short sequences of 10 bp in L10DB and R10DB, and to determine the genotype that corresponds to the length of the SSR between the two short sequences of 10 bp. Finally, the sequence length of all SSR loci in each sample was obtained. Cervus [version 3.07 (Kalinowski et al., 2007)] was used to calculate the frequency of the number of alleles, observed heterozygosity (Ho), expected heterozygosity (He), and polymorphism information content (PIC).

GENETIC RELATIONSHIP ANALYSIS. For each SSR marker, fragment sizes in all samples were collected, and repetitive sizes were removed in Microsoft Excel. Total fragment sizes without repetition were obtained. Then, based on the total fragment sizes, a Perl script was compiled to transform the presence (1) or absence (0) of each fragment size in an accession to binary data. A dendrogram was constructed based on Nei's genetic distances (Nei and Li, 1979) by the NJ method with 200 bootstrap replicates using TREECON [version 1.3b (Van de Peer and De Wachter, 1997)]. The population structure was evaluated with a Bayesian approach using the software STRUCTURE [version 2.3.4 (Evanno et al., 2005; Pritchard et al., 2000)]. This revealed the genetic structure by assigning individuals or predefined groups to clusters. Six runs of STRUCTURE were performed with the number of homogeneous gene pools (K) from 1 to 8. Each run consisted of a burn-in period of 100,000 iterations followed by 100,000 Monte Carlo Markov chain iterations, assuming an admixture model. The results were uploaded to the STRUCTURE HARVESTER website to estimate the most appropriate K value (Earl and Vonholdt, 2012). Replicate cluster analyses of the same data resulted in several distinct estimated assignment coefficients, even though the same starting conditions were used. Therefore, we used CLUMPP software (Jakobsson and Rosenberg, 2007) to average the six independent simulations and illustrated the result graphically using DISTRUCT (Rosenberg, 2004).

Results

IDENTIFICATION OF SSRs AND DISTRIBUTION IN THE PEAR GENOME. The current chromosome-level assembly of the 'Dangshansuli' genome (493 Mbp) was used to identify SSRs. A total of 256,425 loci were identified (520 loci/Mbp) and included mononucleotide repeat motifs (176,463; 68.82%), dinucleotide repeat motifs (63,965; 24.94%), trinucleotide repeat motifs (11,828; 4.61%), tetranucleotide repeat motifs (3317; 1.29%), pentanucleotide repeat motifs (601; 0.23%), and hexanucleotide repeat motifs (251; 0.10%) (Fig. 1A). AG/CT and AT/AT were the dominant dinucleotide repeats,

Fig. 1. The workflow diagram in the identification of genotypes of simple sequence repeat (SSR) loci revealed by whole-genome resequencing data. The sequence of all SSR loci and their left and right flanking region of 120 bp were reserved as a database (SSRDB).

Fig. 2. The type and distribution of microsatellites in the genome of *Pyrus*. (A) Type of repeat motifs. (B) Microsatellite length distribution. (C) Number of microsatellites. (D) Percent of different repeat motifs. P2, dinucleotide repeats; P3, trinucleotide repeats; P456, tetranucleotide repeats, pentanucleotide repeats, and hexanucleotide repeats; SSRs, simple sequence repeats.

accounting for 41.81% and 38.76%, respectively. AAT/ATT and AAG/CTT were the most abundant trinucleotide repeats, accounting for 26.46% and 27.64%, respectively. About 37.76% of isolated SSR loci had a sequence length of less than 21 bp (Fig. 2B), and the number of microsatellites decreased correspondingly with increasing microsatellite length. Removal of mononucleotide-containing SSRs and compound microsatellite loci that contain more than one SSR motif resulted in 49,147 loci. The sequence of these SSR loci and their 120-bp flanking region (≈ 300 bp in the total length) were reserved in a database named SSRDB. The distribution of SSR loci in the *Pyrus* genome showed that all SSR loci were classified into five regions: the gene intron, gene exon, UTR-5' (untranslated regions), UTR-3', and intergenic regions. Most SSR loci (97.6%) were commonly mapped onto intergenic regions (Fig. 2C), which might be related to the sequence length of the intergenic regions. In the *Pyrus* genome, intergenic regions were longer than genes and had more SSR sequences. In five regions, dinucleotide repeats accounted for a larger proportion of the intergenic regions

(Fig. 2D). The remaining regions had more trinucleotide repeats.

ASSESSMENT OF SSR LOCI BASED ON WHOLE-GENOME RESEQUENCING DATA. All reads in the sequencing data of the 30 pear accessions were mapped to the SSRDB using Magic-Blast. The average number of reads in pear accessions was 31 Mbp (sequencing depth, 18.2 $\times$), and almost 2.4-Mbp reads were mapped to the identified SSR loci in each accession (Table 1). A single-copy SSR locus in each pear accession mostly had one or two fragment sizes. The average number of mapped reads in each SSR locus was 50 in this study (Table 1), which satisfied the analysis needs. The mapped reads were analyzed to obtain the length of repeat motifs between two flanking conserved sequences of the SSR locus, and then the different lengths of the fragments in 49,147 loci were found (Supplemental Table 1). In some accessions, fragment sizes were not found in some loci. The median value of the percent of SSR loci with detected fragments in all accessions was 53%, suggesting that some loci were not covered in the genome resequencing data (Fig. 3A). When the percentage of

Fig. 3. The characteristics of the microsatellites in the 30 *Pyrus* accessions. (A) Percentage of simple sequence repeat (SSR) loci detected with fragments. (B) Number of alleles in 23,209 SSR loci. (C) Value of observed heterozygosity (Ho), expected heterozygosity (He), and polymorphism information content (PIC) in 23,209 SSR loci.

Fig. 4. The percentage of simple sequence repeat (SSR) loci with more than two fragments in 30 *Pyrus* accessions.

accessions with detected fragments in a locus was less than 60%, this locus was deleted as a result of lack of information. Finally, a total of 23,209 markers were reserved, and the percent of SSR loci with detected fragments in each accession was up to 95% (Fig. 3A). The number of alleles among the SSR loci ranged from 1 to 26, with a median value of 7 (Fig. 3B). The median values of Ho, He, and PIC were 0.389, 0.758, and 0.7, respectively (Fig. 3C). A large number of loci had no polymorphism, displaying the low average value of Ho and PIC, suggesting that these SSR loci were not useful in pear breeding. In this study, some loci had more than two fragment sizes in some pear accessions. The percent of these SSR loci ranged from 6.3% to 25.7%, suggesting that these loci duplicated during evolution (Fig. 4). The least and greatest values were found in *P. ussuriensis* 'Nanguo' and

P. glabra, respectively. The duplication of SSR loci in wild accessions was more frequent than that in cultivars (Fig. 4).

THE HIGHLY POLYMORPHIC SSR LOCI IN THE CULTIVARS. Although a large number of SSR markers was detected, we did not need this many SSR markers in this application. The selection of highly polymorphic SSR loci was necessary for the cultivars. If three fragment sizes were obtained in a locus in an accession, we considered it a multicopy SSR, and were deleted first in the nine pear cultivars. Then, some loci with fragment sizes detected in 60% of accessions were reserved. To reduce further the number of SSR markers, the number of alleles with more than three markers was used in the analysis. Finally, a total of 702 SSR loci were obtained (Supplemental Table 2). Their average PIC value was 0.67, suggesting that these loci were highly polymorphic. The average marker distance was 1.9

Fig. 5. Genetic relationship among the 30 accessions of *Pyrus* revealed by dendrogram based on the neighbor-joining method and STRUCTURE (Evanno et al., 2005) under $K = 3$ (right) and $K = 2$ (left).

loci/Mbp. The positions of these loci are shown in the *Pyrus* chromosome (Supplemental Fig. 1, Supplemental Table 2), which provides a valuable reference to select suitable markers for molecular breeding in *Pyrus*.

THE GENETIC RELATIONSHIP OF *PYRUS* SPECIES BASED ON THE HIGHLY POLYMORPHIC SSR LOCI. A total of 702 loci in 30 accessions was used to evaluate the genetic relationship of *Pyrus* species and cultivars. One in-house Perl script was used to exchange the fragment sizes to binary data (0, 1) in each SSR locus. First, genetic relationships among the 30 pear accessions were revealed by the NJ clustering approach (Fig. 5). The dendrogram clearly distinguished oriental pears and occidental pears. Seven groups were well clustered. Groups I and II were oriental wild pear accessions and cultivars, respectively. The remaining groups were occidental pears, and these groups were distributed geographically. The species in groups III, IV, and V were from west Asia; the species in groups VI and VII were from Europe and North Africa. Second, the number of K among the 30 accessions was modeled by Bayesian methods using STRUCTURE software. The evaluation of the optimum number of K indicated two maxima for ΔK at $K = 2$ and $K = 3$ (Supplemental Fig. 2), suggesting that a model with two gene pools captured a major split in the data, with the substantial additional resolution provided under the model with $K = 3$. At $K = 2$, the oriental pears, including cultivars and wild accessions, constituted a gene pool (blue, Fig. 5), and the occidental pears constituted the other gene pool (red, Fig. 5). Under this model, the wild pear accessions had two gene pools. The model with three gene pools was also supported by the STRUCTURE results. Under this model, the gene pool in green (Fig. 5) consisted of all occidental pear species. Among the oriental pears, the wild

accessions had a blue gene pool, and the cultivars had an admixture gene pool of blue and red (Fig. 5).

Discussion

Microsatellites are widely distributed in the genome *Pyrus* and are used in cultivar identification (Bao et al., 2007), genetic relationship analysis (Bassil and Postman, 2010; Yue et al., 2018), and quantitative trait locus mapping (Montanari et al., 2016). Although the single nucleotide polymorphism markers based on genome sequencing were popular in the current study, the SSR markers were needed in a long period. The SSR marker was good at detecting the genotype of a single locus because of its easy operation and cheap price. In this study, 256,425 SSR loci were identified from the *Pyrus* chromosomes with a size of 493 Mbp. About 520 loci were found on average in a genome sequence of 1 Mbp (1 locus/2000 bp), which was less than the number of SSR loci found in *Arabidopsis thaliana* (874.8 loci/Mbp) and *Oryza sativa* (807.4 loci/Mbp)

(Lawson and Zhang, 2006). The most abundant microsatellites were mononucleotide repeat motifs, of which the number of A/T motifs was more than C/G motifs, suggesting the A/T repeat was easily mutated (Supplemental Table 1). In microsatellites with 2 to 6 bp (di-, tri-, tetra-, penta-, and hexanucleotides) repeat motifs, dinucleotide repeat motifs accounted for a large proportion. CG/CG was rarely found in the microsatellites, which also occurred in the trinucleotide repeat motifs with a few ACG/CGT and CCG/CGG motifs. The microsatellite distribution is related to selection pressure during evolution and may perform varied functions in different genomic regions. According to their location in the genome, microsatellites are generally divided into intergenic and genic microsatellites. The genic microsatellites are derived from genes, including the gene intron, gene exon, UTR-5', UTR-3', and intergenic regions. A high percentage (97.6%) in 49,147 loci was found in the intergenic regions, which are a subset of noncoding DNA. Most of them have no currently known function. These microsatellites are not regulated by natural selection and are dinucleotide repeat motifs (Fig. 2D). In genic regions, most microsatellites were trinucleotide repeat motifs, and gene exons had a greater proportion of these motifs than the other genic regions. Two or four base deletions and insertions in a gene will cause an open reading frameshift mutation, which causes only trinucleotide repeat motifs in gene exons to be reserved in evolution.

In some previous studies, the SSR markers were isolated from the plant genome, and then PCR with DNA primers was performed to test their polymorphism (Jiao et al., 2012; Pandey et al., 2018). In this study, we identified a large number of SSR markers and developed an approach to evaluate the genotype of microsatellites by genome resequencing data. Thirty pear accessions sequenced by Illumina HiSeq. 2500 + were

analyzed in this study, and the length of the reads was 150 bp. We also tested a large number of genome resequencing data by Illumina HiSeq. 2000 with a short read length (usually <100 bp) from NCBI. Unfortunately, the short reads were limited to obtaining the length of fragments of each SSR locus. The sequencing depth in this study ranged from 12.3 to 80×, and the average number of mapped reads for each SSR locus in a sample was 50, which could cover many fragments in each SSR locus (Table 1). Therefore, the sequencing depth, with more than 12× (total data size 6GB) in a sample, was sufficient to evaluate the genotype of microsatellites in the whole genome. In this study, 49,147 SSR loci were tested. The various lengths of fragments are shown in Supplemental Table 1. We found that no fragments were detected in some loci in the tested accessions, implying that not every region of the genome was covered in the genome resequencing data. The information of all SSR loci in *Pyrus* accessions will be helpful for selecting suitable markers in future work. A large number of loci showed a very low PIC value, suggesting these loci were conserved in the *Pyrus* accessions. The loci with a high PIC value were potentially appropriate markers for the analyses of genetic diversity and relationships of *Pyrus* species.

The genetic relationship of 30 accessions was revealed by 702 highly polymorphic SSR loci. Such a large number of SSR loci has not been used in previous studies. The NJ tree and STRUCTURE analysis (Fig. 5) divided *Pyrus* into two large groups—oriental pears and occidental pears—which is in accordance with the results obtained using amplified fragment length polymorphism (Bao et al., 2008), random amplification of polymorphic DNA (Monte-Corvo et al., 2000), and DNA sequences (Zheng et al., 2014). The NJ tree showed that seven groups could be further divided into all pear accessions. Among the oriental pears, cultivars and wild pear accessions were clustered separately, suggesting that their genetic relationship was distant (Fig. 5). Group I included cultivars from *P. pyrifolia* and *P. ussuriensis*, suggesting that the genetic background of these two species was similar. Cultivars of *P. ussuriensis* were reported to be not monophyletic, and introgression occurred from *P. pyrifolia* (Jiang et al., 2016; Yu et al., 2016). Five accessions from *P. pashia*, *P. betulaeifolia*, *P. calleryana*, and *P. fauriei* were clustered in group II, and the former three species were considered primitive species in oriental pears (Jiang et al., 2016; Zheng et al., 2014). *P. fauriei* originated from Korea and was once treated as a type of *P. calleryana*. The occidental pears were divided into five groups. The species from groups III, IV, and V were distributed in western Asia. *P. regelii*, in group III, were clustered with oriental pears, and the degree of support was up to 82%. *P. regelii* are widely distributed in Kazakhstan, bordering China. Geographic proximity caused a close relationship between *P. regelii* and oriental pears. *P. elaeagnifolia*, *P. syriaca*, and *P. salicifolia* were clustered into group VI. Their close relationship was also revealed by nuclear DNA sequence (Zheng et al., 2014). In group VI, three wild species and two cultivars were clustered. Two cultivars were from *P. communis*, which has a close relationship with *P. nivalis*. Two accessions from *P. mamorensis* originating from North Africa were clustered in group VII.

The population structure results provided two models for the genetic relationships of all pears (Supplemental Fig. 1). Under the model of $K = 2$, two gene pools colored in blue and red were identified (Fig. 5), corresponding to oriental pears and occidental pears, respectively. Some oriental pears were found to contain the gene pool of occidental pears, and some occidental

pear accessions were introgressed by the gene pool found in oriental pears, suggesting that genetic exchange has occurred between these two pear groups (Jiang et al., 2016). The model of $K = 3$ was also supported by this structure. In this model, the oriental pears were further divided into two groups: cultivars and wild pear accessions (Fig. 5). The cultivars and wild pear accessions were the main gene pools and were colored red and blue, respectively (Fig. 5). Some cultivars had two major gene pools, suggesting that gene flow occurred from wild accessions to cultivars. In addition, some wild pear accessions showed gene introgression from cultivated accessions.

Conclusions

In summary, the current analysis isolated 49,147 SSR loci in pears and provided a new approach to test their genotype based on whole-genome resequencing data from 30 *Pyrus* accessions. To our knowledge, such a large number of SSR loci has not been used in previous studies. In our findings, the sequencing depth of 12× (total data size, 6GB) in one sample was sufficient to evaluate the polymorphism of microsatellites in the whole *Pyrus* genome. The genotype of all SSR loci in *Pyrus* will provide a valuable reference for selecting suitable markers in future work, especially in molecular marker-assisted pear breeding.

Literature Cited

- Bailey, L. 1917. *Pyrus*, p. 2865–2878. In: A.W. Leftwich (ed.). The standard cyclopedia of horticulture. Vol. 4. MacMillan, New York, NY.
- Bao, L., K.S. Chen, D. Zhang, Y.F. Cao, T. Yamamoto, and Y. Teng. 2007. Genetic diversity and similarity of pear (*Pyrus* L.) cultivars native to east Asia revealed by SSR (simple sequence repeat) markers. *Genet. Resources Crop Evol.* 54:959–971.
- Bao, L., K.S. Chen, D. Zhang, X.G. Li, and Y. Teng. 2008. An assessment of genetic variability and relationships within asian pears based on AFLP (amplified fragment length polymorphism) markers. *Scientia Hort.* 116:374–380.
- Bassil, N. and J.D. Postman. 2010. Identification of european and asian pears using EST-SSRs from *Pyrus*. *Genet. Resources Crop Evol.* 57:357–370.
- BIG Data Center Members. 2018. Database resources of the BIG Data Center in 2019. *Nucl. Acids Res.* 47:D8–D14.
- Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Cao, Y.F., L.M. Tian, Y. Gao, and F.Z. Liu. 2012. Genetic diversity of cultivated and wild ussuriian pear (*Pyrus ussuriensis* Maxim.) in China evaluated with M13-tailed SSR markers. *Genet. Resources Crop Evol.* 59:9–17.
- Challice, J.S. and M.N. Westwood. 1973. Numerical taxonomic studies of the genus *Pyrus* using both chemical and botanical characters. *Bot. J. Linn. Soc.* 67:121–148.
- Chen, H., Y. Song, L.T. Li, M.A. Khan, X.G. Li, S.S. Korban, J. Wu, and S.L. Zhang. 2015. Construction of a high-density simple sequence repeat consensus genetic map for pear (*Pyrus* spp.). *Plant Mol. Biol. Rpt.* 33:316–325.
- Collard, B.C.Y., M.Z.Z. Jahufer, J.B. Brouwer, and E.C.K. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142:169–196.
- Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19:11–15.
- Earl, D.A. and B.M. Vonholdt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resources* 4:359–361.

- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14:2611–2620.
- Grzegorz, M.B., T.M. Jean, T.M. Danielle, B. Ben, and L.M. Thomas. 2018. Magic-BLAST, an accurate DNA and RNA-seq aligner for long and short reads. *bioRxiv*, doi: 10.1101/390013.
- Jakobsson, M. and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.
- Jiang, S., X. Wang, C. Shi, and J. Luo. 2019. Genome-wide identification and analysis of high-copy-number LTR retrotransposons in asian pears. *Genes (Basel)* 10:156.
- Jiang, S., X. Zheng, P. Yu, X. Yue, M. Ahmed, D. Cai, and Y. Teng. 2016. Primitive genepools of asian pears and their complex hybrid origins inferred from fluorescent sequence-specific amplification polymorphism (SSAP) markers based on LTR retrotransposons. *PLoS One* 11:e0149192.
- Jiao, Y., H.M. Jia, X.W. Li, M.L. Chai, H.J. Jia, Z. Chen, G.Y. Wang, C.Y. Chai, E. van de Weg, and Z.S. Gao. 2012. Development of simple sequence repeat (SSR) markers from a genome survey of chinese bayberry (*Myrica rubra*). *BMC Genomics* 13:201.
- Kalinowski, S.T., M.L. Taper, and T.C. Marshall. 2007. Revising how the computer program Cervus accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16:1099–1106.
- Kimura, T., Y. Shi, M. Shoda, K. Kotobuki, N. Matsuta, T. Hayashi, Y. Ban, and T. Yamamoto. 2002. Identification of asian pear varieties by SSR analysis. *Breed. Sci.* 52:115–121.
- Lawson, M.J. and L. Zhang. 2006. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biol.* 7:R14.
- Montanari, S., L. Perchepped, D. Renault, L. Frijters, R. Velasco, M. Horner, S.E. Gardiner, D. Chagné, V.G.M. Bus, C.E. Durel, and M. Malnoy. 2016. A QTL detected in an interspecific pear population confers stable fire blight resistance across different environments and genetic backgrounds. *Mol. Breed.* 36:1–16.
- Monte-Corvo, L., L. Cabrita, C. Oliveira, and J. Leita. 2000. Assessment of genetic relationships among *Pyrus* species and cultivars using AFLP and RAPD markers. *Genet. Resources Crop Evol.* 47:257–265.
- Naghavi, M.R., M. Mardi, S.M. Pirseyedi, M. Kazemi, P. Potki, and M.R. Ghaffari. 2007. Comparison of genetic variation among accessions of *Aegilops tauschii* using AFLP and SSR markers. *Genet. Resources Crop Evol.* 54:237–240.
- Nei, M. and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* 76:5269–5273.
- Palombi, M. and C. Damiano. 2002. Comparison between RAPD and SSR molecular markers in detecting genetic variation in kiwifruit (*Actinidia deliciosa* A. Chev.). *Plant Cell Rep.* 20:1061–1066.
- Pandey, M., R. Kumar, P. Srivastava, S. Agarwal, S. Srivastava, N.S. Nagpure, J.K. Jena, and B. Kushwaha. 2018. WGSSAT: A high-throughput computational pipeline for mining and annotation of SSR markers from whole genomes. *J. Hered.* 109:339–343.
- Perchepped, L., D. Leforestier, E. Ravon, P. Guérif, C. Denancé, M. Tellier, S. Terakami, T. Yamamoto, M. Chevalier, Y. Lespinasse, and C.E. Durel. 2015. Genetic mapping and pyramiding of two new pear scab resistance QTLs. *Mol. Breed.* 35:197.
- Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, and S. Tingey. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2:225–238.
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Rosenberg, N.A. 2004. DISTRUCT: A program for the graphical display of population structure. *Mol. Ecol. Notes* 4:137–138.
- Rubstov, G.A. 1944. Geographical distribution of the genus *Pyrus* and trends and factors in its evolution. *Am. Nat.* 78:358–366.
- Sharma, S., S. Ciuffo, E. Starchenko, D. Darji, L. Chlumsky, I. Karsch-Mizrachi, and C.L. Schoch. 2018. The NCBI bioCollections database. *Database (Oxford)*, doi: 10.1093/database/bay006.
- Teng, Y. and K. Tanabe. 2004. Reconsideration on the origin of cultivated pears native to east Asia. *Acta Hort.* 634:175–182.
- Teng, Y., K. Tanabe, F. Tamura, and A. Itai. 2002. Genetic relationships of *Pyrus* species and cultivars native to East Asia revealed by randomly amplified polymorphic DNA markers. *J. Amer. Soc. Hort. Sci.* 127:262–270.
- Thiel, T., W. Michalek, R.K. Varshney, and A. Graner. 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 de Peer, Y. and R. De Wachter. 1997. Construction of evolutionary distance trees with TREECON for Windows: Accounting for variation in nucleotide substitution rate among sites. *Comput. Appl. Biosci.* 13:227–230.
- Wu, J., Y. Wang, J. Xu, S.S. Korban, Z. Fei, S. Tao, R. Ming, S. Tai, A.M. Khan, J.D. Postman, C. Gu, H. Yin, D. Zheng, K. Qi, Y. Li, R. Wang, C.H. Deng, S. Kumar, D. Chagné, X. Li, J. Wu, X. Huang, H. Zhang, Z. Xie, X. Li, M. Zhang, Y. Li, Z. Yue, X. Fang, J. Li, L. Li, C. Jin, M. Qin, J. Zhang, X. Wu, Y. Ke, J. Wang, H. Yang, and S. Zhang. 2018. Diversification and independent domestication of asian and european pears. *Genome Biol.* 19:77.
- Wu, J., Z. Wang, Z. Shi, S. Zhang, R. Ming, S. Zhu, M.A. Khan, S. Tao, S.S. Korban, H. Wang, N.J. Chen, T. Nishio, X. Xu, L. Cong, K. Qi, X. Huang, Y. Wang, X. Zhao, J. Wu, C. Deng, C. Gou, W. Zhou, H. Yin, G. Qin, Y. Sha, Y. Tao, H. Chen, Y. Yang, Y. Song, D. Zhan, J. Wang, L. Li, M. Dai, C. Gu, Y. Wang, D. Shi, X. Wang, H. Zhang, L. Zeng, D. Zheng, C. Wang, M. Chen, G. Wang, L. Xie, V. Sovero, S. Sha, W. Huang, S. Zhang, M. Zhang, J. Sun, L. Xu, Y. Li, X. Liu, Q. Li, J. Shen, J. Wang, R.E. Paull, J.L. Bennetzen, J. Wang, and S. Zhang. 2013. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* 23:396–408.
- Xue, H., P. Zhang, T. Shi, J. Yang, L. Wang, S. Wang, Y. Su, H. Zhang, Y. Qiao, and X. Li. 2018. Genome-wide characterization of simple sequence repeats in *Pyrus bretschneideri* and their application in an analysis of genetic diversity in pear. *BMC Genomics* 19:473.
- Yamamoto, T., T. Kimura, Y. Sawamura, K. Kotobuki, Y. Ban, T. Hayashi, and N. Matsuta. 2001. SSRs isolated from apple can identify polymorphism and genetic diversity in pear. *Theor. Appl. Genet.* 102:865–870.
- Yamamoto, T., S. Terakami, N. Takada, S. Nishio, N. Onoue, C. Nishitani, M. Kunihiisa, E. Inoue, H. Iwata, T. Hayashi, A. Itai, and T. Saito. 2014. Identification of QTLs controlling harvest time and fruit skin color in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed. Sci.* 64:351–361.
- Yu, P., S. Jiang, X. Wang, S. Bai, and Y. Teng. 2016. Retrotransposon-based sequence-specific amplification polymorphism markers reveal that cultivated *Pyrus ussuriensis* originated from an interspecific hybridization. *Eur. J. Hort. Sci.* 81:264–272.
- Yue, X., X. Zheng, Y. Zong, S. Jiang, C. Hu, P. Yu, G. Liu, Y. Cao, H. Hu, and Y. Teng. 2018. Combined analyses of chloroplast DNA haplotypes and microsatellite markers reveal new insights into the origin and dissemination route of cultivated pears native to East Asia. *Front. Plant Sci.* 9:591.
- Zhao, H., L. Yang, Z. Peng, H. Sun, X. Yue, Y. Lou, L. Dong, L. Wang, and Z. Gao. 2015. Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys*. *Sci. Rpt.* 5:8018.
- Zheng, X., D. Cai, D. Potter, J. Postmand, J. Liu, and Y. Teng. 2014. Phylogeny and evolutionary histories of *Pyrus* L. revealed by phylogenetic trees and networks based on data from multiple DNA sequences. *Mol. Phylogenet. Evol.* 80:54–65.
- Zietkiewicz, E., A. Rafalski, and D. Labuda. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20:176–183.