

RESEARCH ARTICLE

Open Access



Application of genome-wide insertion/deletion markers on genetic structure analysis and identity signature of *Malus* accessions

Xuan Wang¹, Fei Shen^{1,2}, Yuan Gao³, Kun Wang³, Ruiting Chen^{1,4}, Jun Luo⁵, Lili Yang⁵, Xi Zhang¹, Changpeng Qiu¹, Wei Li¹, Ting Wu¹, Xuefeng Xu¹, Yi Wang¹, Peihua Cong³, Zhenhai Han¹ and Xinzhong Zhang^{1*}

Abstract

Background: Apple (*Malus* spp.), one of the most important temperate fruit crops, has a long cultivation history and is economically important. To identify the genetic relationships among the apple germplasm accessions, whole-genome structural variants identified between *M. domestica* cultivars 'Jonathan' and 'Golden Delicious' were used.

Results: A total of 25,924 insertions and deletions (InDels) were obtained, from which 102 InDel markers were developed. Using the InDel markers, we found that 942 (75.3%) of the 1251 *Malus* accessions from 35 species exhibited a unique identity signature due to their distinct genotype combinations. The 102 InDel markers could distinguish 16.7–71.4% of the 331 bud sports derived from 'Fuji', 'Red Delicious', 'Gala', 'Golden Delicious', and other cultivars. Five distinct genetic patterns were found in 1002 diploid accessions based on 78 bi-allele InDel markers. Genetic structure analysis indicated that *M. domestica* showed higher genetic diversity than the other species. *Malus* underwent a relatively high level of wild-to-crop or crop-to-wild gene flow. *M. sieversii* was closely related to both *M. domestica* and cultivated Chinese cultivars.

Conclusions: The identity signatures of *Malus* accessions can be used to determine distinctness, uniformity, and stability. The results of this study may also provide better insight into the genetic relationships among *Malus* species.

Keywords: *Malus*, InDel, Bud sports, Genetic structure, Germplasm

Background

Apple (*Malus* spp.), one of the most commonly cultivated fruit crops, supports many local economies in temperate zones. *Malus* is extremely rich in diversity, with 25 to 78 species in the genus depending on the taxonomic classifications [51, 56]. High levels of interspecific hybridization occur naturally, which generates genetic admixtures, contributing to the diversity within the genus [6, 7, 12]. In addition to the natural diversification

of the genus, anthropogenic activities, including selection and cross breeding, have led to approximately 10,000 cultivars worldwide [8, 21, 65]. Identification of the distinctness of the germplasm would be beneficial to the successful conservation and efficient utilization of genetic resources.

The genetic variability and allelic diversity in these accessions are usually examined to reveal their distinctness. Identification of population structure and kinship within germplasm collections is a fundamental prerequisite for identifying robust marker-trait associations [68]. There are also possibly duplicates, synonyms,

* Correspondence: zhangxinzhong999@126.com

¹College of Horticulture, China Agricultural University, Beijing, China
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

homonyms, or materials with missing names that must be carefully examined among the living collections [40]. For example, a previous study identified 330 apple cultivars or abandoned trees that could be either grafted clones or 'own rooted seedlings' using nine SSR markers [24]. In addition, the test for distinctness, uniformity, and stability (DUST) is a statutory requirement to release a new cultivar (International Union for the Protection of New Varieties of Plants (UPOV) Convention Articles 5–9, 991)[64]. Limited by the fact that traditional field tests are time-consuming, laborious, and greatly influenced by the environment, DNA markers are used in DUST in many species [25, 58]. Therefore, it is also necessary to develop an efficient marker-assisted DUST protocol in *Malus* plants.

Owing to the co-dominant inheritance and because they are often multi-allelic, simple sequence repeat (SSR) markers have been widely used in apples to evaluate genetic diversity, population structure, and to analyze parentage [16, 23, 30, 32, 40, 43, 65]. However, the disadvantages of SSRs are frequently reported. The instability of SSRs increased dramatically with plant age [22]. Certain chemicals or radiation may cause DNA double-strand breaks, and the repair of these breaks usually results in small insertions or deletions (InDels) at the break site. These InDels presumably contribute to the instability of SSRs [22]. Errors can also be found in the documented parentage of some accessions by comparing the SSR profiles to show parent-offspring similarity [16, 46].

Single nucleotide polymorphism (SNP) markers are commonly used in large-scale, high-throughput automated detection of genetic variation because of their large number and wide distribution in the genome [44, 67]. A previous study used an 8 K apple SNP array [5] to identify cryptic relationships between accessions, analyze population structure, and calculate the linkage disequilibrium in apple [68]. Similarly, 3704 confident SNPs were used to analyze a core collection of cider and dessert French apple cultivars [34].

In addition to SSR and SNP markers, InDels have been recognized as an ideal source for marker development due to their high-density, co-dominance, robust stability, and genotyping efficiency [28, 74]. InDel markers have been used to identify the specificity of germplasm resources and provide information for breeding in chickpea (*Cicer arietinum* L.) [28], cotton (*Gossypium hirsutum* L.) [74], pepper (*Capsicum spp.* L.) [26], *Carapa guianensis* [63], mung bean (*Vigna radiata* (L.) Wilczek) [35], and cucumber (*Cucumis sativus* L.) [36]. In addition, InDel markers have been successfully used to distinguish somatic variations in apple [33], although InDel markers have not been as widely used in apple as SSRs and SNPs.

The objectives of the current study were to develop a set of stable co-dominant InDel markers and to identify

Malus accessions. Genome-wide InDels were robustly used for analysis of distinctness, genetic structure, genetic composition, and the parentage of 1251 *Malus* accessions. The results provided insight into *Malus* germplasm resources and may facilitate the future utilization of germplasm in apple breeding.

Results

Genome-wide structural variant (SV) calling and selection of InDel markers

The next generation resequencing data from the two apple founder cultivars, 'Jonathan' and 'Golden Delicious', resulted in an average read depth of 43.44 and have been deposited in the NCBI Sequence Read Archive (SRA) with the accession number PRJNA392908 [60]. A total of 66,841 genome-wide SVs between 'Jonathan' and 'Golden Delicious' were obtained using the apple genome v1.0 as reference [69], including 16,130 deletions (DEL), 9794 insertions (INS), 430 inversions (INV), 1132 intra-chromosomal translocations (ITX), and 39,355 inter-chromosomal translocations (CTX) (Fig. 1a-c). Our results showed that InDels were more well-distributed on chromosomes than the other types of SVs (Fig. 1c). The length of the majority of DEL (78.15%) and INS (99.53%) ranged from 50 bp to 400 bp (Fig. 1b), indicating that InDels are more representative for genome-wide marker development because of their large number and frequent distribution throughout the genome.

Of the 170 InDels chosen throughout the genome (10 per chromosome), 102 were validated for further analyses (example of validated InDel in Fig. 2a and b, list of validated InDels in Table S1). These InDels were combined into nine fluorescence multiplex PCR groups; each group contained three to 24 InDel markers (Table S2).

Genotyping of the selected InDel markers and generation of identity signatures of accessions

Of the 1251 *Malus* accessions included in this study, 942 exhibited unique genotype combinations (Table S3). Three hundred and nine accessions shared genotype combinations with at least one of the other accessions, which comprised of 76 distinct patterns (Tables S3 and S4). Sixty-one accessions were found to be synonyms (including 2 with known alternative names), 78 were homonyms, and 8 were replicated collections (Table S5). The genotypes of two tetraploids, 'Zumi Crab 4x' (B-21) and 'Gala 4x' (XC-4), were the same as their diploid progenitors, 'Zumi Crab' (B-30) and 'Ruihong' (QD-25), respectively (Table S4). There were 199 accessions that were mutants of nine known cultivars (Table S4) and 22 accessions had registration errors or incorrect names (Table S5). Excluding synonyms and accessions with errors or incorrect names, 1018 accessions were identified as unique by the 102 InDel markers (Table S3). The InDel identity signature was then generated for each of

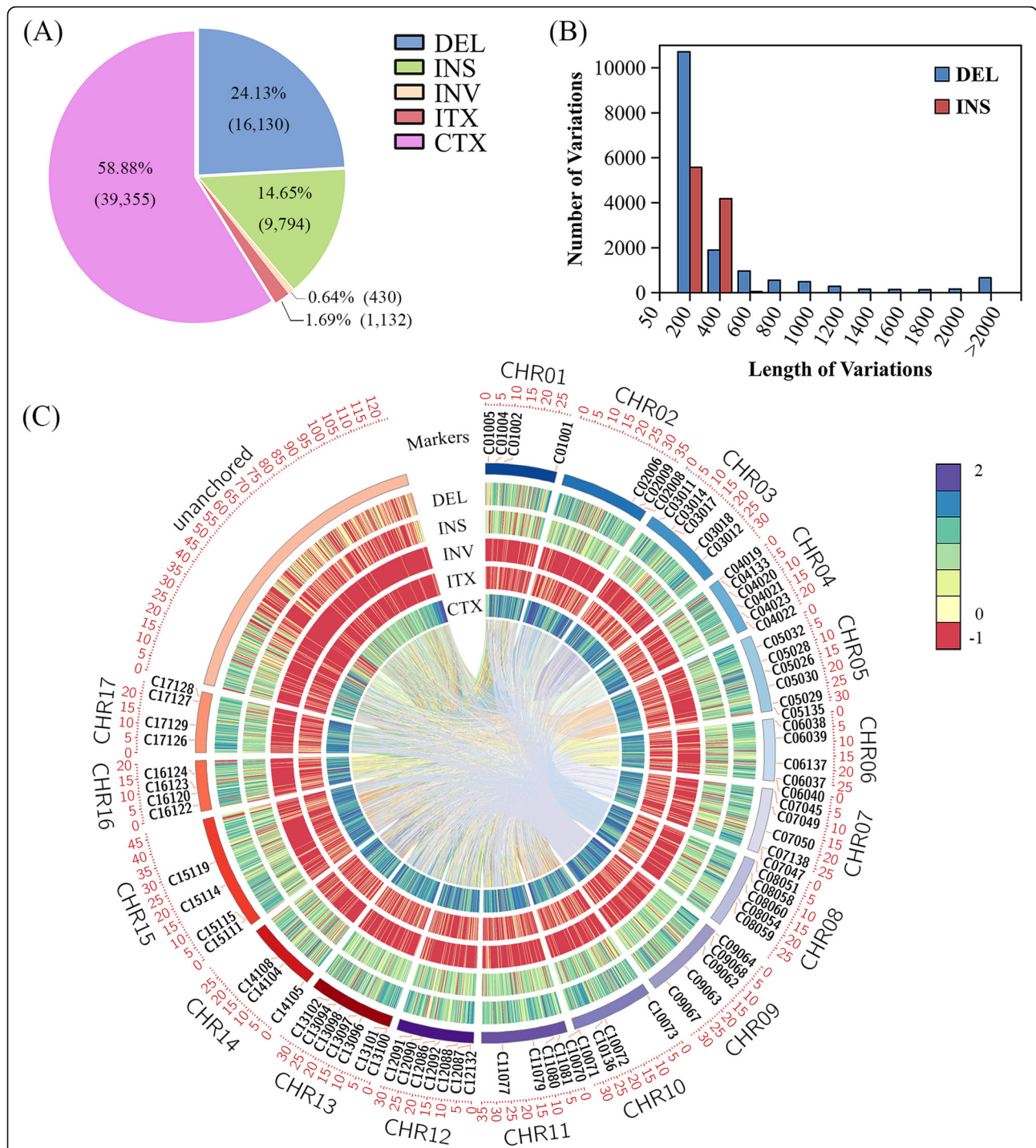
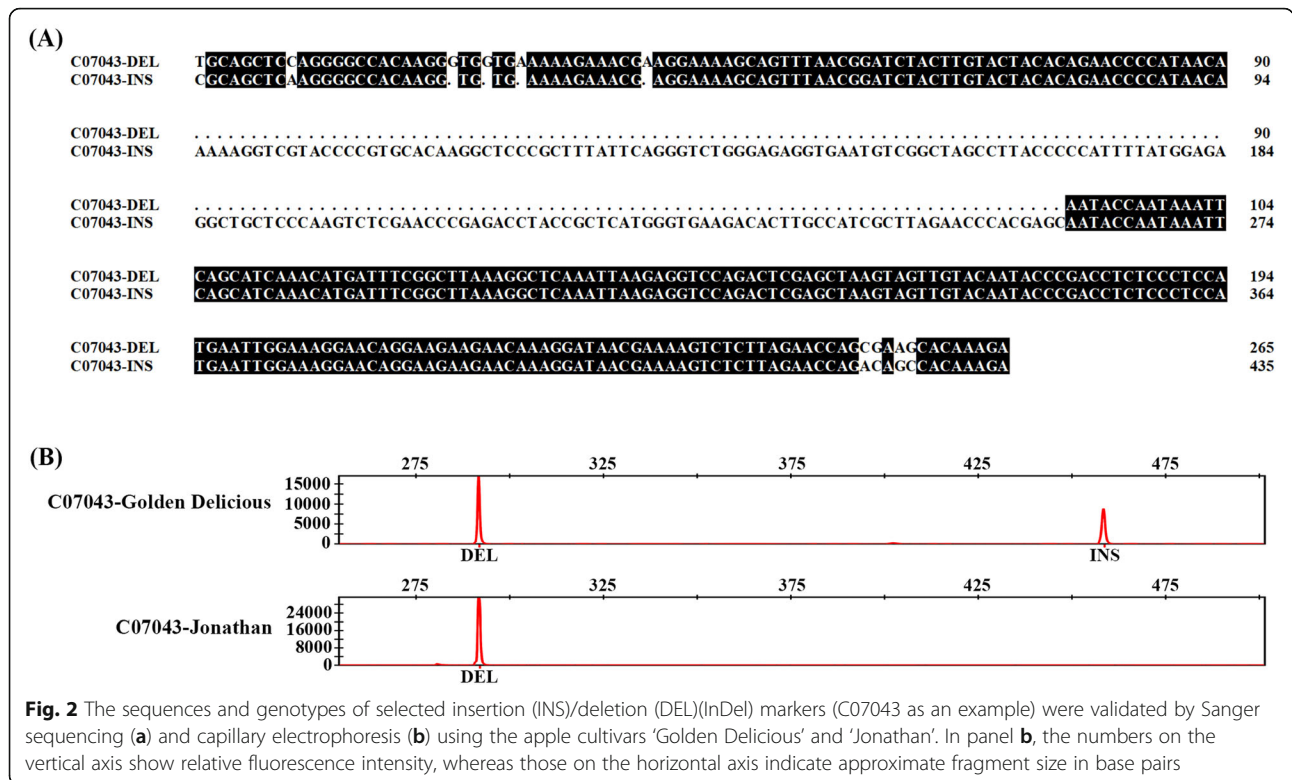


Fig. 1 The properties of structural variants (SVs) and the genome-wide distribution of insertion/deletion (InDel) markers selected between apple (*Malus domestica*) cultivars ‘Golden Delicious’ and ‘Jonathan’. **a** Proportion of each type of SV. DEL: deletion; INS: insertion; INV: inversion; ITX: intra-chromosomal translocation; CTX: inter-chromosomal translocation. Percentages and numerals in brackets indicate the proportion and number of different types of SV, respectively. **b** The fragment length of INS and DEL. **c** The genome-wide distribution of SVs and the 102 selected InDel markers. The rectangles in the outer-most whirl represent the chromosomes, the SVs cannot be reliably unanchored to any chromosome were marked by ‘unanchored’. The chromosome number and the physical position are labeled on the edges of the plot. The inner whirls represent the distribution of DEL, INS, INV, ITX, and CTX on each chromosome. The lines connecting in the center of the figure indicate the corresponding positions before and after the shifts due to ITX and CTX. The value corresponding to the chromaticity bar represents the logarithm of the number of SVs in the range of 0.2 Mb on the chromosome. ‘-1’ on the chromaticity bar corresponds to no SVs in the range of 0.2 Mb



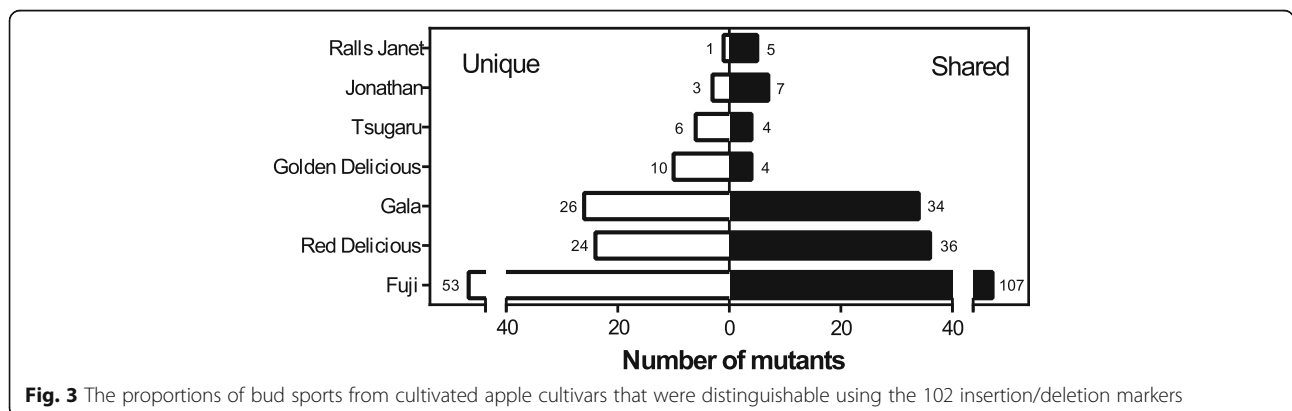
these 1018 accessions with a 2-dimensional bar code (QR code) conveying the 102 InDel marker genotypes (Supplementary File S1).

Identification of somatic mutants

In this study, 331 of the included 981 *Malus domestica* Borkh. accessions were bud sports derived from commercially used cultivars. These mutants included 160 bud sports of ‘Fuji’, 60 bud sports of ‘Red Delicious’, 60 bud sports of ‘Gala’, and 40 bud sports derived from each of the following cultivars: ‘Golden Delicious’, ‘Tsugaru’, ‘Jonathan’, and ‘Ralls Janet’ (Fig. 3).

Fifty-three (33.1%) of the 160 ‘Fuji’ mutants were distinguished from the other ‘Fuji’ bud sports. The remaining

107 ‘Fuji’ bud sports were classified into 11 subgroups, each composed of two to 75 accessions (Fig. 3; Tables S3 and S4). Similarly, among the ‘Red Delicious’ bud sports, the genotypes of 24 (40.0%) of the 60 accessions were distinct from the other ‘Red Delicious’ bud sports, whereas the other 36 bud sports shared three genotype combinations (Fig. 3; Tables S3 and S4). Twenty-six (43.3%) of the 60 ‘Gala’ bud sports were distinct using these InDel markers; 34 bud sports showed six genotype combinations (Fig. 3; Tables S3 and S4). Regarding the bud sports of ‘Golden Delicious’, ‘Tsugaru’, ‘Jonathan’, and ‘Ralls Janet’, 10 of the 14 (71.4%), six of the 10 (60.0%), three of 10 (30.0%), and one of the six (16.7%) were uniquely distinguished, respectively (Fig. 3; Tables S3 and S4).



Furthermore, we compared the marker genotypes of 24 bud sports derived from ‘Fuji’, ‘Gala’, and ‘Red Delicious’ with the corresponding wildtype cultivars. The wildtype cultivar (e.g. ‘Starking’) of a certain bud sport (e.g. ‘Starkrimson’) refers to the cultivar from which the bud sport has been selected. A wildtype cultivar (e.g. ‘Starking’) can sometimes also be a bud sport of an older cultivar (e.g. ‘Red Delicious’). The genotypes of the 102 InDel markers of five bud sports were identical to the corresponding wildtype cultivars (Table S6). Polymorphisms in at least one marker were detected in 19 bud sports compared with the corresponding wildtype cultivars.

Genetic composition of the InDel markers

Five genotype distribution patterns were detected among the 78 biallelic InDel markers using the unique 1002 diploid accessions (Fig. 4; Table S7). Pattern I (38 markers) was characterized by the relatively low frequency (7.0%) of homozygous INS in *M. domestica* compared with the extremely high frequency (71.9%) of homozygous DEL genotypes. In other species than *M. domestica*, much lower frequency (2.0%) of homozygous INS genotypes were detected and the frequency of genotypes with heterozygous DEL:INS was also relatively low (Fig. 4; Table S7). Four markers showed pattern II genotype distribution, where the homozygous DEL genotypes were detected in low frequencies in *M. domestica* and were rare or completely absent in other species (Fig. 4; Table S7). Pattern III (11 markers) exhibited no obvious distortion in marker genotype frequency distribution but few marker/species combinations complied with Hardy-Weinberg equilibrium (Fig. 4; Table S7). Pattern IV (9 markers) was characterized by extremely high frequency (80.0%) of heterozygous DEL:INS genotype in every species, except for five markers in *M. baccata*, of which the frequency of homozygous DEL genotypes was higher (Fig. 4; Table S7). Pattern V (16 markers) showed the same pattern as pattern III (Fig. 4; Table S7).

Parentage analysis

The parentage analysis allowed for the identification of the parent-offspring relationships among the accessions. The parentage of 66 cultivars was confirmed (Table S8), and the documented parentage of six accessions was found to be incorrect (Table 1). The cultivar ‘53–205’, which was believed to be a hybrid from ‘Jonathan’ × ‘Golden Delicious’, was found to be a first-generation offspring of ‘Jonathan’ × ‘Miyazaki Spur Fuji’. Two supposed full-siblings, ‘33–018’ and ‘33–101’, hybrids with the parents ‘Zisai Pearl’ × ‘Golden Delicious’, were identified to be half-siblings instead; the male parents were ‘Miyazaki Spur Fuji’ and unknown, respectively. Similarly, the parentage of ‘H5–101’, ‘50–32’, and ‘62–45’ was corrected (Table 1). The unknown parents of seven

cultivars were hypothesized based on the parent-offspring relationships. For example, ‘Harlikar’ was selected in Japan from an open pollinated progeny of ‘Golden Delicious’. Herein, we propose that the paternal parent was ‘Oregon Spur II’ or a related somatic mutant of ‘Red Delicious’ (Table 1).

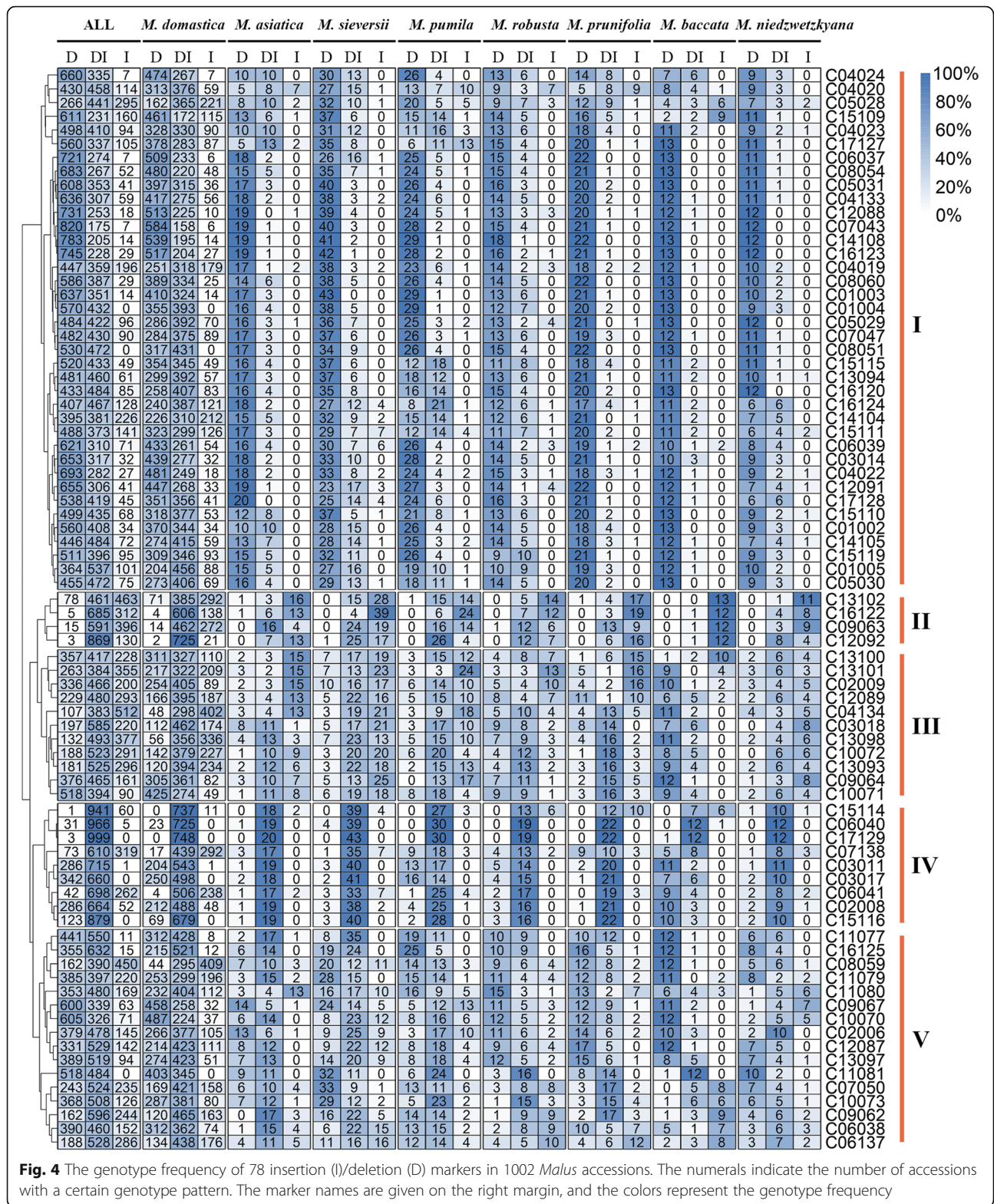
Genetic structure analysis

A genetic structure analysis was generated based on 173 accessions of seven *Malus* species (Table S9). All seven *Malus* species showed relatively low inbreeding coefficients, indicating a low level of population structure within these species (Table 2). Both the highest expected (*He*) and the highest observed heterozygosity (*Ho*) were obtained in *M. domestica*. Conversely, the lowest level of genetic diversity was detected in *M. baccata*, as shown by the lowest *He* and *Ho* (Table 2). Similarly, the highest and the lowest average number of effective alleles were observed in *M. domestica* and *M. baccata*, respectively (Table 2).

To estimate the genetic differentiation between *Malus* species, pairwise differentiation (*Fst*) values were calculated and all *Fst* values were highly significant ($P < 0.001$) (Table 3). The highest level of genetic differentiation was found between *M. baccata* and all of the other species ($Fst = 0.061–0.129$). The differentiation between *M. domestica* and the six other species ($Fst = 0.033–0.129$) was higher compared to the other five species ($Fst = 0.02–0.037$) (Table 3).

Genetic discrimination between the seven species was confirmed through a multivariate Principal Component Analysis (PCA) (Fig. 5a). In the bi-dimensional plot, we found that the two species *M. domestica* and *M. baccata* were completely separate (Fig. 5a). *M. asiatica* was divided into two groups; one was distributed in the lower right corner adjacent to *M. sieversii*, while the other was admixed with *M. domestica*. Most accessions of *M. pumila* admixed with *M. sieversii*, whereas *M. robusta* and *M. prunifolia* were scattered with other species (Fig. 5a).

Relationships among the accessions of the seven different *Malus* species were also depicted using a phylogenetic analysis (Fig. 5b). Our results showed that most accessions of *M. sieversii*, *M. pumila*, and *M. baccata* formed separate clades. Twelve of the 20 *M. asiatica* accessions were typically found to be closely related to *M. sieversii*. *M. robusta* and *M. prunifolia* were largely clustered in the same clade (Fig. 5b). *M. baccata* was found to be basal to the other six species, whereas *M. domestica* was at the distal end of the phylogenetic tree (Fig. 5b). A subset of 13 *M. sieversii* and 12 *M. robusta* accessions were clustered close to *M. domestica*. Several accessions of *M. asiatica*, *M. robusta*, *M. prunifolia*, *M. sieversii*, and *M. pumila* were scattered in the large *M. domestica* clade (Fig. 5b).



Finally, relationships among the *Malus* species were explored using ADMIXTURE cross-validation, which indicated that K=6 was a sensible modeling choice; the other inflection points were K=4 and K=7 (Figure S1).

Thus, the three Q estimates (K=4, 6, and 7) were plotted separately (Fig. 5c). At both K=6 and K=4, *M. domestica*, *M. sieversii*, *M. pumila*, and *M. baccata* were clustered into separate gene pools. *M. sieversii*

Table 1 Newly proposed parentage of 13 *Malus domestica* accessions (> 0.98 confidence)

Accession number	Accession name	Documented parentage	Newly proposed parentage
YX-33-018	33-018	Zisai Pearl × Golden Delicious	Zisai Pearl × Miyazaki spur Fuji
YX-33-101	33-101	Zisai Pearl × Golden Delicious	Zisai Pearl × Unknown
YX-53-205	53-205	Jonathan × Golden Delicious	Jonathan × Miyazaki Spur Fuji
CL-5	H5-101	Golden Delicious × Jonathan	Golden Delicious × Fuji
CL-4	50-32	Miyazaki Spur Fuji × Starkrimson	Miyazaki Spur Fuji × Unknown
XY-68	62-45	Hanfu × Yueshuai	Fuji × Yueshuai
WH-8	Harlikar	Golden Delicious × Unknown	Golden Delicious × <i>Oregon spur II</i>
BK-DANXIA	Danxia	Golden Delicious × Unknown	Golden Delicious × <i>Red Delicious</i>
BK-YG	Yoko	Golden Delicious × Unknown	Golden Delicious × <i>Jonathan</i>
7--23	Youyi	Jonathan × Unknown	Jonathan × <i>Summer Pearmain</i>
B-2	Jiping 1	Mato 1 × Unknown	Mato 1 × Gala
4--7	Fuhong	Unknown	<i>Jonathan</i> × <i>Red Astrachan</i>
8--19	Shennong 2	Unknown	<i>Golden Delicious</i> × Unknown

Note: The italic font indicates the newly proposed parentage which was unknown before, and the bold font indicates the error in the documented parentage was corrected

differentiated into two subdivisions, one of which (blue) exhibited introgression into *M. domestica* and *M. robusta*. The other subdivision (yellow) of *M. sieversii* showed apparent gene flow into *M. asiatica* and *M. prunifolia* (Fig. 5c). Introgression was also detected from *M. baccata* into other species, especially *M. prunifolia* and *M. robusta*. When $K = 7$, *M. prunifolia* clustered into a separate gene pool and showed gene flow (orange) into *M. robusta* (Fig. 5c).

Discussion

Benefited from the high quality assemblies of the apple genome [13, 69, 75], large scale of SNP and InDel markers can be easily obtained [48]. In this study, we detected a total of 25,924 InDels between the two cultivars, 'Jonathan' and 'Golden Delicious'. These InDels provided a large reservoir for high performance PCR-based DNA markers for the *Malus* genus [41, 66]. One hundred and two of these InDel markers were applied in this study to the following analyses in *Malus* accessions.

Table 2 Summary of genetic variation in seven *Malus* species

	Na	Ne	I	Ho	He	F
<i>M. domestica</i>	2.000	1.774	0.618	0.508	0.429	-0.173
<i>M. sieversii</i>	1.987	1.544	0.484	0.367	0.320	-0.097
<i>M. asiatica</i>	1.987	1.543	0.481	0.404	0.318	-0.172
<i>M. pumila</i>	2.000	1.581	0.502	0.396	0.335	-0.122
<i>M. robusta</i>	2.000	1.598	0.532	0.383	0.354	-0.063
<i>M. prunifolia</i>	1.936	1.467	0.418	0.315	0.274	-0.042
<i>M. baccata</i>	1.795	1.277	0.294	0.180	0.181	-0.010

Na: average number of alleles; Ne: average number of effective alleles; I: Shannon's diversity index; Ho: observed heterozygosity; He: expected heterozygosity; F: inbreeding coefficient

The application of genome-wide InDel markers on the genetic structure analysis of *Malus* accessions

Seventy-eight bi-allele InDel markers were used in investigating the relationship between the seven *Malus* species. Lower H_e and H_o values, as well as lower average number of effective alleles were detected in the other six species than in *M. domestica* with the lowest values detected in *M. baccata*. Although these lower levels of H_e , H_o , and N_e could indicate lower levels of genetic diversity in these species, low values could also be observed because the InDel markers were developed from two *M. domestica* cultivars. The lowest H_e , H_o and N_e values were detected in *M. baccata* indicating a low level of genetic diversity in this species. Additionally, the phylogenetic analysis and the structure analysis showed less genetic relatedness of *M. baccata* accessions to the other species. The two subdivisions of *M. sieversii*, however, showed gene introgression into *M. domestica* or *M. robusta* and *M. asiatica* or *M. prunifolia*, respectively. These data were highly consistent with the bi-directional gene flow of *M. sieversii*, which is believed to be the common ancestral species of *M. domestica* and ancient Chinese apple cultivars [15, 69]. *M. domestica* was domesticated in Central Asia from *M. sieversii*, but as it migrated westwards, it hybridized with the European crabapple *M. sylvestris* and/or *M. orientalis*, from which modern apples are descended [10]. However, the DNA ITS1 sequences and genomic regions used in previous studies were not informative for discriminating the samples of *M. sylvestris*, *M. sieversii*, and *M. domestica* [8, 59]. When ancient *M. domestica* moved eastwards, it hybridized with several local wild or semi-cultivated relatives to create Chinese domesticated landrace cultivars, such as 'Nai', which is highly similar to *M. sieversii*

Table 3 Pairwise differentiation (*Fst*) between the seven *Malus* species

	<i>M. domestica</i>	<i>M. sieversii</i>	<i>M. asiatica</i>	<i>M. pumila</i>	<i>M. robusta</i>	<i>M. prunifolia</i>	<i>M. baccata</i>
<i>M. domestica</i>	–						
<i>M. sieversii</i>	0.051	–					
<i>M. asiatica</i>	0.054	0.028	–				
<i>M. pumila</i>	0.046	0.028	0.034	–			
<i>M. robusta</i>	0.033	0.033	0.028	0.033	–		
<i>M. prunifolia</i>	0.074	0.033	0.020**	0.037	0.026	–	
<i>M. baccata</i>	0.129	0.107	0.091	0.104	0.069	0.061	–

Note: All *Fst* values were significant at $P < 0.0001$, except for the number marked with ** which indicates $P < 0.005$

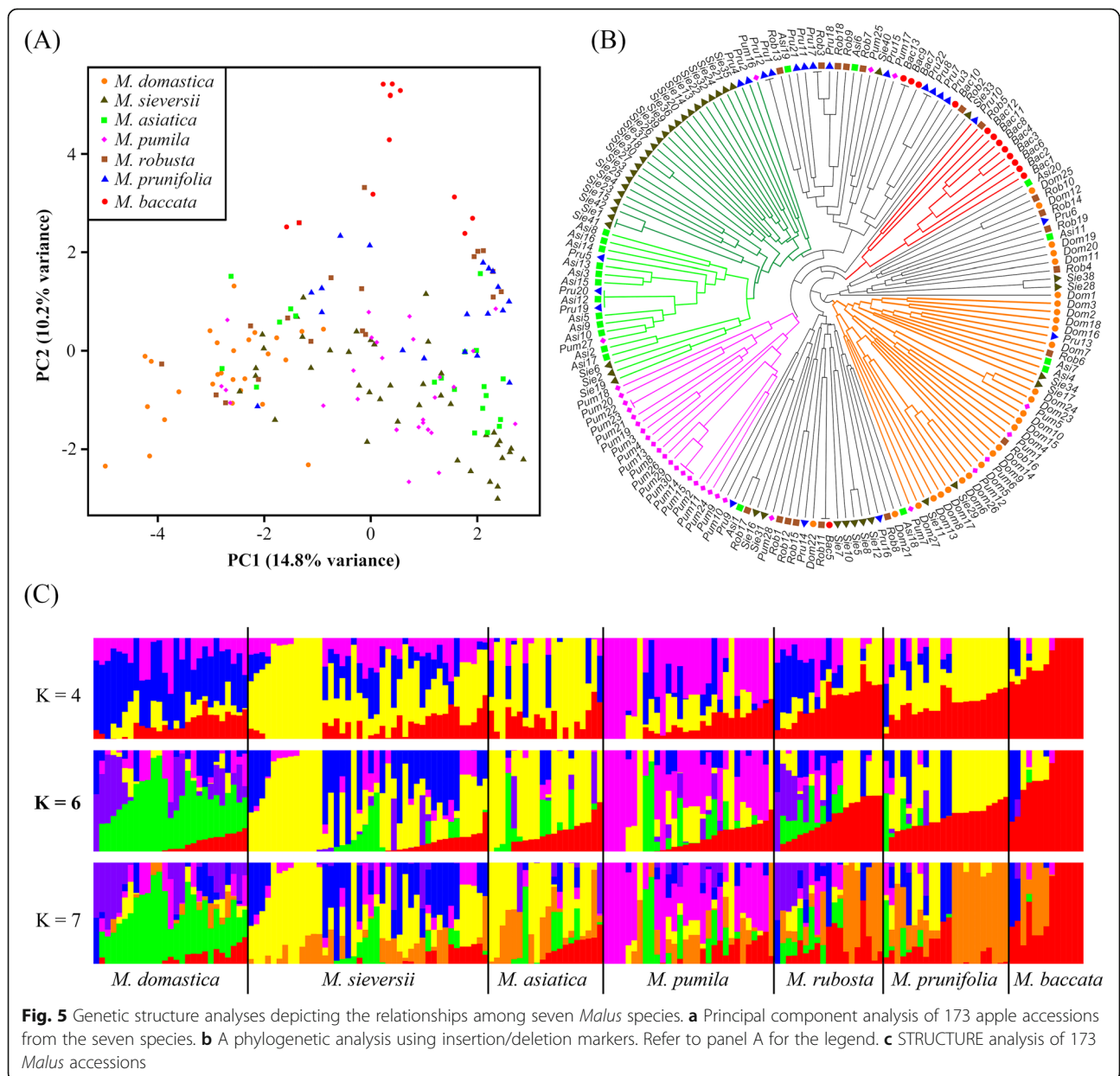


Fig. 5 Genetic structure analyses depicting the relationships among seven *Malus* species. **a** Principal component analysis of 173 apple accessions from the seven species. **b** A phylogenetic analysis using insertion/deletion markers. Refer to panel A for the legend. **c** STRUCTURE analysis of 173 *Malus* accessions

and contained a small signature from other wild apple species [15, 39]. The close genetic relatedness of *M. asiatica* or *M. prunifolia* to *M. baccata* and *M. sieversii* identified in this study supports the previous hypothesis that Chinese native species, such as *M. asiatica* and *M. prunifolia*, are very likely to be hybrids between *M. baccata* and *M. sieversii* [15].

Genetic diversity in domesticated species is often affected by intentional artificial selection and unintentional genetic bottlenecks [9]. Over the last 800 years, *M. domestica* showed no significant reduction in genetic diversity [23], which can possibly be explained by the wild-to-crop introgression [12]. Interspecific hybridization may be an important mechanism for germplasm diversification, and similar genes across multiple species underlies parallel/convergent phenotypic evolution between taxa [53]. The highest level of genetic diversity among the seven *Malus* species was observed in *M. domestica*, indicated by the highest *He* and *Ho*. During domestication and evolution, both the modern deliberate selection and past natural selection may gradually change the genetic composition of a species [53]. We found the genetic composition differed among the InDel markers and *Malus* species.

In this study, the low inbreeding coefficients of all the seven species were consistent with the high level of gametophytic self-incompatibility in *Malus* [14, 42, 73]. The lowest inbreeding coefficient was detected for *M. domestica* and *M. asiatica*, which could be explained by the artificial selection for cultivars with high levels of heterozygosity [12].

The highest inbreeding coefficient observed among the seven species in this study was in *M. baccata*. This observation is likely an artifact from our marker development panel, which consisted only of *M. domestica* accessions. Because *M. baccata* is rather distantly related to *M. domestica* [12], our markers were likely not as informative in this species.

In this study, we found relatively high differentiation between *M. domestica* and the other species. While wild-to-crop gene flow may occur naturally, anthropogenic factors, such as apple production and the variations in apple flower visitors, significantly impact wild-to-crop gene flow [9]. We observed that several accessions of *M. asiatica*, *M. robusta*, *M. prunifolia*, *M. sieversii*, and *M. pumila* scattered in the *M. domestica* clade (Fig. 5b). This is similar to previous findings that showed high levels of introgression from *M. domestica* detected in *M. orientalis* (3.2% of hybrids), *M. sieversii* (14.8%), and *M. sylvestris* (36.7%) [11]. Conversely, gene flow from domesticated-to-wild accessions or escapes from cultivated *M. domestica* threatens the fitness and the genetic integrity of wild

relatives; therefore, it is important to conserve wild germplasm resources [6, 20].

The application of genome-wide InDel markers to delineate the identity signature of *Malus* accessions

Identity signatures of 1018 *Malus* accessions were created as QR codes using the 102 InDel markers in this study. These QR codes can not only used for DUST within *Malus*, but also can distinguish some of the bud sports of apple cultivars. Early studies attempted to distinguish bud sports of apple cultivars with amplified fragment length polymorphism markers; however, the efficiency was low [37, 76]. Recent studies have had limited success distinguishing clonal mutants because the high levels of clonality or homogeneity among cultivars derived from bud sports [12]. A previous study used two InDel markers to efficiently and specifically distinguish ‘Fuji’ and its somatic variant ‘Benishogun’ from four other bud sport cultivars [33]. In this study, the 102 InDel markers discriminated successfully 33.1, 40.0, 43.2, and 71.4% of bud sports of ‘Fuji’, ‘Red Delicious’, ‘Gala’, and ‘Golden Delicious’. There would be three reasons why the bud sports cannot be fully distinguishable. The first is that some bud sports are genetically identical due to the parallel or reproducible occurrence of somatic variations in fruit crops [3, 29]. The second aspect that hinders the genetic identification of bud sports is chimeric forms of somatic variation in fruit crops [17–19, 70]. Epigenetic variations may be the third source of clonal differences that have been difficult to be detected genetically [45, 61, 71].

The application of genome-wide InDel markers for lineage tracing of *Malus* accessions

Many apple cultivars, such as ‘Red Delicious’, ‘Golden Delicious’, and ‘Ralls Janet’, originated from chance seedlings and one or even both parents of these cultivars are unknown [44]. Lineage tracing back of cultivars with unknown parentage has been pioneered in ‘Honeycrisp’ by SSR markers and SNP linkage maps [4, 27]. Most recently, the parent-offspring relationships of 1400 apple cultivars were analyzed with whole-genome SNPs [44]. By using the 102 InDel markers in this study, the previously reported parentage of 66 cultivars was corrected, whereas previously unknown parents of seven cultivars, such as ‘Harlikar’, were identified (Table 1). To elucidate the pedigree or the genetic background of cultivars with unknown parentage, the cost of using these 102 InDel markers should be lower than the available SNP arrays ([2, 5, 27, 38];). However, it would be impossible for these InDels to compose haplotypes, as has been done in

some previous studies (e.g. [27]), due to the marker density being too low.

Conclusions

One hundred and two stable co-dominant long InDel markers were developed in *Malus*. Identity signatures of 1018 *Malus* accessions were created as QR codes using these markers. The QR codes can not only be used for DUST, but also can efficiently distinguish some bud sports of apple cultivars. These markers were also used in the analysis of parent-offspring relationship to determine the previously unknown parentage. The application of these InDel markers on the genetic structure analysis also provided insight into the genetic relationships among *Malus* species.

Methods

Plant materials

We sampled and analyzed a collection of 1251 *Malus* accessions, including 981 accessions of *M. domestica* Borkh., 49 accessions of *M. sieversii* (Ledeb.) Roem., 20 accessions of *M. asiatica* Nakai, 31 accessions of *M. pumila* Mill., 21 accessions of *M. robusta* Rehder, 25 accessions of *M. prunifolia* (Wild.) Borkh., 13 accessions of *M. baccata* (L.) Borkh., and 111 other species (Table S3). All the plant materials are originally collected and possessed by China Agricultural University and Chinese Academy of Agricultural Science, respectively. The experiments on plants including field investigation and sample collection were performed under institutional guidelines in accordance with local legislation. Young leaf samples were collected and stored on silica gel. The genomic DNA was extracted using the modified CTAB protocol [62].

Calling of SV from previous resequencing data of 'Jonathan' and 'Golden delicious'

SV were called using Delly (version 0.8.1) software [54]. The BAM files from the cultivars 'Jonathan' (SRX4380657) and 'Golden Delicious' (SRX4380658) ([60]; <https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA392908>) were fed into the Delly call function with default parameters to call SVs. The distribution of the obtained SVs in the genome of the accessions was presented using Circos (version 0.69–8) software [31].

Selection and genotyping of InDel markers for all accessions

One hundred and seventy InDels with 50–400 bp polymorphic fragments were selected to develop InDel markers, ten InDels were selected in each of the 17 chromosomes. The InDel fragments were validated between 'Jonathan' and 'Golden Delicious' by Sanger sequencing and capillary electrophoresis analysis. Only the

markers that were confirmed to produce unique, valid amplified products and were used for further analysis.

The multi-PCR forward primers of each InDel markers were labeled with the fluorescent dyes FAM, HEX, NED, and PET (Table S2). Multi-PCR was performed in a final volume of 10 μ L containing 1 μ L of DNA template (200 ng), 1 μ L of primer mix, 4 μ L of 2.5 \times Master Mix I (Beijing Microread Genetic Co., Ltd., Beijing, China), and 4 μ L of double distilled water (ddH₂O). The thermocycler conditions were set as follows: pre-incubation at 95 $^{\circ}$ C for 5 min; followed by 35 cycles of 30 s for denaturing at 95 $^{\circ}$ C, 90 s for annealing at 55 $^{\circ}$ C, and 90 s for elongation at 72 $^{\circ}$ C; and a final extension 72 $^{\circ}$ C for 15 min. Amplified products were stored at 12 $^{\circ}$ C until analysis with an ABI3730 XL sequencing system (Applied Biosystems, Foster City, CA, USA). Fragment and sizing analyses were carried out using GeneMapper v.5.0 software (Applied Biosystems, Foster City, CA, USA), and chromatograms were independently read by two operators.

The identity signature of the accessions was represented by the unique genotype combination of the 102 InDel markers. Then the genotype information from the accessions was used to create a QR code using an online tool (<https://cli.im/>).

Identification of genetic composition of the InDel markers

For the genetic composition analysis, only the unique 1002 diploid accessions were used and the 78 biallelic InDel markers were selected from all markers and were used in the analysis. The results of genetic composition were visualized by a heatmap using the pheatmap package (<https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12>) with default clustering method in R.

Genetic structure analysis

For the genetic structure analysis, the 78 bi-allele InDel markers were used and 173 *Malus* germplasm accessions with unique genotype combinations were selected, including 27 randomly chosen from *M. domestica*, and all accessions in relative species (42 *M. sieversii*, 20 *M. asiatica*, 30 *M. pumila*, 19 *M. robusta*, 22 *M. prunifolia*, and 13 *M. baccata*) (Table S9). Known polyploid accessions were not included here to ensure bi-allele genetic composition. *He* and *Ho* were estimated with GenAEx 6.5 [49, 50]. *Fst* between species was assessed in exact tests using GENEPOP 4.0 [55, 57].

To elucidate the genetic relationship among accessions, a PCA was performed using the pca3d (version 0.10) package in R [72]. A phylogenetic tree was built using the ape (version 5.3) package in R [47]. A population structure analysis was performed using the block relaxation algorithm implemented in ADMIXTURE

(version 1.3) software [1]. We generated the associated support files using PLINK (version 1.90) software [52].

Parentage analysis

To determine the parentage of some *M. domestica* cultivars, the parent-offspring relationships of accessions with one or two unknown parents were analyzed based on the genotype data of the 102 InDel markers using a custom Python script, AppleParentage1.0 software (<https://github.com/wangx321/AppleParentage1.0>). The confidence parameters were set to > 0.98 (Threshold = 1).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-020-02744-2>.

Additional file 1: Supplementary File S1. The QR code giving the molecular ID of 1018 *Malus* accessions.

Additional file 2: Supplementary Table S1. Sequences, PCR product sizes and primers of the 102 InDel markers.

Additional file 3: Supplementary Table S2. Fluorescent labelled multiplex PCR matching schemes of the 102 insertion/deletion markers.

Additional file 4: Supplementary Table S3. Genotypes of the 102 InDel markers for 1,251 *Malus* accessions.

Additional file 5: Supplementary Table S4. *Malus* accessions with shared genotype combinations of the 102 InDel markers.

Additional file 6: Supplementary Table S5. Synonyms, homonyms and other *Malus* accessions with incorrect names detected by the 102 InDel markers.

Additional file 7: Supplementary Table S6. Comparison of genotypes of InDel markers between bud sports and corresponding wild-type *Malus domestica* cultivars.

Additional file 8: Supplementary Table S7. Genotype and allele frequencies of 78 bi-allele InDel markers in 1,002 diploid accessions from eight *Malus* species.

Additional file 9: Supplementary Table S8. Parentage analysis of 66 *Malus domestica* cultivars using InDel markers.

Additional file 10: Supplementary Table S9. *Malus* accessions used for genetic structure analysis.

Additional file 11: Supplementary Figure S1. Cross-validation plot for the InDel dataset.

Abbreviations

CTAB: Cetyltrimethylammonium bromide; CTX: Inter-chromosomal translocation; DEL: Deletion; DUST: Test for distinctness, uniformity, and stability; *Fst*: Fixation index 'F-statistics'; *He*: Expected heterozygosity; *Ho*: Observed heterozygosity; InDel: Insertion and deletion; INS: Insertion; INV: Inversion; ITS1: Internal transcribed spacer 1; ITX: Intra-chromosomal translocation; PCA: Principal component analysis; PCR: Polymerase chain reaction; QR code: 2-dimensional bar code; QTL: Quantitative trait loci; SNP: Single nucleotide polymorphism; SRA: Sequence Read Archive; SSR: Simple sequence repeat; SV: Structural variant

Acknowledgements

The authors thank the Key Laboratory of Stress Physiology and Molecular Biology for Fruit Trees in Beijing Municipality, the Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (Nutrition and Physiology), Ministry of Agriculture, and the Construction of Beijing Science and Technology Innovation and Service Capacity in Top Subjects (CEFF-PXM2019_014207_000032).

Authors' contributions

X.Z. and Z.H. initiated and designed the experiments. F.S. and X.W. performed the bioinformatic analysis. X.W., R.C., Y.G., K.W. and P.C. collected and preserved plant materials. X.W., R.C., Y.W., X.X., T.W., W.L., C.Q. and Xi Z. performed the experiments. X.W., J.L. and L.Y. performed the parentage analysis. X.W. and X.Z. wrote the paper. All authors have read and approved the manuscript.

Funding

This work was funded by the earmarked fund of the China Agriculture Research System (CARS-27). The views expressed in this work are the sole responsibility of the authors. The funding body did not play any roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All DNA re-sequencing reads are freely available and have been upload to Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA392908>) already.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Horticulture, China Agricultural University, Beijing, China. ²Beijing Agro-biotechnology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China. ³Research Institute of Pomology, Chinese Academy of Agricultural Science, Xingcheng, Liaoning, China. ⁴Present Address: Shaanxi Haisheng Fruit Industry Development Co., Ltd., Shaanxi, Xian, China. ⁵College of Information and Electrical Engineering, China Agricultural University, Beijing, China.

Received: 29 June 2020 Accepted: 17 November 2020

Published online: 30 November 2020

References

- Alexander H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
- Bianco L, Cestaro A, Sargent DJ, Banchi E, Derdak S, Di Guardo M, Salvi S, Jansen J, Viola R, Gut I, Laurens F, Chagne D, Velasco R, van de Weg E, Troglio, M. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus × domestica*Borkh.). *PLoS One.* 2014;9:e110377.
- Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martina C. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell.* 2012;24:1242–55.
- Cabe PR, Baumgarten A, Onan K, Luby JJ, Bedford DS. Using microsatellite analysis to verify breeding records: a study of 'Honeycrisp' and other cold-hardy apple cultivars. *HortScience.* 2005;40:15–7.
- Chagne D, Crowhurst RN, Troglio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees JD, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner SE, Bassil N, Peace C. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS One.* 2012;7:e31745.
- Coart E, van Glabeke S, de Loose M, Larsen AS, Roldan-Ruiz I. Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). *Mol Ecol.* 2006;15:2171–82.
- Cornille A. Diversification the genus *Malus*. PhD thesis. Orsay: French National Centre for Scientific Research, University of Paris Sud; 2012.
- Cornille A, Antolin F, Garcia E, Vernesi C, Fietta A, Brinkkemper O, Kirleis W, Schlumbaum A, Roldan-Ruiz I. A multifaceted overview of apple tree domestication. *Trends Plant Sci.* 2019;24:770–82.

9. Cornille A, Feurtey A, Gelin U, Ropars J, Misvanderbrugge K, Gladieux P, Giraud T. Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples. *Evol Appl.* 2015;8:373–84.
10. Cornille A, Giraud T, Smulders MJ, Roldan-Ruiz I, Gladieux P. The domestication and evolutionary ecology of apples. *Trends Genet.* 2014;30:57–65.
11. Cornille A, Gladieux P, Giraud T. Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evol Appl.* 2013;6:737–48.
12. Cornille A, Gladieux P, Smulders MJM, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyanyan A, Clavel A, Olonova M, Feugey L, Gabrielyan I, Zhang X-G, Tenaillon MJ, Giraud T. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet.* 2012;8:e1002703.
13. Daccord N, Celton JM, Linsmith G, Becker C, Choise N, Schijlen E, van de Geest H, Bianco L, Micheletti D, Velasco R, Di Pierro EA, Gouzy J, Rees DJG, Guerif P, Muranty H, Durel CE, Laurens F, Lespinasse Y, Gaillard S, Aubourg S, Quesneville H, Weigel D, van de Weg E, Troggio M, Bucher E. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet.* 2017;49:1099–106.
14. De Franceschi P, Pierantoni L, Dondini L, Grandi M, Sansavini S, Sanzoli J. Evaluation of candidate F-box genes for the pollen S of gametophytic self-incompatibility in the Pyrinae (*Rosaceae*) on the basis of their phylogenomic context. *Tree Genet Genomes.* 2011;7:663–83.
15. Duan NB, Bai Y, Sun HH, Wang N, Ma YM, Li MJ, Wang X, Jiao C, Legall N, Mao LY, Wan SB, Wang K, He TM, Feng SQ, Zhang ZY, Mao ZQ, Shen X, Chen XL, Jiang YM, Wu SJ, Yin CM, Ge SF, Yang L, Jiang SH, Xu HF, Liu JX, Wang DY, Qu CZ, Wang YC, Zuo WF, Xiang L, Liu C, Zhang DY, Gao Y, Xu YM, Xu KN, Chao T, Fazio G, Shu HR, Zhong GY, Cheng LL, Fei ZJ, Chen XS. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat Commun.* 2017;8:249.
16. Evans KM, Patocchi A, Rezzonico F, Mathis F, Durel CE, Fernandez-Fernandez F, Boudichevskaia A, Dunemann F, Stankiewicz-Kosyl M, Gianfranceschi L, Komjanc M, Lateur M, Madduri M, Noordijk Y, van de Weg WE. Genotyping of pedigreed apple breeding material with a genome-covering set of SSRs: trueness-to-type of cultivars and their parentages. *Mol Breed.* 2011;28:535–47.
17. Fernandez L, Chaib J, Martinez-Zapater J-M, Thomas MR, Torregrosa L. Mis-expression of a PISTILLATA-like MADS box gene prevents fruit development in grapevine. *Plant J.* 2013;73:918–28.
18. Fernandez L, Doligez A, Lopez G, Thomas MR, Bouquet A, Torregrosa L. Somatic chimerism, genetic inheritance, and mapping of the fleshless berry (flb) mutation in grapevine (*Vitis vinifera* L.). *Genome.* 2006;49:721–8.
19. Fernandez L, Romieu C, Moing A, Bouquet A, Maucourt M, Thomas MR, Torregrosa L. The grapevine fleshless berry mutation: a unique genotype to investigate differences between fleshy and nonfleshy fruit. *Plant Physiol.* 2006;140:537–47.
20. Feurtey A, Cornille A, Shykoff JA, Snirc A, Giraud T. Crop-to-wild gene flow and its fitness consequences for a wild fruit tree: towards a comprehensive conservation strategy of the wild apple in Europe. *Evol Appl.* 2017;10:180–8.
21. Foster TM, Aranzana MJ. Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. *Hortic Res.* 2018;5:44.
22. Golubov A, Yao Y, Maheshwari P, Bilichak A, Boyko A, Belzile F, Kovalchuk I. Microsatellite instability in *Arabidopsis* increases with plant development. *Plant Physiol.* 2010;154:1415–27.
23. Gross BL, Henk AD, Richards CM, Fazio G, Volk GM. Genetic diversity in *Malus × domestica* (*Rosaceae*) through time in response to domestication. *Am J Bot.* 2014;101:1770–9.
24. Gross BL, Wedger MJ, Martinez M, Volk GM, Hale C. Identification of unknown apple (*Malus × domestica*) cultivars demonstrates impact local breeding program on cultivar diversity. *Genet Resour Crop Evol.* 2018;65:1317–27.
25. Guan JJ, Zhang P, Huang QM, Wang JM, Yang XH, Chen QB, Zhang JH. SNP markers potential applied in DUS testing of maize. *Int J Agric Biol.* 2020;23:417–22.
26. Guo GJ, Zhang GL, Pan BG, Diao WP, Liu JB, Ge W, Gao CZ, Zhang Y, Jiang C, Wang SB. Development and application of InDel markers for *Capsicum* spp. based on whole-genome re-sequencing. *Sci Rep.* 2019;9:3691.
27. Howard NP, van de Weg E, Bedford DS, Peace CP, Vanderzande S, Clark MD, Teh SL, Cai L, Luby JJ. Elucidation of the 'Honeycrisp' pedigree through haplotype analysis with a multi-family integrated SNP linkage map and a large apple (*Malus × domestica*) pedigree-connected SNP data set. *Hortic Res.* 2017;4:7.
28. Jain A, Roorkiwal M, Kale S, Garg V, Yadala R, Varshney RK. InDel markers: an extended marker resource for molecular breeding in chickpea. *PLoS One.* 2019;16:e0213999.
29. Kobayashi S, Yamamoto NG, Hirochika H. Retrotransposon-induced mutations in grape skin color. *Science.* 2004;304:982.
30. Kitahara K, Matsumoto S, Yamamoto T, Soejima J, Kimura T, Komatsu H, Abe K. Parent identification of eight apple cultivars by S-RNase analysis and simple sequence repeat markers. *HortScience.* 2005;40:314–7.
31. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19:1639–45.
32. Lassois L, Denance C, Ravon E, Guyader A, Guisnel R, Hibrand-Saint-Oyant L, Poncet C, Lasserre-Zuber P, Feugey L, Durel CE. Genetic diversity, population structure, parentage analysis, and construction of core collections in the French apple germplasm based on SSR markers. *Plant Mol Biol Rep.* 2016;34:827–44.
33. Lee HS, Kim GH, Kwon SI, Kim JH, Kwon YS, Choi C. Analysis of 'Fuji' apple somatic variants from next-generation sequencing. *Genet Mol Res.* 2016;15:gmr.15038185.
34. Leforestier D, Ravon E, Muranty H, Cornille A, Lemaire C, Giraud T, Durel CE, Branca A. Genomic basis of the differences between cider and dessert apple varieties. *Evol Appl.* 2015;8:650–61.
35. Li QS, Chen JB, Gu HP, Yuan XX, Chen X, Cui J. Genetic diversity and fingerprint analysis of mungbean varieties from China based on InDel markers. *J Plant Genet Recour.* 2019;20:122–8.
36. Li SG, Shen D, Liu B, Qiu Y, Zhang XH, Zhang ZH, Wang HP, Li XX. Development and application of cucumber InDel markers based on genome resequencing. *J Plant Genet Recour.* 2013;14:278–83.
37. Li YJ, Tang ML, Yu Q, Liu MY, Song LQ, Han ZH. Optimization of AFLP system for Fuji and identification early-maturing sport from 'Changfu 2'. *Acta Hortic Sin.* 2009;36:327–32.
38. Luo FX, van de Weg E, Vanderzande S, Norelli JL, Flachowsky H, Hanke V, Peace C. Elucidating the genetic background of the early-flowering transgenic genetic stock T1190 with a high-density SNP array. *Mol Breed.* 2019;39:21.
39. Ma B, Liao L, Peng Q, Fang T, Zhou H, Korban SS, Han Y. Reduced representation genome sequencing reveals patterns of genetic diversity and selection in apple. *J Integr Plant Biol.* 2017;59:190–204.
40. Marconi G, Ferradini N, Russi L, Concezzi L, Veronesi F, Albertini E. Genetic characterization of the apple germplasm collection in Central Italy: the value of local varieties. *Front Plant Sci.* 2018;9:1460.
41. Markkandan K, Yoo S, Cho Y-C, Lee DW. Genome-wide identification of insertion and deletion markers in Chinese commercial rice cultivars, based on next-generation sequencing data. *Agronomy-Basel.* 2018;8:36.
42. Matsumoto D. Gametophytic self-incompatibility. *Encyclopedia of Applied Plant Sciences 2nd Edition*, vol. 2; 2017. p. 275–80.
43. Moriya S, Iwanami H, Yamamoto T, Abe K. A practical method for apple cultivar identification and parent-offspring analysis using simple sequence repeat markers. *Euphytica.* 2011;177:135–50.
44. Muranty H, Denance C, Feugey L, Crepin JL, Barbier Y, Tartarini S, Ordidge M, Troggio M, Lateur M, Nybom H, Papstein F, Laurens F, Durel C. Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. *BMC Plant Biol.* 2020;20:2.
45. Ong-Abdullah M, Ordway JM, Jiang N, Ooi SE, Kok SY, Sarpan N, Azimi N, Hashim AT, Ishak Z, Rosli SK, Malike FA, Abu Bakar NA, Marjuni M, Abdullah N, Yaakub Z, Amiruddin MD, Nookiah R, Singh R, ETL L, Chan KL, Azizi N, Smith SW, Bacher B, Budiman MA, Van Brunt A, Wischmeyer C, Beil M, Hogan M, Lakey N, Lim CC, Arulandoo X, Wong CK, Choo CN, Wong WC, Kwan YY, SSRS A, Sambanthamurthi R, Martienssen RA. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature.* 2015;525:533–7.
46. Ordidge M, Kirdwichai P, Baksh MF, Venison EP, Gibbings JG, Dunwell JM. Genetic analysis of a major international collection of cultivated apple varieties reveals previously unknown historic heteroploid and inbred relationships. *PLoS One.* 2018;13:e0202405.

47. Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019;35:526–8.
48. Peace CP, Bianco L, Troggio M, van de Weg E, Howard NP, Cornille A, Durel CE, Myles S, Migicovsky Z, Schaffer RJ, Costes E, Fazio G, Yamane H, van Nocker S, Gottschalk C, Costa F, Chagne D, Zhang XZ, Patocchi A, Gardiner SE, Hardner C, Kumar S, Laurens F, Bucher E, Main D, Jung S, Vanderzande S. Apple whole genome sequences: recent advances and new prospects. *Hortic Res*. 2019;6:59.
49. Peakall R, Smouse PE. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol Ecol Notes*. 2006;6:288–95.
50. Peakall R, Smouse PE. GenALEX 6.5: genetic analysis in excel. Population genetic software for teaching and research - an update. *Bioinformatics*. 2012;28:2537–9.
51. Phipps JB, Robertson KR, Smith PG, Rohrer JR. A checklist of the subfamily *Maloideae* (*Rosaceae*). *Can J Bot*. 1990;68:2209–69.
52. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maier J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
53. Purugganan MD. Evolutionary insights into the nature of plant domestication. *Curr Biol*. 2019;29:R705–14.
54. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28:1333–9.
55. Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered*. 1995;86:248–9.
56. Robinson JP, Harris SA, Juniper BE. Taxonomy of the genus *Malus* mill. (*Rosaceae*) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Syst Evol*. 2001;226:35–58.
57. Rousset F. Genepop' 007: a complete re-implementation of the genepop software for windows and Linux. *Mol Ecol Resour*. 2008;8:103–6.
58. Saccomanno B, Wallace M, O'Sullivan DM, Cockram J. Use of genetic markers for the detection of off-types for DUS phenotypic traits in the inbreeding crop, barley. *Mol Breed*. 2020;40:13.
59. Savelyeva EN, Boris KV, Kochieva EZ, Kudryavtsev AM. Analysis of sequences of ITS1 internal transcribed spacer and 5.8S ribosome gene of *Malus* species. *Russ J Genet*. 2013;49:1175–82.
60. Shen F, Huang ZY, Zhang BG, Wang Y, Zhang X, Wu T, Xu XF, Zhang XZ, Han ZH. Mapping gene markers for apple fruit ring rot disease resistance using a multi-omics approach. *G3 Genes Genomes Genet*. 2019;9:1663–78.
61. Telias A, Kui LW, Stevenson DE, Cooney JM, Hellens RP, Allan AC, Hoover EE, Bradeen JM. Apple skin patterning is associated with differential expression of *MYB10*. *BMC Plant Biol*. 2011;11:93.
62. Tel-Zur N, Abbo S, Myslabodski D, Mizrahi Y. Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera *Hylocereus* and *Selenicereus* (*Cactaceae*). *Plant Mol Biol Rep*. 1999;17:249–54.
63. Tysklind N, Blanc-Jolivet C, Mader M, Meyer-Sand BRV, Paredes-Villanueva K, Coronado ENH, Garcia-Davila CR, Sebbenn AM, Caron H, Troispoux V, Guichoux E, Degen B. Development of nuclear and plastid SNP and INDEL markers for population genetic studies and timber traceability of Carapa species. *Conserv Genet Resour*. 2019;11:337–9.
64. UPOV (1991) International convention for the protection of new varieties of plants. International Union for the Protection of new varieties of plants, Geneva (online). Available at: <http://www.upov.int/upovlex/en/conventions/1991/act1991.html>.
65. Urrestarazu J, Denance C, Ravon E, Guyader A, Guisnel R, Feugey L, Poncet C, Lateur M, Houben P, Ordidge M, Fernandez-Fernandez F, Evans KM, Paprstein F, Sedlak J, Nybom H, Garkava-Gustavsson L, Miranda C, Gassmann J, Kellerhals M, Suprun I, Pikunova AV, Krasova NG, Torutaeva E, Dondini L, Tartarini S, Laurens F, Durel CE. Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biol*. 2016;16:130.
66. Vali U, Brandstrom M, Johansson M, Ellegren H. Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. *BMC Genet*. 2008;9:8.
67. Vanderzande S, Howard NP, Cai LC, Linge CD, Antanaviciute L, Bink MCAM, Krusselbrink JW, Bassil N, Gasic K, Iezzoni A, Van de Weg E, Peace C. High-quality genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. *PLoS One*. 2019;14:e0210928.
68. Vanderzande S, Micheletti D, Troggio M, Davey MW, Keulemans J. Genetic diversity, population structure, and linkage disequilibrium of elite and local apple accessions from Belgium using the IRSC array. *Tree Genet Genomes*. 2017;13:125.
69. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoli M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchiotti A, Kater MM, Masiero S, Lasserre P, Lepinasse Y, Allan AC, Bus V, Chagne D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouze P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R. The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat Genet*. 2010;42:833–9.
70. Walker AR, Lee E, Robinson SP. Two new grape cultivars, bud sports of cabernet sauvignon bearing pale-coloured berries, are the result of deletion of two regulatory genes of the berry colour locus. *Plant Mol Biol*. 2006;62: 623–35.
71. Wang Z, Meng D, Wang A, Li T, Jiang S, Cong P, Li T. The methylation of the *PcMYB10* promoter is associated with green-skinned sport in max red Bartlett pear. *Plant Physiol*. 2013;162:885–96.
72. Weiner J 3rd, Parida SK, Maertzdorf J, Black GF, Reipsilber D, Telaar A, Mohney RP, Arndt-Sullivan C, Ganoza CA, Faé KC, Walz G, Kaufmann SHE. Biomarkers of inflammation, immunosuppression and stress are revealed by metabolomic profiling of tuberculosis patients. *PLoS One*. 2012;7:e40221.
73. Wu J, Gu C, Khan MA, Wu J, Gao Y, Wang C, Korban SS, Zhang S. Molecular determinants and mechanisms of gametophytic self-incompatibility in fruit trees of *Rosaceae*. *Crit Rev Plant Sci*. 2013;32:53–68.
74. Wu M, Wang N, Shen C, Huang C, Wen TW, Lin ZX. Development and evaluation of InDel markers in cotton based on whole-genome re-sequencing data. *Acta Agron Sin*. 2019;45:196–203.
75. Zhang LY, Hu J, Han XL, Li JJ, Gao Y, Richards CM, Zhang CX, Tian Y, Liu GM, Gul EA, Wang DJ, Tian Y, Yang CX, Meng MH, Yuan GP, Kang GD, Wu YL, Wang K, Zhang HT, Wang DP, Cong PH. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat Commun*. 2019;10:1494.
76. Zhu J, Wang T, Zhao YJ, Zhang W, Li GC. Identification of apple varieties with AFLP molecular markers. *Acta Hortic Sin*. 2000;27:102–6.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

