



Microsatellite Marker Development and Population Structure Analysis in Japanese Apricot (*Prunus mume* Sieb. et Zucc.)

Numaguchi, Koji ; Ishio, Shogo ; Kitamura, Yuto ; Nakamura, Kentaro ; Ishikawa, Ryo ; Ishii, Takashige

(Citation)

The Horticulture Journal, 88(2):222-231

(Issue Date)

2019

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

© 2019 The Japanese Society for Horticultural Science (JSHS), All rights reserved.

(URL)

<https://hdl.handle.net/20.500.14094/90008009>



Microsatellite Marker Development and Population Structure Analysis in Japanese Apricot (*Prunus mume* Sieb. et Zucc.)

Koji Numaguchi^{1,3}, Shogo Ishio², Yuto Kitamura³, Kentaro Nakamura², Ryo Ishikawa¹ and Takashige Ishii^{1*}

¹Graduate School of Agricultural Science, Kobe University, Kobe 657-8501, Japan

²Tsukuba Research Institute, Sumitomo Forestry Co., Ltd., Tsukuba 300-2646, Japan

³Japanese Apricot Laboratory, Wakayama Fruit Tree Experiment Station, Minabe, Wakayama 645-0021, Japan

Japanese apricot (*Prunus mume* Sieb. et Zucc.) is one of the major fruit tree crops in Japan. However, a paucity of molecular tools has limited studies on the species' genetic diversity and clone identification. Therefore, we newly designed 201 microsatellite markers using the *P. mume* reference genome and selected 20 highly polymorphic markers. The markers showed higher polymorphism detectability than those previously developed using peach and apricot genomes. They were used successfully for fingerprinting most of the *Prunus* cultivars examined (124 *P. mume* accessions and one accession each of *P. armeniaca*, *P. salicina*, *P. persica*, and *P. dulcis*), and the resulting genotype data were used to examine the genetic differentiation of six Japanese apricot cultivar groups, including those producing normal fruit, small-fruit, and ornamental flowers, as well as Taiwanese cultivars, putative hybrids of *P. armeniaca* and *P. mume*, and putative hybrids of *P. salicina* and *P. mume*. Phylogenetic cluster analysis showed three clades with high support values; one clade comprised the putative *P. armeniaca* × *P. mume* hybrids, and the two others included Taiwanese and ornamental cultivars. The rest of the accessions were grouped into two wide clusters, but not clearly divided into the respective cultivar groups. These complex relationships were supported by the principal coordinate and STRUCTURE analyses. Since Japanese apricot is thought to have originated in China, many factors such as human preference, geographical separation, introgression, and local breeding, may have been involved to form the present complex genetic structure in Japanese apricot.

Key Words: clone identification, domestication, fruit cultivars, ornamental cultivars.

Introduction

Japanese apricot (*Prunus mume* Sieb. et Zucc.) is a stone fruit species that is found in the same genus as peach [*P. persica* (L.) Batsch], almond [*P. dulcis* (Mill.) D. A. Webb.], apricot (*P. armeniaca* L.), Japanese plum (*P. salicina* Lindl.), European plum (*P. domestica* L.), and sweet cherry (*P. avium* L.). Japanese apricot is thought to have originated in China and has been cultivated for over 3000 years with over 500 cultivars (Horiuchi et al., 1996). It is believed that the species was introduced to Japan ~2000 years ago (Horiuchi

et al., 1996; Mega et al., 1988), and the annual production of the fruit in Japan recently reached 111,400 tones (MAFF, 2016). The fruits of the Japanese apricot are generally processed into pickles (i.e., “Umeboshi”), fruit liqueurs (i.e., “Ume-shu”), syrups, jams, and other processed foods.

Conventionally, Japanese apricot is divided into two major groups: fruit and flower ornamental cultivars (Horiuchi et al., 1996; Mega et al., 1988). The two groups differ in a variety of morphological traits, tree architecture, petal number, flower color, and fruit size, but clear genetic differences have yet to be reported. In addition, there are some minor groups with characteristic traits. For example, cultivars in the small-fruit group literally bear very small-sized fruits (approximately less than 10 g), and mostly exhibit self-compatibility (Yaegaki et al., 2003). Taiwanese cultivars generally show a very weak chilling requirement for endodor-

Received; May 23, 2018. Accepted; November 26, 2018.

First Published Online in J-STAGE on January 18, 2019.

This work was supported by JSPS KAKENHI Grant Number JP18K14449.

* Corresponding author (E-mail: tishii@kobe-u.ac.jp).

mancy release, which is a suitable survival strategy in subtropical climates (Kitamura et al., 2018). Inter-specific hybrid groups derived from the crosses with apricot or Japanese plum also have very specific morphological and physiological characteristics (Mehlenbacher et al., 1991; Yaegaki et al., 2012; Yoshida and Yamanishi, 1988).

In general, the direction of domestication is influenced by human preference (Zeder et al., 2006). In peach, for example, a distinct genetic differentiation was detected between fruit and ornamental cultivars based on genome-wide SNPs analysis (Akagi et al., 2016). However, there are not enough molecular markers for genetic analysis in Japanese apricot. Previously, microsatellite markers were designed using the peach and cherry nucleotide sequences (Gao et al., 2004), but most of them do not give informative band patterns. Hayashi et al. (2008) examined genetic diversity in Japanese apricot using microsatellite markers based on the peach and apricot genomic sequences, but the marker numbers and their genomic information are limited. Recently, Zhang et al. (2012) reported the genome sequences of eight linkage groups from a Chinese *P. mume* accession.

Here, we designed 201 new microsatellite markers using the *P. mume* reference genome and selected 20 highly polymorphic markers that were more capable of detecting polymorphisms than markers previously developed using peach and apricot genomes. The markers were used to fingerprint 124 *P. mume* accessions (mainly Japanese cultivars), as well as four other *Prunus* species (*P. armeniaca*, *P. salicina*, *P. persica*, and *P. dulcis*). The resulting genotype data were used to evaluate the genetic differentiation of Japanese apricot cultivars.

Materials and Methods

Plant materials

The plant materials used in this study are listed in Tables 1 and S1. A total of 124 *P. mume* accessions

were used, including 46 fruit (F) cultivars, 10 small-fruit (FS) cultivars, 49 ornamental (O) cultivars, five Taiwanese (T) cultivars, 10 putative *P. armeniaca* × *P. mume* (AM) hybrids, and four putative *P. salicina* × *P. mume* (SM) hybrids. In addition, one accession each of apricot (Pa), Japanese plum (Ps), peach (Pp), and almond (Pd) were used as outgroup species. Of these, 61 accessions had multiple entries of different trees, and 79 accessions were the same as previously used by Hayashi et al. (2008) (Table S1). All the plant materials were maintained in the experimental orchards of the Japanese Apricot Laboratory, Wakayama Fruit Tree Experiment Station, Minabe, Wakayama, Japan. Genomic DNA was extracted from leaves using a DNeasy plant mini kit (Qiagen, Hilden, Germany).

Microsatellite primer design

Microsatellite primers were designed using the *P. mume* reference genome (Zhang et al., 2012). The nucleotide sequences of the eight linkage groups (NC_024126.1 for LG1, NC_024127.1 for LG2, NC_024128.1 for LG3, NC_024129.1 for LG4, NC_024130.1 for LG5, NC_024131.1 for LG6, NC_024132.1 for LG7, and NC_024133.1 for LG8) were downloaded from the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov/genome/13911), and surveyed for microsatellite regions using a Tandem repeats finder (Benson, 1999). Microsatellite regions with ~20 GA/CT or AG/TC repeats were selected at each 1 Mb interval (a total of 201 regions). Primer pairs were then designed using Primer 3 (Untergasser et al., 2012) with the following parameters: 17–22 nucleotides long, T_m approximately 60°C, and product size in the range of 100–250 bp.

Microsatellite marker screening

The 201 microsatellite markers (designated as JAM, Table S2) were subjected to the following three screenings for the marker availability: 1) amplification ability

Table 1. Number of the *Prunus* accessions and trees used in this study.

Species	Code	Group description	No. accessions	No. trees ^z
<i>P. mume</i>	F	Cultivars for fruits	46	95
	FS	Cultivars for small fruits (called “Ko-ume”)	10	18
	O	Ornamental cultivars	49	50
	T	Taiwanese cultivars	5	7
	AM	Putative hybrids (<i>P. mume</i> × <i>P. armeniaca</i>)	10	15
	SM	Putative hybrids (<i>P. mume</i> × <i>P. salicina</i>)	4	8
<i>P. armeniaca</i>	Pa	Apricot	1	2
<i>P. salicina</i>	Ps	Japanese plum	1	2
<i>P. persica</i>	Pp	Peach	1	1
<i>P. dulcis</i>	Pd	Almond	1	1
Total			128	199

^z A total of 61 accessions had multiple tree entries.

with a standard cultivar of 'Nanko', 2) a wide range amplification check using eight *Prunus* accessions (three F cultivars and one each of FS, O, T, AM, and Pa), 3) polymorphism examination with 16 *P. mume* cultivars (10 F, three FS, and three O cultivars). For all screenings, PCR was performed in a 20 μ L volume using *Ex Taq* (Takara Bio, Shiga, Japan) according to the manufacturer's instructions. The PCR conditions were as follows: an initial denaturation of 5 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 55°C, 1 min at 72°C, and 7 min at 72°C for final extension. PCR products were separated by 4% polyacrylamide gel electrophoresis and visualized by a silver staining method (Panaud et al., 1996).

Microsatellite marker analysis

Using a selected set of 20 highly polymorphic microsatellite markers, a total of 128 *Prunus* accessions (199 tree entries) were genotyped. In addition, 11 microsatellite markers previously reported by Hayashi et al. (2008) were also employed. To precisely determine the lengths of amplified fragments, PCR was performed using the post-labeling method described by Schuelke (2000) with minor modifications. To generate fluorescent PCR products, the U-19 universal primer (5'-GTTTCCCCAGTCACGACGT-3') was labeled with four kinds of fluorescent molecules (6-FAM, VIC, NED, and PET), as well as with 2-bp barcodes at the 3' ends (TG for 6-FAM-labeled U-19, AC for VIC-labeled U-19, CA for NED-labeled U-19, and GT for PET-labeled U-19). Non-labeled forward primers, with 7-bp 5' pig-tails (5'-GTTTCTT-3'), and U-19-fused reverse primers were also synthesized and used to perform multiplex PCR. The 20 μ L PCR mixtures were prepared using a Type-it Microsatellite PCR kit (Qiagen), with 50–100 ng template DNA, 0.2 μ M each labeled U-19 primer, 0.2 μ M each pig-tailed forward primer, and 0.04 μ M each U-19-fused reverse primer. The PCR conditions were as follows: initial denaturation of 5 min at 95°C; followed by 10 cycles of 30 s at 95°C, 90 s at 60°C with a decrease of 0.5°C in each cycle, and 30 s at 72°C; followed by 22 cycles of 30 s at 95°C, 90 s at 55°C, and 30 s at 72°C, and a final extension step of 30 min at 60°C. Finally, the fragments were analyzed using an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) using LIZ-600 size standard (Applied Biosystems). Alleles were defined on the basis of fragment size (in nucleotides) and scored using Genemapper 4.1 (Applied Biosystems).

Genetic diversity and population structure analysis

Statistical analyses were performed after excluding two triploid cultivars ('Horyukaku' and 'Takasago') and the four accessions of other *Prunus* species (*P. armeniaca* 'Heiwa' was employed as an outgroup only for phylogenetic tree construction). Different genotypes found in the same accessions were independ-

ently examined in the analyses. For accessions that had identical genotypes at all loci, only one representative accession was used in the analyses, except for phylogenetic tree construction. Marker scores with null or more than three alleles were treated as missing data. However, for only the phylogenetic analysis, the score sets of loci with missing data were excluded. The number of alleles, observed heterozygosity (H_o), expected heterozygosity (H_e), number of effective alleles (N_e), fixation index (F_{IS}), Nei's genetic distance (Nei et al., 1983), and G'_{st} (Hedrick, 2005) were calculated using GenAlEx v. 6.502 (Peakall and Smouse, 2012). Here, G'_{st} is defined as a standardized genetic differentiation measure of G_{st} (Nei, 1972) for highly polymorphic markers (Hedrick, 2005). Polymorphism information content (PIC) was calculated using POLYSAT (Clark and Jasieniuk, 2011) in R (R Development Core Team, 2008). Power of discrimination (PD) (Kloosterman et al., 1993) was computed manually using the following formula: $PD = 1 - \sum g_i^2$, where g_i is the frequency of the i th genotype. Meanwhile, within-group genetic variation was assessed using the gene diversity value (D) (Nei, 1987). Statistical significance for F_{IS} and G'_{st} was tested using GENEPOP 4.2 (Raymond and Rousset, 1995; Rousset, 2008) and GenAlEx v. 6.502 (Peakall and Smouse, 2012), respectively. The levels of linkage disequilibrium (LD) were calculated as the squared allele-frequency correlation (r^2) values using MIDAS (Gaunt et al., 2006). The r^2 values were calculated between the most frequent alleles at each two markers, and among all combinations of alleles at each two markers and summed up with weight of allele frequencies, according to Iwata et al. (2013).

The population structure of the *P. mume* accessions was analyzed using three methods, namely principal coordinate analysis using GenAlEx v. 6.502 (Peakall and Smouse, 2012), neighbor-joining (NJ) tree construction using Poptree 2 with 1000 bootstrap replicates (Takezaki et al., 2010), and individual-based Bayesian clustering using STRUCTURE 2.3.4 with Markov chain Monte Carlo simulations (Pritchard et al., 2000). The NJ tree was visualized using MEGA6 (Tamura et al., 2013). In the STRUCTURE analysis, the AM and SM hybrids were excluded in order to simplify the simulations. In addition, the number of genetic clusters (K) was set from 1–15, and each analysis was allowed to run for 200,000 simulations after a burn-in period of 200,000 iterations. Ten runs were performed for each K value, and the optimal value was determined using $L(K)$, $|L''(K)|$, and ΔK (Evanno et al., 2005), which were calculated using Structure Harvester (Earl and vonHoldt, 2012). A representative bar plot for each K was selected based on $\text{LnP}(D)$ values generated by STRUCTURE software, and visualized using Structure Plot v2.0 (Ramasamy et al., 2014).

Results

Selection of highly polymorphic microsatellite markers

Among the 201 microsatellite markers designed, 188 (93.5%) were successfully amplified using the template DNA of ‘Nanko’ (Table S2). Of these, 128 markers were randomly selected for second amplification screening. As a result, 59 markers gave clear bands for all eight *Prunus* accessions. They were preliminarily checked for the level of polymorphism using 16 *P. mume* cultivars. Their PIC values were compared and a total of 20 highly polymorphic markers were selected: at least two markers were selected from those giving the highest PIC values in each linkage group. They were renamed as PMKS markers (Table S2).

Allele combinations at 20 polymorphic microsatellite loci among Prunus accessions

In order to examine the allelic diversity, a total of 128 *Prunus* accessions (199 tree entries) were genotyped with a set of 20 highly polymorphic markers (Table S3). Identical genotypes for all the loci were observed among the trees in 56 out of 61 accessions having multiple entries, whereas different allele combinations were observed in the other five accessions: ‘Fudono’ (F4), ‘Naniwa’ (F24), ‘PM1-1’ (SM1), ‘Tsuyakane’ (SM4), and ‘Heiwa’ (Pa) (Table S3).

Among the 124 *P. mume* accessions, a total of 107 allele combinations were observed, with 92 of the combinations being unique to single accessions and 15 being shared by multiple accessions. Thirteen of the redundant genotypes were shared by two accessions each, and genotypes of ‘B’ and ‘J’ were observed for seven and three accessions, respectively (Table S3). In addition to the triploid cultivars (‘Horyukaku’ and ‘Takasago’), several other accessions also gave more than three alleles at certain loci.

Efficacy of microsatellite markers in P. mume

The efficacy (i.e., polymorphism detection ability) of the 20 microsatellite markers was assessed with the allele combinations found in *P. mume* (Table 2). The PIC values ranged from 0.63 to 0.90 (mean = 0.79) and the PD values ranged from 0.84 to 0.97 (mean = 0.93). Using the same *P. mume* accessions, 11 microsatellite markers previously reported by Hayashi et al. (2008) were also evaluated (Tables 2 and S2). The average values of all the polymorphism indices (number of alleles, Ho, He, PIC, and PD) for the newly developed marker set were greater than those for previously reported, thereby reflecting the powerful detection ability of the new marker set. For example, ‘Hakuo’ (FS2), ‘Koshu Saisho’ (FS4), and ‘Purple Queen’ (FS8) shared a single genotype when using the previously developed set, whereas the newly developed markers were able to distinguish ‘Hakuo’ from the other two. Similarly, the new marker set was also able to distinguish ‘Ikuyonezame’

(O14) and ‘Kinko’ (O19). The F_{IS} values ranged from 0.01 to 0.46 (mean = 0.13) and 16 of 20 loci showed statistical significance at the 5% level (Table 2). The levels of LD were low among all combinations of 20 loci (Table S4).

Genetic variation within cultivar groups

Genetic diversity indices were calculated to evaluate variation within the six cultivar groups (F, FS, O, T, AM, and SM; Table 3), using the same set of allele combinations without duplicates. Although the H_o and D values of the six *P. mume* groups were similar, a relatively high N_e values were observed for the F and O groups (Table 3).

Genetic differentiation among cultivar groups

Genetic differentiation among the *P. mume* cultivar groups was first evaluated using principal coordinate analysis with individual genotype data. The first principal coordinate axis (PC1) seemed to separate the F and O cultivars (Fig. 1), whereas the second axis (PC2) distinguished the T cultivars from the others, and also seemed to explain within-group variation. However, even though the analysis revealed genetic differentiation among the *P. mume* accessions, the percentages of variation explained by PC1 and PC2 were relatively low (9.0% and 6.7%, respectively).

We next calculated pairwise Nei’s genetic distance and G'_{st} values for the six cultivar groups (Table 4). Significant ($P < 0.01$) G'_{st} values were obtained for all pairs. Among the six groups, the T cultivars showed relatively high genetic distance values (1.053–2.042), followed by the two putative hybrid groups (SM and AM). However, in contrast to the results of the principal coordinate analysis, the genetic distance between the F and O cultivars was quite low (0.238). A similar low value (0.275) was also observed between the F and FS groups, although they have major phenotypic differences in terms of fruit size and harvesting time.

Phylogenetic and STRUCTURE analyses

Phylogenetic analysis was performed with all the *P. mume* accessions with no missing data and *P. armeniaca* ‘Heiwa’ as an outgroup (Fig. 2). Three of the clades were strongly supported (bootstrap value > 70). One, which was clearly separated from the others, included the AM hybrids and the *P. armeniaca* ‘Heiwa’, whereas the two others contained T (T clade) and O cultivars, including ‘China mume’ (O-1 clade). The rest of the accessions belonged to two wider clusters that mainly included F and O cultivars (Fig. 2).

STRUCTURE analysis was performed using 94 genotypes from the four *P. mume* cultivar groups (F, FS, O, and T). The highest ΔK value was obtained at $K = 2$, followed by $K = 7$ (Fig. 3A). In the bar plot for $K = 2$, all T cultivars showed a single orange-colored cluster (Fig. 3B), whereas the other three groups gave a mix-

Table 2. Polymorphism indices calculated for the present (A) and the previous marker sets (B).

(A)	Name	Linkage group ^z	N ^y	No. alleles	Ho ^x	He ^x	F _{IS} ^x	PIC ^x	PD ^x	Reference ^w
	PMKS15	LG1	105	17	0.76	0.86	0.11* ^v	0.84	0.96	1, 2
	PMKS21	LG1	105	20	0.84	0.87	0.04	0.86	0.96	1, 2
	PMKS49	LG2	105	15	0.70	0.84	0.16**	0.82	0.94	1, 2
	PMKS59	LG2	105	21	0.80	0.81	0.01	0.80	0.95	1, 2
	PMKS68	LG3	105	14	0.80	0.85	0.05*	0.83	0.95	1, 2
	PMKS75	LG3	105	23	0.89	0.91	0.02**	0.90	0.97	1, 2
	PMKS99	LG4	105	12	0.70	0.85	0.18**	0.83	0.94	1, 2
	PMKS113	LG4	105	12	0.76	0.81	0.06*	0.79	0.93	1, 2
	PMKS121	LG5	105	9	0.67	0.78	0.15**	0.76	0.92	1, 2
	PMKS131	LG5	98	15	0.46	0.85	0.46**	0.84	0.93	1, 2
	PMKS133	LG5	105	18	0.76	0.90	0.15**	0.89	0.96	1, 2
	PMKS149	LG6	105	17	0.70	0.75	0.08	0.73	0.91	1, 2
	PMKS164	LG6	105	8	0.60	0.72	0.17*	0.68	0.88	1, 2
	PMKS175	LG7	89	15	0.68	0.79	0.13**	0.77	0.93	1, 2
	PMKS179	LG7	105	14	0.74	0.78	0.05**	0.77	0.93	1, 2
	PMKS187	LG8	101	12	0.56	0.76	0.26**	0.74	0.88	1, 2
	PMKS191	LG8	105	14	0.64	0.74	0.14**	0.72	0.88	1, 2
	PMKS193	LG8	104	15	0.78	0.85	0.09	0.84	0.96	1, 2
	PMKS197	LG8	105	12	0.68	0.78	0.13*	0.76	0.92	1, 2
	PMKS201	LG8	105	10	0.56	0.65	0.14**	0.63	0.84	1, 2
	Mean			14.7	0.70	0.81	0.13	0.79	0.93	
(B)	Name	Linkage group ^z	N ^y	No. alleles	Ho ^x	He ^x	F _{IS} ^x	PIC ^x	PD ^x	Reference ^w
	UDP96-001	LG1	105	7	0.69	0.71	0.04	0.66	0.87	3, 7
	pchgms3	LG2	105	13	0.71	0.76	0.06	0.73	0.91	4, 7
	MA007a	LG5	105	14	0.80	0.84	0.05**	0.82	0.94	5, 7
	MA017a	Unknown	104	11	0.44	0.79	0.44**	0.77	0.90	5, 7
	MA040a	LG1	105	6	0.40	0.50	0.20	0.45	0.69	5, 7
	M6a	LG6	103	12	0.39	0.50	0.22**	0.48	0.67	5, 7
	M7a	LG6	105	10	0.38	0.71	0.47**	0.67	0.84	5, 7
	PaCITA4	LG4	105	17	0.87	0.89	0.02**	0.88	0.96	6, 7
	PaCITA7	LG2	104	19	0.88	0.91	0.03	0.91	0.98	6, 7
	PaCITA19	LG5	105	11	0.65	0.78	0.17**	0.75	0.92	6, 7
	PaCITA21	LG7	103	12	0.37	0.65	0.43**	0.62	0.79	6, 7
	Mean			12.0	0.60	0.73	0.19	0.70	0.86	

^z After *P. mume* reference genome sequences by Zhang et al. (2012).

^y Number of genotypes examined. Genotypes giving null or more than three alleles were excluded.

^x Ho: observed heterozygosity, He: expected heterozygosity, F_{IS}: fixation index, PIC: polymorphism information content, PD: power of discrimination.

^w 1: The present study, 2: Ishio et al., patent pending, 3: Testolin et al. (2000), 4: Sosinski et al. (2000), 5: Yamamoto et al. (2002), 6: Lopes et al. (2002), 7: Hayashi et al. (2008).

^v * and **: significant at 5% and 1% level, respectively.

ture of two clusters. Yellow and orange clusters seemed to be dominant in the F and O cultivar groups, respectively. On the other hand, for $K = 7$, T cultivars could be clearly separated, and most FS cultivars showed a single red cluster. A purple cluster seemed to be dominant in the F group. O cultivars consisted of many admixture types, but cultivars belonging to the O-1 clade in the phylogenetic tree shared the same navy-colored cluster.

Discussion

A new marker set of highly polymorphic microsatellites

Studies on genetic diversity and clone identification in *P. mume* have been relatively limited owing to a lack of effective molecular tools. Therefore, we designed 201 new microsatellite markers using the *P. mume* reference genome reported by Zhang et al. (2012), and selected 20 highly polymorphic markers for further analysis. As shown in Tables 2 and S2, the present marker set exhibited greater polymorphism detection

ability than those previously developed using the peach and apricot genomes (Hayashi et al., 2008) and was successfully used to fingerprint most of the *Prunus* cultivars (128 accessions, 199 trees). However, 15 redun-

Table 3. Average genetic diversity indices for six cultivar groups in *P. mume*.

Group ^z	N ^y	No. alleles	Ho ^x	D ^x	Ne ^x
F	40	8.8	0.68	0.74	4.25
FS	9	5.3	0.71	0.67	3.49
O	40	9.0	0.71	0.78	4.91
T	5	4.5	0.64	0.65	3.35
AM	8	5.6	0.79	0.71	3.75
SM	5	3.6	0.70	0.61	3.12

^z Group description is shown in Table 1.

^y Number of genotypes excluding the duplicates within each group.

^x Ho: observed heterozygosity, D: gene diversity, Ne: Number of effective alleles.

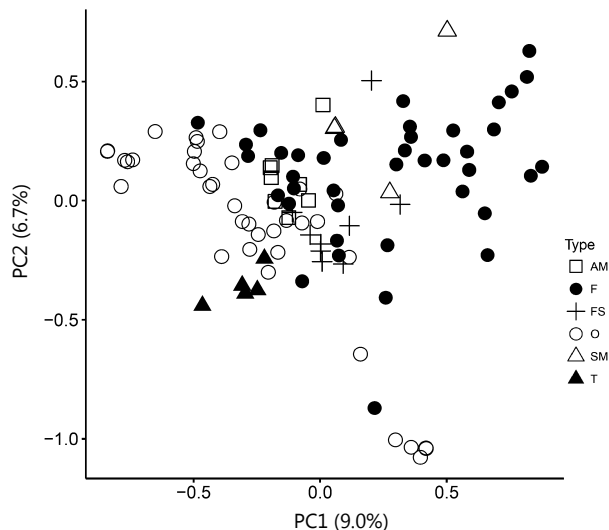


Fig. 1. Principal coordinate analysis of the six Japanese apricot cultivar groups. Values in parentheses indicate the proportion of the total variance explained by each principal coordinate. The abbreviations AM, F, FS, O, SM, and T indicate *P. armeniaca* × *P. mume* hybrids, fruit, small-fruit, ornamental cultivars, *P. salisina* × *P. mume* hybrids, and Taiwanese cultivars, respectively.

dant genotypes (designated as A to O in Table S3) were shared by two or more accessions. Hayashi et al. (2008) also reported that identical genotypes were detected among cultivars having very similar morphological characteristics, namely, among the three cultivars of ‘Shirokaga’ (F36), ‘Gojiro’ (F7), and ‘Gyokuei’ (F8), two of ‘Kodama’ (F20) and ‘Kinyuji’ (F19), and two of ‘Hanakami’ (F10) and ‘Rinshu (Nara)’ (F32). Since these accessions may be recent somaclonal variants, resulting from bud sports with slight mutations, it is quite difficult to distinguish them using microsatellite markers. Therefore, next generation sequencing will likely be needed to detect cultivar-specific SNPs.

On the other hand, different allele combinations were observed among tree entries in the five accessions: ‘Fudono’ (F4), ‘Naniwa’ (F24), ‘PM1-1’ (SM1), ‘Tsuyakane’ (SM4), and ‘Heiwa’ (Pa) (Table S3). Two different genotypes between trees were observed in ‘Fudono’ and ‘Naniwa’. Since they are old local varieties, some chance seedlings showing similar phenotypes may have been cultivated as the same varieties. For the modern varieties ‘PM1-1’ and ‘Tsuyakane’, length mutations were observed between two tree entries only at the PMKS113 locus, suggesting that they were generated by recent somaclonal variation. Two ‘Heiwa’ trees were identical at 15 of the 20 loci. ‘Heiwa’ is quite an old variety. Probably, the difference was also caused by somaclonal variation.

In recent years, Japanese fruit crops (strawberry, sweet cherry, and grapevine) have been taken overseas and cultivated. To control such dissemination, domestic fruit cultivars should be protected by international patent or registration with fingerprinting data. The highly polymorphic marker set developed here may be suitable for identifying cultivars of Japanese apricot, and moreover, 201 markers designed to cover whole genome linkage groups in *P. mume* (Table S2) can be useful for linkage and QTL analyses to improve Japanese apricot.

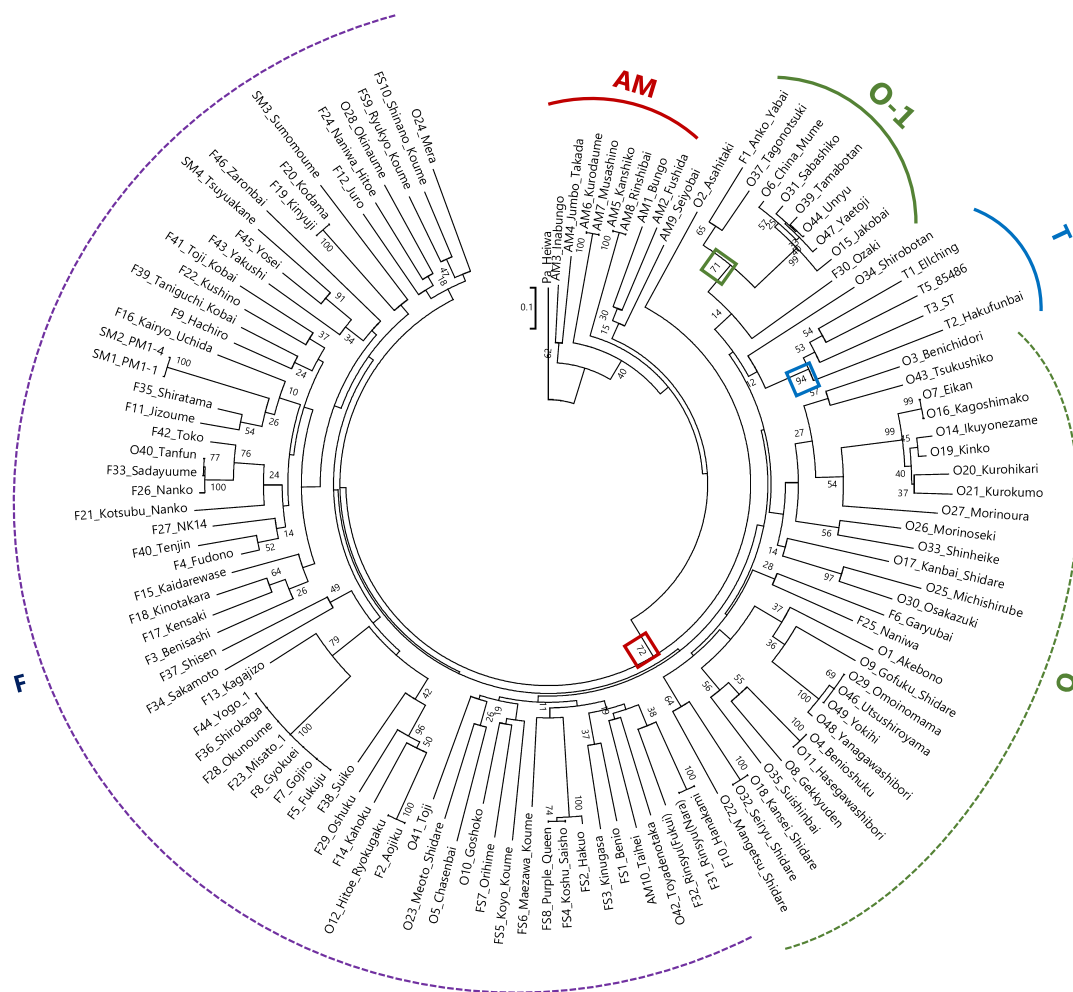
Core collections have been reported for many crops, such as rice (Ebana et al., 2008), soybean (Kaga et al., 2012), and strawberry (Wada et al., 2017). They allow us to maintain a minimum number of genetic resources with maximum genetic variation. The present marker

Table 4. Pairwise genetic differentiation measure, G'_{st} (below diagonal) and Nei’s genetic distance value (above diagonal) between six cultivar groups.

Group ^z	F	FS	O	T	AM	SM
F		0.275	0.238	1.336	0.624	0.487
FS	0.193** ^y		0.265	1.111	0.583	0.872
O	0.201**	0.184**		1.053	0.568	0.860
T	0.730**	0.642**	0.631**		1.688	2.042
AM	0.438**	0.388**	0.393**	0.800**		1.022
SM	0.334**	0.546**	0.553**	0.859**	0.606**	

^z Group description is shown in Table 1.

^y **: Significant at 1% level.



set can be utilized to develop a core collection of Japanese apricot to enhance efficient breeding.

Among the *P. mume* accessions, the average observed heterozygosity of 20 PMKS markers was 0.70, which was lower than that of expected heterozygosity (0.81, Table 2), and most of the F_{IS} values showed significance at the 5% level. These results are due to the fact that the present group of Japanese accessions is not a natural population with random mating. Probably, Japanese accessions went through strong artificial selection, introgression and specific breeding. The levels of LD were low (Table S4), suggesting no association of alleles at 20 loci.

STRUCTURE analysis).

Among six cultivar groups, putative Taiwanese (T) cultivars were clearly distinguished from others in both phylogenetic and STRUCTURE analyses (Figs. 2 and 3B). The results are consistent with those of Hayashi et al. (2008) using SSRs derived from peach and apricot, and Shimada et al. (1994) using RAPD markers. The T cultivars are uniquely characterized by their weak bud dormancy, which enables them to adapt in subtropical climates (Kitamura et al., 2018). The AM cultivars except ‘Taihei’ (AM10) were clearly distinguished in the phylogenetic analysis (Fig. 2). They were derived from interspecific hybrids between *P. mume* and *P. armeniaca* (Mehlenbacher et al., 1991; Yoshida and Yamanashi, 1988). Although ‘Taihei’ (AM10) showed evidence for inheritance of *P. armeniaca* characteristics (Hayashi, 2009; Hayashi et al., 2008), it was included in the huge cluster of Japanese apricot. Probably, ‘Taihei’ went through a subsequent cross with Japanese apricot cultivars. Other F, FS, and O cultivars were not clearly divided into the respective cultivar

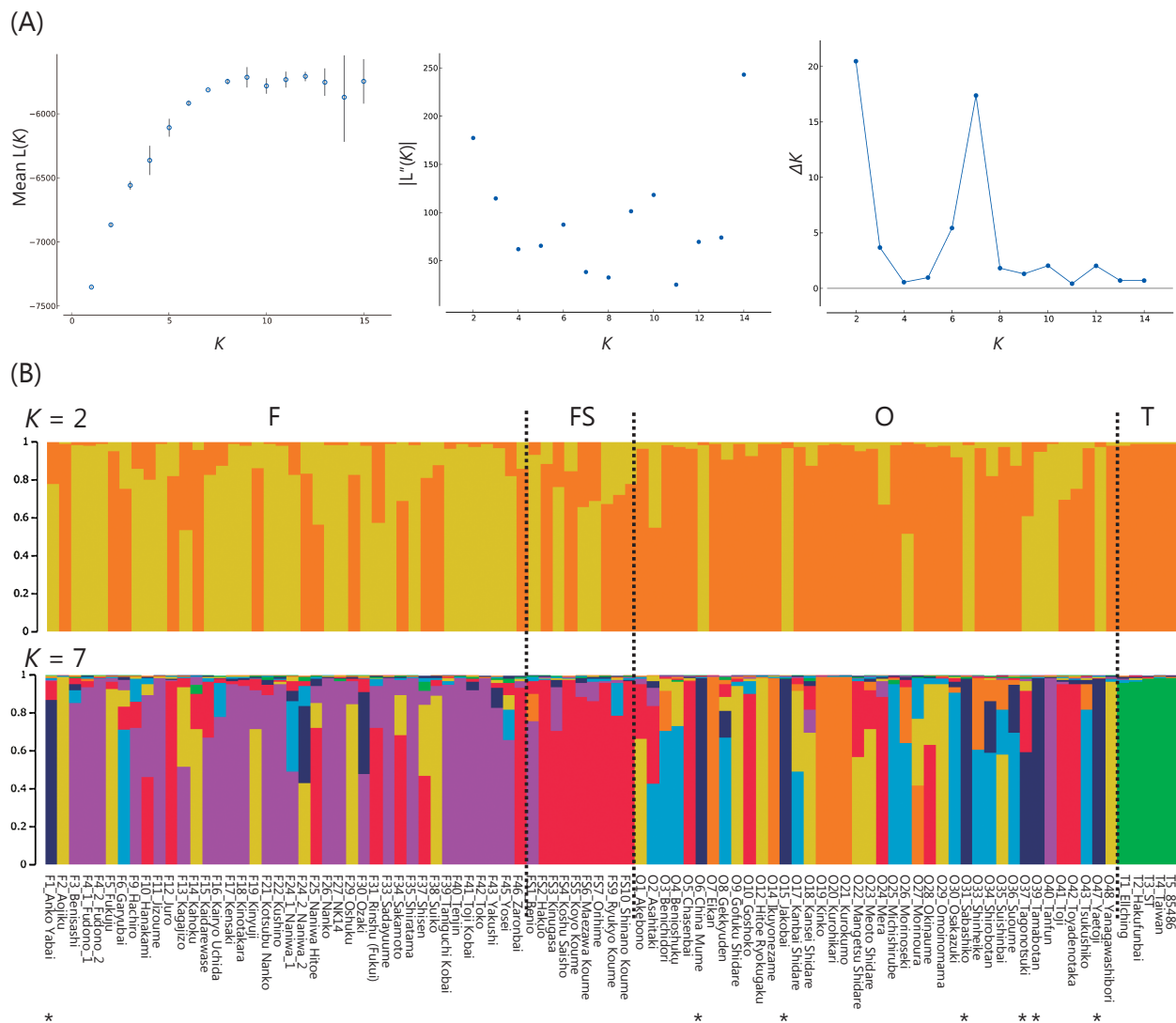


Fig. 3. STRUCTURE analysis of *P. mume* cultivars. Fruit (F), small-fruit (FS), flower ornamental (O), and Taiwanese (T) cultivars were employed for the simulations. (A) Values of mean $L(K)$, $|L''(K)|$, and ΔK . The highest value was obtained at $K = 2$, implying the existence of two clusters. The ΔK value was also high at $K = 7$, indicating the existence of 7 clusters. (B) Bar plots for each value of K , at $K = 2$ and 7. Representative bar plots were selected based on $\text{LnP}(D)$ values generated by STRUCTURE software. Cultivars belonging to the O-1 clade (Fig. 1) are indicated by asterisks.

groups. They formed a complex genetic structure in the Japanese population.

In this study, the genetic differentiation among Japanese apricot was examined by three analyses. However, we do not know the exact sequence of the differentiation process because Japanese apricot is thought to have originated in China and many cultivars may have been occasionally introduced to Japan. Many factors, such as human preference, geographical separation, introgression, and local breeding, may be involved to form the complex genetic structure in Japanese apricot. More comprehensive analytical methods (e.g., genome-wide SNP survey with wider genetic resources including Chinese cultivars and wild relatives) will shed light on the details of this species' evolution, domestication, and improvement history.

Acknowledgements

We are deeply grateful to the researchers at the Japanese Apricot Laboratory, Wakayama Fruit Tree Experiment Station, for maintaining the Japanese apricot cultivars.

Literature Cited

- Akagi, T., T. Hanada, H. Yaegaki, T. M. Gradziel and R. Tao. 2016. Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication. *DNA Res.* 23: 271–282.
- Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27: 573–580.
- Clark, L. and M. Jasieniuk. 2011. POLYSAT: an R package for polyploid microsatellite analysis. *Mol. Ecol. Resour.* 11: 562–566.

- 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. Resour.* 4: 359–361.
- Ebana, K., Y. Kojima, S. Fukuoka, T. Nagamine and M. Kawase. 2008. Development of mini core collection of Japanese rice landrace. *Breed. Sci.* 58: 281–291.
- Evanno, G., S. Regnaut and J. Coudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611–2620.
- Gao, Z. H., Z. J. Shen, Z. H. Han, J. G. Fang, Y. M. Zhang and Z. Zhang. 2004. Microsatellite markers and genetic diversity in Japanese apricot (*Prunus mume*). *HortScience* 39: 1571–1574.
- Gaunt, T. R., S. Rodriguez, C. Zapata and I. N. M. Day. 2006. MIDAS: software for analysis and visualization of interallelic disequilibrium between multiallelic markers. *BMC Bioinformatics* 7: 227. DOI: 10.1186/1471-2105-7-227. <<https://bmcbioinformatics.biomedcentral.com/>>.
- Hayashi, K. 2009. Genetic diversity analysis in Japanese apricot using DNA markers (In Japanese). Ph.D. Thesis. Univ. of Tsukuba, Tsukuba.
- Hayashi, K., K. Shimazu, H. Yaegaki, M. Yamaguchi, H. Iketani and T. Yamamoto. 2008. Genetic diversity in fruiting and flower-ornamental Japanese apricot (*Prunus mume*) germplasms assessed by SSR markers. *Breed. Sci.* 58: 401–410.
- Hedrick, P. W. 2005. A Standardized genetic differentiation measure. *Evolution* 59: 1633–1638.
- Horiuchi, S., M. Yoshida, H. Kariya, T. Nakamura, H. Hasebe, T. Suzuki and T. Sakitani. 1996. *Nihonnoume Sekainoume* (In Japanese). Yokendo, Tokyo.
- Iwata, H., T. Hayashi, S. Terakami, N. Takada, Y. Sawamura and T. Yamamoto. 2013. Potential assessment of genome-wide association study and genomic selection in Japanese pear *Pyrus pyrifolia*. *Breed. Sci.* 63: 125–140.
- Kaga, A., T. Shimizu, S. Watanabe, Y. Tsubokura, Y. Katayose, K. Harada, D. A. Vaughan and N. Tomooka. 2012. Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. *Breed. Sci.* 61: 566–592.
- Kitamura, Y., T. Habu, H. Yamane, S. Nishiyama, K. Kajita, T. Sobue, T. Kawai, K. Numaguchi, T. Nakazaki, A. Kitajima and R. Tao. 2018. Identification of QTLs controlling chilling and heat requirements for dormancy release and bud break in Japanese apricot (*Prunus mume*). *Tree Genet. Genomes* 14: 33.
- Kloosterman, A. D., B. Budowle and P. Daselaar. 1993. PCR amplification and detection of the human D1S80 VNTR locus—Amplification conditions, population-genetics and application in forensic analysis. *Int. J. Leg. Med.* 105: 257–264.
- Lopes, M. S., K. M. Sefc, M. Laimer and A. da Camara Machado. 2002. Identification of microsatellite loci in apricot. *Mol. Ecol. Notes* 2: 24–26.
- Ministry of Agriculture, Forestry and Fisheries (MAFF). 2016. Fruits and Nuts. In: The 90th Statistical yearbook of Ministry of Agriculture, Forestry and Fisheries. <www.maff.go.jp/e/data/stat/90th/index.html>.
- Mega, K., E. Tomita, S. Kitamura, S. Saito and S. Mizukami. 1988. Ume. p. 289–300 (In Japanese). In: T. Aoba et al. (eds.). The grand dictionary of horticulture. Shogakukan, Tokyo.
- Mehlenbacher, S. A., V. Cocin and L. F. Hough. 1991. Apricots (*Prunus*). p. 65–107. In: J. N. Moore and J. R. Ballington Jr. (eds.). Genetic resources of temperate fruit and nut crops. The International Society for Horticultural Science, Wageningen.
- Nei, M. 1972. Genetic distance between populations. *Am. Nat.* 106: 283–292.
- Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.
- Nei, M., F. Tajima and Y. Tatenno. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol.* 19: 153–170.
- Panaud, O., X. Chen and S. R. McCouch. 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol. Gen. Genet.* 252: 597–607.
- Peakall, R. and P. E. Smouse. 2012. GenA1Ex 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28: 2537–2539.
- Pritchard, J. K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- R Development Core Team. 2008. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ramasamy, R. K., S. Ramasamy, B. B. Bindroo and V. G. Neik. 2014. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springer-Plus* 3: 431. DOI: 10.1186/2193-1801-3-431.
- Raymond, M. and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity* 86: 248–249.
- Rousset, F. 2008. Genepop'007: a complete re-implementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8: 103–106.
- Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18: 233–234.
- Shimada, T., T. Haji, M. Yamaguchi, T. Takeda, K. Nomura and M. Yoshida. 1994. Classification of Mume (*Prunus mume* Sieb. et Zucc.) by RAPD assay. *J. Jpn. Soc. Hortic. Sci.* 63: 543–551.
- Sosinski, B., M. Gannavaraqu, L. D. Hanger, L. E. Beck, G. J. King, C. D. Ryder, S. Rajapakse, W. V. Baird, R. E. Ballard and A. G. Abbott. 2000. Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. *Theor. Appl. Genet.* 101: 421–428.
- Takezaki, N., M. Nei and K. Tamura. 2010. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol. Biol. Evol.* 27: 747–752.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30: 2725–2729.
- Testolin, R., T. Marrazzo, G. Cipriani, R. Quarta, I. Verde, M. T. Dettori, M. Pancaldi and S. Sansavini. 2000. Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genomic origin of cultivars. *Genome* 43: 512–520.
- Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm and S. G. Rozen. 2012. Primer 3—new capabilities and interfaces. *Nucleic Acids Res.* 40: e115. DOI: 10.1093/nar/gks596.
- Wada, T., Y. Noguchi, S. Isobe, M. Kuniyoshi, T. Sueyoshi and K. Shimomura. 2017. Development of a core collection of strawberry cultivars based on SSR and CAPS marker polymorphisms. *Hort. J.* 86: 365–378.
- Yaegaki, H., T. Haji, Y. Nakamura, M. Miyake, K. Nishimura, H. Kyotani and M. Yamaguchi. 2003. Varietal and yearly varia-

- tions in fruit and endocarp weights and their ratio in Japanese apricot. J. Japan. Soc. Hort. Sci. 72: 546–550 (In Japanese with English abstract).
- Yaegaki, H., M. Yamaguchi, T. Haji, Y. Suesada, M. Miyake, T. Kihara, K. Suzuki and M. Uchida. 2012. New Japanese apricot cultivar ‘Tsuyukane’. Bull. NARO Inst. Fruit Tree Sci. 12: 1–6 (In Japanese with English abstract).
- Yamamoto, T., K. Mochida, T. Imai, Y. Z. Shi, I. Ogiwara and T. Hayashi. 2002. Microsatellite markers in peach [*Prunus persica* (L.) Batsch] derived from an enriched genomic and cDNA libraries. Mol. Ecol. Notes 2: 298–301.
- Yoshida, M. and H. Yamanishi. 1988. Apricot cultivars in Japan. Acta Hortic. 209: 69–81.
- Zeder, N. A., E. Emshwiller, B. D. Smith and D. G. Bradley. 2006. Documenting domestication: the intersection of genetics and archaeology. Trends Genet. 22: 139–155.
- Zhang, Q., W. Chen, L. Sun, F. Zhao, B. Huang, W. Yang, Y. Tao, J. Wang, Z. Yuan, G. Fan, Z. Xing, C. Han, H. Pan, X. Zhong, W. Shi, X. Liang, D. Du, F. Sun, Z. Xu, R. Hao, T. Lv, Y. Lv, Z. Zheng, M. Sun, L. Luo, M. Cai, Y. Gao, J. Wang, Y. Yin, X. Xu, T. Cheng and J. Wang. 2012. The genome of *Prunus mume*. Nat. Commun. 3: 1318. DOI: 10.1038/ncomms2290.