Interspecific introgression and natural selection in the evolution of Japanese apricot (*Prunus mume*)

Koji Numaguchi^{1,2*}, Takashi Akagi^{3*}, Yuto Kitamura², Ryo Ishikawa¹ and Takashige Ishii¹

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

²Japanese Apricot Laboratory, Wakayama Fruit Tree Experiment Station, Minabe,

Wakayama 645-0021, Japan; ³Graduate School of Environmental and Life Science,

Okayama University, Okayama 700-8530, Japan

*Authors for correspondence:

Koji Numaguchi

Tel: +81 739 74 3780

Email: numaguchi_k0001@pref.wakayama.lg.jp

Takashi Akagi

Tel: +81 86 251 8337

Email: takashia@okayama-u.ac.jp

Total word count (excluding summary, references and legends):	5581	No. of figures:	7 (Figs 1–7 in color)
Summary:	198	No. of Tables:	0
Introduction:	784	No. of Supporting Information Files:	2 (Figs S1–S9; Tables S1–S7)
Materials and Methods:	1650		
Results:	1853		
Discussion:	1228		
Acknowledgements:	66		

Summary

- Domestication and population differentiation in crops involve considerable phenotypic changes. The logs of these evolutionary paths, including natural/artificial selection, can be found in the genomes of the current populations. However, these profiles have been little studied in tree crops, which have specific characters, such as long generation time and clonal propagation, maintaining high levels of heterozygosity.
- We conducted exon-targeted resequencing of 129 genomes in the genus *Prunus*, mainly Japanese apricot (*Prunus mume*), and apricot (*P. armeniaca*), plum (*P. salicina*), and peach (*P. persica*). Based on their genome-wide single nucleotide polymorphisms merged with published resequencing data of 79 Chinese *P. mume* cultivars, we inferred complete and ongoing population differentiation in *P. mume*.
- Sliding window characterization of the indexes for genetic differentiation identified interspecific fragment introgressions between *P. mume* and related species (plum and apricot). These regions often exhibited strong selective sweeps formed in the paths of establishment or formation of substructures of *P. mume*, suggesting that *P. mume* has frequently imported advantageous genes from other species in the subgenus *Prunus* as adaptive evolution.
- These findings shed light on the complicated nature of adaptive evolution in a tree crop that has undergone interspecific exchange of genome fragments with natural/artificial selection.

Key words: fruit tree, introgression, population structure, *Prunus mume*, selective sweep, targeted resequencing.

Introduction

Domestication and population differentiation often involve considerable phenotypic changes in crops (Diamond, 2002; Purugganan & Fuller, 2009; Zeder, 2015). Mainly natural mutations are thought to drive these changes, while occasional interspecific introgression can also potentially contribute (Baack & Rieseberg, 2007; Harrison & Larson, 2014; Gaut *et al.*, 2015; Suarez-Gonzalez *et al.*, 2018). For instance, domesticated apple (*Malus domestica*), which originated from a wild apple species distributed in Central Asia (*M. sieversii*), has recently experienced additional genomic introgression from another wild species (*M. sylvestris*) (Cornille *et al.*, 2012). Citrus and olive are also suggested to have complicated evolutionary pathways involving interspecific introgression from related or ancestral wild species (Wu *et al.*, 2014, 2018; Diez *et al.*, 2015). Hexaploid bread wheat (*Triticum aestivum*) is a notable example of herbaceous crops with a drastic domestication process. This species was developed from a dynamic hybridization between tetraploid emmer wheat (*T. turgidum*) and diploid Tausch's goatgrass (*Aegilops tauschii*) and subsequent introgression from other species promoted cultivar differentiation (Molnár-Láng *et al.*, 2015; He *et al.*, 2019). Owing to the evolutionary importance of interspecific introgression, much previous research on a variety of species has inferred the presence of interspecific introgression in current populations. Notwithstanding, few studies have further estimated the genome-wide distribution of introgressed fragments and their importance for domestication/population differentiation events.

Introduced mutations favorable to environmental adaptation or human preference might be subjected to natural or artificial selection. When a particular locus

experiences a strong selection pressure, the genetic diversity of adjacent genomic regions is reduced as well as the targeted locus itself, which is known as a “selective sweep” (Stephan, 2019). Therefore, we can estimate the genetic factors playing important roles in the formation of current populations by characterizing genome-wide selective sweep profiles (Clark *et al.*, 2004; Sabeti *et al.*, 2007; Kosova *et al.*, 2010; Ishii *et al.*, 2013; Akagi *et al.*, 2016; Lee *et al.*, 2016; Pankin *et al.*, 2018; Nadachowska-Brzyska *et al.*, 2019). Selective sweep profiles have been well studied especially in annual crops such as rice (*Oryza sativa*). In annual crops, where selected alleles are thought to be fixed in a homozygous state, patterns of selective sweeps have been identified mainly using site frequency spectrum (SFS)-based methods using the reduction in genetic diversity as the index. Conversely, perennial crops (or tree crops) have more complicated genomic/genetic conditions, mainly due to vegetative propagation, frequent outcrossing, and long generation time. Therefore, a selected allele in perennial crops is expected to be maintained in a heterozygous manner, which would be quite similar to animal (including human) genomes, requiring haplotype-based detection of selective sweeps (Voight *et al.*, 2006; Sabeti *et al.*, 2007).

The genus *Prunus* includes a wide variety of major tree crops consumed worldwide, such as peach (*P. persica*), sweet cherry (*P. avium*), plum (*P. salicina*), apricot (*P. armeniaca*), and almond (*P. dulcis*). Japanese apricot (*P. mume*) is also a major fruit/flower crop in East Asia commonly known as Chinese “mei” or “mume”. Japanese apricot is believed to have been domesticated firstly in China several thousand years ago, transferred into Japan ca. 2,000 years ago, originally for ornamental purposes (Mega *et al.*, 1988; Horiuchi *et al.*, 1996; Faust *et al.*, 2011). Currently, cultivars are widely diversified mainly based on their usage, such as for pickles (“umeboshi”), syrups/liquors,

and ornamental flowers. However, in contrast to historical implication and conventional categorization, the genetic background of this species remains little known. Japanese apricot is morphologically similar to apricot and plum, and they are all nested in the subgenus *Prunus* (Bortiri *et al.*, 2001). Species of the subgen. *Prunus* are partially compatible for interspecific crossing (Yamaguchi *et al.*, 2018; Morimoto *et al.*, 2019), and some Japanese apricot cultivars have been traditionally thought to carry genetic factors of apricot (*P. armeniaca*) (Mega *et al.*, 1988; Horiuchi *et al.*, 1996), which was later supported by molecular marker analyses (Shimada *et al.*, 1994; Hayashi *et al.*, 2008; Numaguchi *et al.*, 2019). Furthermore, recent breeding programs often utilize interspecific crossing of the subgen. *Prunus*, such as “Sumomo-ume” (*P. salicina* × *P. mume*), or “Pluot” (*P. salicina* × *P. armeniaca*) (Kyotani *et al.*, 1988; Brantley, 2004; Yaegaki *et al.*, 2012).

Given the above information, it would appear that Japanese apricot, and related species in the subgen. *Prunus*, have undergone complicated evolutionary processes, involving natural or artificial selection and potential introgressions among them. Here, to clarify the evolutionary paths to establish the current subgen. *Prunus*, mainly for *P. mume*, we analyzed genome-wide single nucleotide polymorphisms (SNPs) based on targeted resequencing of ca. 15,000 exons in East Asian *P. mume* cultivars. An integrative analysis of selective sweeps and the transition of fragmental genetic structures successfully inferred the importance of interspecific introgressions and lineage-specific selection during the evolution of *P. mume*.

Materials and Methods

Plant materials (Japanese and Taiwanese cultivars and *Prunus* relatives)

We used 112 Japanese and 5 Taiwanese cultivars of Japanese apricot (*P. mume*), and 7 apricot (*P. armeniaca*), 4 Japanese plum (*P. salicina*), and 1 peach (*P. persica*) cultivars (Table S1). For Japanese cultivars of *P. mume*, we used 55 fruit and 45 ornamental cultivars, 8 hybrids between *P. mume* and *P. armeniaca*, and 4 hybrids between *P. mume* and *P. salicina*. Cultivar categorization was based on previous reports (Hayashi *et al.*, 2008; Numaguchi *et al.*, 2019). All plant materials were maintained at the Japanese Apricot Laboratory, Wakayama Fruit Experiment Station (Minabe, Wakayama, Japan) as previously reported by Numaguchi *et al.* (2019) (Table S1).

Target capture sequencing

Genomic DNA was extracted from young leaves using Nucleon PhytoPure (GE Healthcare, Chicago, IL, USA) and subjected to phenol/chloroform purification. We employed a KAPA HyperPlus kit (Kapa Biosystems, Wilmington, MA, USA) to construct gDNA-seq libraries for an Illumina platform. Libraries were barcoded for each sample using single 8-bp NEXTflex adaptors (Bioo Scientific, Austin, TX, USA) and enriched by PCR using PrimeSTAR Max (Takara Bio, Shiga, Japan) with the following protocol: 3 min at 95°C, followed by eight cycles of 10 s at 95°C, 30 s at 65°C, 30 s at 72°C, and final extension for 5 min at 72°C.

To selectively retrieve libraries with exons, a myBaits Custom design kit was used to design 1–20-K probes (Arbor Biosciences, Ann Arbor, MI, USA), which uses biotinylated RNA probes to concentrate fragments carrying sequences of interest (Gnirke

et al., 2009), based on the published genomic and coding sequences of *P. mume* (Zhang *et al.*, 2012). We selected 29,621 non-redundant coding loci showing single hits with BLAST+ (MEGABLAST with -p 70 option) against the *P. mume* genome, for the subsequent bait designing. A 120-mer bait with 25–55 GC% per locus was randomly designed for each locus, and finally we obtained a bait set targeting 15,171 coding loci. We pooled an equal amount of constructed Illumina libraries (eight samples per tube). Pooled libraries were purified by AMPure XP (Beckman Coulter, Indianapolis, IN, USA) and then electrophoresed on 1% agarose gel. We cut out a 300–700-bp area of DNA bands to re-extract libraries using a FastGene Gel/PCR extraction kit (NIPPON Genetics, Tokyo, Japan). Libraries were then subjected to target capture hybridization using myBaits Custom designed probes (Arbor Biosciences). Captured libraries were sequenced using the HiSeq 4000 platform (Illumina, San Diego, CA, USA) (paired-end 100 bp).

SNP calling

In addition to our original sequencing data, we also used published sequencing data. Of the 348 *P. mume* cultivars in the whole-genome sequencing data reported by Zhang *et al.* (2018), 79 derived from China were selected and downloaded from Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) (Table S2). The data were selected to evenly contain all the P1 to P16 phylogenetic clusters reported by Zhang *et al.* (2018). Importantly, the clade P1 contains interspecific hybrids such as *P. mume* × *P. armeniaca* and *P. mume* × *P. salicina*. Raw reads were trimmed with no-demultiplex-allprep-8 (<https://github.com/Comai-Lab/allprep>) to select the reads with high quality (Phred score > 20 over a 5-bp window, length > 35-bp) and containing no ‘N’ and adapter sequences.

The selected reads were mapped against LG1–8 of the peach (*P. persica*) v2.0 reference genome (Verde *et al.*, 2017) using BWA-MEM with default parameters (Li, 2013). PCR duplicates were removed with OverAmp-3 (<http://comailab.genomecenter.ucdavis.edu/index.php/Bwa-doall>). SNPs were called using Samtools mpileup (Li *et al.*, 2009) and VarScan2 mpileup2snp (Koboldt *et al.*, 2009, 2012) with default settings (here, referred to as “Primary_set”). From Primary_set, we removed loci with >20% missing genotyping rate with PLINK (Purcell *et al.*, 2007). Then, missing genotypes were imputed using Beagle 5.0 (Browning *et al.*, 2018) with default settings (“Imputed_set”). We also prepared “Cap_set” by removing the Chinese cultivars from Primary_set. Cap_set was subjected to sequencing depth estimation with VCFtools (Danecek *et al.*, 2011). Imputed_set and Cap_set were used to estimate annotated genomic locations using SnpEff (Cingolani *et al.*, 2012).

Population structure analysis

SNP sets for population structure analyses were prepared based on Imputed_set. We first extracted cultivars of interest and removed loci with a minor allele frequency (MAF) <0.03 and that violated the Hardy–Weinberg equilibrium ($P < 0.0001$) using PLINK. We further pruned SNPs with high ($r^2 > 0.5$) linkage disequilibrium (LD) within a 50-SNP window with 3 SNPs shifting using PLINK (--indep 50 3 2).

Population structure was estimated using three methods: principal component analysis (PCA), Bayesian clustering, and maximum likelihood phylogenetic analysis. PCA was performed using smartpca of EIGENSOFT (Patterson *et al.*, 2006). ADMIXTURE (Alexander *et al.*, 2009) was used for Bayesian clustering. We assumed K

= 2–10, and 10 simulations were carried out for each K value. We then compiled the results of 10 simulations for each K using CLUMPP (Jakobsson & Rosenberg, 2007). The most optimal K was estimated based on cross-validation error (CVE) values calculated according to the ADMIXTURE manual. In the present study, the most optimal K was estimated to be four (CVE = 0.328). A maximum likelihood (ML) phylogenetic tree was constructed using SNPhylo (Lee *et al.*, 2014) with 1,000 bootstrap replications.

For detection of linkage disequilibrium, based on Imputed_set, we removed samples Chi_30, 202, 270, 283, 396, Jap_AM1–8, and SM1–4, which were previously considered to be interspecific hybrids (Hayashi *et al.*, 2008; Zhang *et al.*, 2018; Numaguchi *et al.*, 2019), and additionally Chi_250 and Jap_O2, which were newly classified as “Admixed” in the present study. Pairwise LD was computed using PopLDdecay (Zhang *et al.*, 2019) with -MaxDist 10000, -MAF 0.03, and -Het 0.75 options. The Plot_Multipop function was then used to calculate moving averages of LD for each 10-kb bin.

For detection of genetic differentiation and identity by descent (IBD), the same SNP set as that used for the above population structure analyses was used. Pairwise Weir and Cockerham weighted F_{ST} was calculated using VCFtools. IBD was estimated using pairwise pi-hat values from PLINK.

Identification of selective sweeps

To estimate genomic regions that experienced natural or artificial selection, we used the following methods: composite likelihood ratio (CLR) (Nielsen *et al.*, 2005), nSL (Ferrer-Admetlla *et al.*, 2014), and XP-EHH (Sabeti *et al.*, 2007) tests. Of these, the CLR test is

a method based on SFS, which detects deviations in allele frequency from neutrality at each site. This assumes that a selected allele is fixed in a population (as in annual crops). Conversely, nSL and XP-EHH analyses are based on extended haplotype homozygosity (EHH), which detects elongated linkage disequilibrium blocks around a selected core allele. EHH-based methods work well if a selected allele is not completely fixed in a population but is maintained in a heterozygous state (as in trees, humans, and other animals) (Sabeti *et al.*, 2007; Kosova *et al.*, 2010; Akagi *et al.*, 2016; Lee *et al.*, 2016; Nadachowska-Brzyska *et al.*, 2019). XP-EHH compares EHH values between paired populations (e.g., ancestral and derived populations) and can detect selective sweeps related to population differentiation.

SweeD (Pavlidis *et al.*, 2013) (with -grid 500 flag) was used for calculation of CLR values. Neutral thresholds were determined according to Nielsen *et al.* (2005). We first generated 1,000 simulated neutral genotype datasets using ms (Hudson, 2002), based on the observed number of polymorphic sites (S) and sample size (n) of each subpopulation. Next, we ran SweeD with -grid 500 flag using simulated genotype sets to obtain neutral CLR values. Neutral thresholds were determined as 99% percentile values for each subpopulation. We used selscan v 1.2.0a (Szpiech & Hernandez, 2014) to perform nSL and XP-EHH analyses. We assumed that genetic position was equal to physical position in the XP-EHH analysis. An SNP dataset was generated based on Imputed_set. We first extracted genotype sets for each subpopulation (China, Japan, Taiwan, ornamental, fruit, and small-fruit). Loci with MAF <0.03 were filtered for each dataset, and subsequently, haplotypes were phased using Beagle 5.0. Unstandardized nSL and XP-EHH values were Z-scored using the norm function of selscan and transformed to *P* values.

Genetic differentiation in genome fragments among subgen. *Prunus* species

For SNP data preparation, we first removed admixed cultivars from Imputed_set and subsequently paired Chinese, Japanese, and Taiwanese cultivars with apricots (*P. armeniaca*) or Japanese plums (*P. salicina*). We then filtered loci with MAF <0.03 with PLINK. Here, we attempted to estimate introgressed genomic positions based on sliding window characterization for indices of genetic differentiation. To do this, we assessed the transition of three indices for population differentiation: (i) value of the first principal component (PC1) in PCA, (ii) Q value of the ADMIXTURE analysis with $K = 2$, and (iii) Jost's D value (Jost, 2008), with the sliding window approach. We conducted "Bin-PCA" and "Bin-Admixture" analyses, which refer PCA from scikit-learn (Pedregosa *et al.*, 2011) and ADMIXTURE (Alexander *et al.*, 2009), to consecutively calculate PC1 and Q values, respectively. We used 1-Mb bin and 500-kb walking size in Bin-PCA and Bin-Admixture analyses. PC1 values were Z-transformed based on the equation: $zPC1 = (PC1 - \mu PC1) / \sigma PC1$. Here, $\mu PC1$ and $\sigma PC1$ indicate the average and standard deviation of PC1, respectively. Q values of Bin-Admixture were transformed into absolute values of the difference between each Q value of *P. mume* individuals (Pm_indv.) and average Q values for *P. armeniaca* or *P. salicina* (related_ave.) as follows: $dQ = |Q_{Pm_indv.} - Q_{related_ave.}|$. For calculating Jost's D , we used vcfWindowedFstats in pypgen 0.2.1 (<https://pypi.org/project/pypgen/>) with a window size of 1 Mb.

Detailed analysis on the loci with selective sweep and interspecific introgression

To further examine the inferred region with characteristic selection or introgression, we focused on some specific regions. Especially, we narrowed the possibly introgressed regions on chromosome 8 with Bin-Admixture setting the 50-kb bin and 25-kb walking size. We removed bins with <10 SNPs. Neighbor-joining phylogenetic analysis was performed to visualize the allelic evolution in the specific regions with TASSEL 5.0 (Bradbury *et al.*, 2007), using extracted SNPs of interested regions from Imputed_set. Roots for phylogenetic trees were determined at midpoints using FigTree v1.4.4 (<https://github.com/rambaut/figtree/releases>).

Results

Efficacy of targeted resequencing in the subgenus Prunus

From the targeted resequencing of 129 *Prunus* cultivars (117 *P. mume*, 7 *P. armeniaca*, 4 *P. salicina*, and 1 *P. persica*), 1,096,007,397 of a total 1,177,780,940 reads (93.1%) (deposited at DRA009691; Table S1) were mapped onto LG1–8 of the peach v2.0 genome, including 402,859,421 uniquely mapped reads (34.2%). In Cap_set, we could identify a total of 489,420 SNPs with an average depth of 29.8×, of which each cultivar ranged from 15.2× to 60.2×. SnpEff analysis revealed that total 94.4% of SNPs were located on a genic region (35.3%) or its upstream and downstream regions (26.8% and 32.3%, respectively) (Table S3). In Imputed_set, we obtained a total of 148,953 SNPs, of which the SnpEff result was almost consistent with that of Cap_set (Table S3). Thus, we could successfully and cost-effectively obtain SNPs based on exon capture in subgen. *Prunus* cultivars.

Definition of population structure

A total of 14,310 selected SNPs were used for PCA, ADMIXTURE and ML phylogenetic analyses to reveal the population structure among 208 *Prunus* cultivars (79 Chinese cultivars were added to the 129 cultivars). In all three analyses, Japanese apricot (*P. mume*), apricot (*P. armeniaca*), and Japanese plum (*P. salicina*) were clearly found to form species specific clusters (Figs. 1, 2). This was also supported by the F_{ST} values among the species (Table S4). Hypothetical interspecific hybrids of *P. mume* (Jap_AM1–8, SM1–4, Chi_30, 202, 270, and 396) (Hayashi *et al.*, 2008; Zhang *et al.*, 2018; Numaguchi *et al.*, 2019) were positioned between *P. mume* and the other *Prunus* species, supporting that they are “Admixed” individuals (Figs 1, 2). Importantly, Chinese and Japanese cultivars of *P. mume* were clustered into separate groups, whereas Taiwanese cultivars were clustered with Japanese cultivars (Figs 1a, b, S1). The ML tree (Fig. 2) also supported that the *P. mume* cluster was largely divided into Chinese and Japanese (with Taiwanese) clades with statistical support (bootstrap >60 in ML), including minor exceptions (Jap_O6, 9, 14, 18, 27, 42, 44, and 45; green stars in Fig. 2). In the IBD analysis, some pairs of Japanese and Taiwanese cultivars were inferred to be in first- or second-degree relationships (π -hat = 0.25–0.5; Fig. S2). Conversely, some pairs of Chinese and Japanese cultivars showed first-degree relationships (π -hat = 0.5), but most combinations were genetically distinct (Fig. S2). The highest F_{ST} value was observed between Chinese and Taiwanese cultivar groups, in contrast to the geographical proximity (Table S4). These results differ from conventional (or empirical) observations, which have indicated that Japanese cultivars of *P. mume* were originally introduced from China to Japan relatively recently (ca. 2,000 years ago) via human activities (Horiuchi *et al.*,

1996). Based on the ML tree, Japanese cultivars showed weak differentiation depending on their characteristics or application by humans (Fig. 2). Although subpopulations of fruit, small-fruit, and ornamental cultivars was not clearly divided in PCA, ADMIXTURE, and F_{ST} (Figs S1, 1b, Table S5), in the ML tree, the majority of fruit (36 of 45 cultivars), small-fruit (9 of 10 cultivars), and ornamental (25 of 45 cultivars) cultivars belonged to the same cluster (Fig. 2). This supports the possibility that human preference triggered a recent differentiation of Japanese population from the same genetic resources (Horiuchi *et al.*, 1996; Hayashi *et al.*, 2008; Numaguchi *et al.*, 2019).

LD mostly decayed within ca. 100 kb in all the *P. mume* groups surveyed in the present study (Figs 3, S3), which is much longer than in *P. armeniaca* but shorter than in *P. persica* (Mariette *et al.*, 2016; Akagi *et al.*, 2016; Yu *et al.*, 2018). The extent of LD was slightly different among cultivar groups. For example, LD decayed slower in Japanese cultivars than in the others (Fig. 3). Within the Japanese cultivars, ornamental cultivars exhibited further slower LD decay (Fig. S3), presumably due to their narrow genetic resources and frequent utilization of bud-sport for development of new cultivars especially after the Edo Period (Mega *et al.*, 1988; Horiuchi *et al.*, 1996).

Identification of selective sweeps related to population differentiation in P. mume

Alleles that have undergone positive selection showed, i) reduction in genetic diversity, and ii) extension of haplotype, in adjacent genetic regions and a targeted locus itself (Stephan, 2019). We applied an SFS-based method, SweeD, to detect i), whereas we used extended haplotype homozygosity (EHH)-based method, nSL and XP-EHH, to identify ii).

For Chinese and Japanese populations, the results for SweeD and single-population EHH-based nSL analyses were different (Figs 4a, S4, Table S6). Especially in the Chinese population, no SweeD peak exceeded the neutral threshold (Fig. S4). Tree crops have specific characters, such as long generation time and frequent vegetative propagation, suggesting that selected alleles may have not been completely fixed, like in humans (Voight *et al.*, 2006; Akagi *et al.*, 2016). Sites with only SFS-based peaks (with no EHH-based peaks) are thus suspected to be derived from the occasional reduction of nucleotide diversity, namely, the substitution ratio in the *P. mume* genome, distortion of the availability in SNPs, or simple drift (Akagi *et al.*, 2016). Therefore, in *P. mume*, we mainly focused on the results of the EHH-based analyses in the following sections. Only in the Taiwanese cultivar group was the pattern of significant peaks of SweeD similar to that of nSL (Figs 4a, S4, Table S6), suggesting that Taiwanese cultivars may have experienced stronger selection pressure than the other cultivar groups.

In the nSL analysis, we could identify a strong peak ($P < 1e^{-4}$) on chromosome 8 (ca. 6.7–6.8 Mb), which was common in all Chinese, Japanese, and Taiwanese cultivar groups of *P. mume* (Fig. 4a). A peak on chromosome 6 was also common in three geographic groups, but the Japanese peak was not significant. Strong peaks common in Chinese and Japanese cultivars were observed on chromosome 2. Geographically specific peaks were found on chromosomes 2 and 8 in Chinese, chromosomes 3 and 4 in Japanese, and chromosomes 1, 2, 4, 5, 7, and 8 in Taiwanese populations. These results suggest that the common ancestor of *P. mume* underwent certain positive selection, such as on chromosomes 6 and 8, and thereafter established the three populations based on geographical separation and independent selection.

Furthermore, we conducted two population-based XP-EHH analyses in, (i) the

Chinese cultivars (reference) vs. the Japanese cultivars (derived), (ii) the Japanese cultivars (reference) vs. the Taiwanese cultivars (derived), (iii) the ornamental cultivars (reference) vs. fruit cultivars (derived), and (iv) the fruit cultivars (reference) vs. the small-fruit cultivars (derived) (Fig. 4b). The XP-EHH can focus on the selected alleles that are highly differentiated between populations (Sabeti *et al.*, 2007). Here, negative and positive normalized XP-EHH values indicated that an extended haploblock was observed in the reference and derived populations, respectively (Szpiech & Hernandez, 2014). Most of the strong peaks ($P < 1e^{-4}$; Table S7) were not overlapped with those detected in the nSL analysis (Fig. 4b). In the Chinese vs Japanese analysis, significant peaks were observed on all chromosomes except for chromosome 5 (Fig. 4b). The genomic positions of these peaks were different from those in SweeD (Fig. S4, Table S7), indicating that they have not yet perfectly fixed in the population. Most strong XP-EHH peaks in Japanese vs Taiwanese groups showed negative values, except for ca. 33.9 Mb of chromosome 1 (Table S7), indicating that Japanese cultivars underwent much more extensive selection in the differentiation from Taiwanese cultivars. In the analyses of Japanese ornamental vs fruit cultivars and fruit vs small-fruit, we could also find many significant peaks, presumably involving ongoing selection in favor of human preference in Japan. Especially, peaks of ca. 3.9 Mb in chromosome 6 overlapped with the SweeD peak (Figs 4b, S5, Table S7), suggesting strong selection pressure for the small-fruit trait in Japanese cultivars. Accordingly, tests for selection in *P. mume* populations identified its tree-crop-specific patterns for natural or artificial selection.

Fragmental interspecific introgression in the subgen. *Prunus*

We detected, in 1-Mb bins, the zPC1 (Bin-PCA) and proportion of the Q values in Admixture with $K = 2$ (Bin-Admixture) to scan for genome-wide transition of genetic differentiation or potential introgressions for *P. mume* vs *P. armeniaca* or *P. salicina*. The Bin-PCA and Bin-Admixture showed mostly consistent results (Figs S6, S7). In most of the chromosomes, *P. mume* genomes showed signs of fragmental interspecific introgression (or no clear differentiation between species) from *P. armeniaca* (Fig. S6) or *P. salicina* (Fig. S7). When the three geographic groups were compared, Japanese cultivars showed the most frequent signals of interspecific introgressions from *P. armeniaca* or *P. salicina* (Figs S6b, S7b), while Taiwanese cultivars rarely showed them (Figs S6c, S7c). We also calculated a distance-matrix-based Jost's D statistic in 1-Mb bins. Jost's D values tended to be low in the genomic regions with interspecific introgression signals in Bin-PCA and Bin-Admixture (Figs S6, S7). However, unlike Bin-PCA and Bin-Admixture, the transition of Jost's D values substantially fluctuated in most chromosomes (Figs S6, S7). The overlapped region of Bin-PCA, Bin-Admixture, and Jost's D signals may be a strong signature of interspecific introgressions, indicating the especially low allelic divergence between *P. mume* and relatives.

Importantly, some signals of interspecific introgressions were overlapped with the selective sweep (nSL peaks) (Fig. 5; hereafter, "introgression-sweep"), indicating that introgressed regions may have been positively selected in the evolution of *P. mume* populations. They were located on chromosomes 6 and 8 in Chinese cultivars, 2, 3, 4, 6, and 8 in Japanese cultivars, and 6 and 8 in Taiwanese cultivars. In chromosomes 6 and 8, nSL signals harbor fragment introgressions from both *P. armeniaca* and *P. salicina* (Fig. 5a-c). In chromosome 2, 3, and 4 of Japanese cultivars, introgression signals were accompanied by nSL peaks independently of other groups (Fig. 5b). These results suggest

that interspecific introgressions independently contributed to the establishment of not only *P. mume* but also of each geographical group following the divergence of these subpopulations.

It is worth noting that the strongest introgression-sweep signal was detected around 6 Mb on chromosome 8 (Fig. 6a,b). Fine assessment of the Q values with shorter bins (50 kb) increased the resolution for the genomic region with overlap of Bin-Admixture and nSL signals around 6.7–7.1 Mb (Fig. 6c). Next, we compared evolutionary topologies constructed from the SNPs in region 1 (6.0–6.7 Mb), region 2 (6.7–7.1 Mb), region 3 (7.1–8.5 Mb), and in the whole chromosome 8, according to an approach by Choi & Purugganan (2018) (Fig. 6d–g). The topology for the whole chromosome 8 (Fig. 6d) was mostly consistent with that for the whole genome (Fig. 2) in accordance with the divergence of species and populations. For regions 1–3, only region 2 showed a distinct topology with alleles of *P. mume* and related species in the subgenus *Prunus* grouped together, while alleles putatively introgressed and under selection were nested in a single clade with *P. salicina* alleles (Fig. 6f), showing very small genetic differentiation (alleles with green band in Fig. 6f). Consequently, region 2 (6.7–7.1 Mb) is thought to have been exposed to positive selection pressure, which may be associated with the adaptive evolution of *P. mume* to import advantageous alleles from other species.

Discussion

Completed and ongoing population differentiation in P. mume

In the present study, we revealed that Chinese and Japanese cultivars of *P. mume* showed distinct population differentiation (Figs 1, 2). Conversely, Taiwanese cultivars belonged to the Japanese clade but clustered independently from the Japanese cultivars, consistent with the results of previous studies (Hayashi *et al.*, 2008; Numaguchi *et al.*, 2019). These results suggest that the differentiation of Chinese and Japanese populations predated that of the Taiwanese population, which is inconsistent with the conventional belief that *P. mume* cultivars were derived from Chinese ones, and recently (ca. 2000 years ago), people have been introduced to other regions (Mega *et al.*, 1988; Faust *et al.*, 2011). There is a record describing wild *P. mume* accessions collected from western Japan that have been clonally maintained (<https://agriknowledge.affrc.go.jp/RN/3030041889>; in Japanese). Those samples may have important genetic information related to the origin of Japanese cultivars. We may directly investigate the population structure of domesticated crops using ancient DNA (aDNA) extracted from ancient plant remains (Frantz *et al.*, 2016; Mascher *et al.*, 2016; Kistler *et al.*, 2018; Narasimhan *et al.*, 2019; Allaby *et al.*, 2019; Smith *et al.*, 2019). Thus, collaboration between geneticists and archaeologists will make rapid progress in studies on crop domestication.

Contribution of genomic fragments undergoing interspecific introgressions and positive selections in the evolution of *P. mume*

We observed a higher number of strong (and successive) peaks in EHH-based scans than in an SFS-based SweeD analysis. This suggested that, in tree (or perennial) crops, most selected alleles are maintained in a heterozygous state, as suggested previously by Akagi *et al.* (2016). Since most SNPs called in our exon capture approach

were located around protein coding regions (Table S3), it is expected that we may be able to detect functional nucleotide polymorphisms (FNPs) even with low sequencing coverage for each individual (Bamshad *et al.*, 2011; Kaur & Gaikwad, 2017). We could detect candidate genes potentially involved in environmental adaptation (Fig. 4, Tables S6, S7). For instance, in the regions with strong selective sweep, leucine rich repeat containing proteins (e.g., Prupe1G161800, Prupe4G157900, Prupe8G046600, Prupe.8G012000, and Prupe.8G012800) and receptor-like kinases (e.g., Prupe.6G183600 and Prupe.6G261400) would commonly contribute to pathogen recognition pathways (Ellis *et al.*, 2000). The BTB/POZ-MATH-TRAF-like protein (Prupe.1G107200) is potentially associated with virus resistance in *P. armeniaca* (Mariette *et al.*, 2016). Genes potentially involved in stress response (e.g., Prupe.2G089100, Prupe.2G145200, and Prupe.3G110300) (Vij & Tyagi, 2008; Cheng *et al.*, 2011) were also identified (Tables S6, S7). These results suggest that selection on biotic or abiotic stress responsive genes may have contributed to the geographic separation of Chinese, Japanese, and Taiwanese cultivars.

According to the results of Bin-PCA, Bin-Admixture, and Jost's *D* analyses, it was suggested that substantial fractions of *P. mume* genomes have frequently exchanged genomic fractions with related species of the subgen. *Prunus*. Large fractions of introgression may indicate that *P. mume*, especially in Japanese cultivars, have experienced a limited number of generations since the interspecific hybridization. Two genomic regions on chromosomes 6 and 8 were detected to have interspecific introgressions in Chinese, Japanese, and Taiwanese cultivar groups (Figs 5, 6). Particularly, a 6.74–6.80-Mb region in the region 2 on chromosome 8 contained high nSL peaks commonly detected among three groups (Table S6). Although this region carries

no genes in the reference peach (*P. persica*) genome, we can propose the two following possibilities: 1) only in the *P. mume* genome, the selected haploblock in the corresponded region harbors candidate genes, and 2) this region includes *cis*-regulatory elements affecting gene expression in the flanking regions. For hypothesis 2), often, *cis*-elements were located distantly upstream (>10 kb) of the genes (Clark *et al.*, 2004; Konishi *et al.*, 2006; Ishii *et al.*, 2013; Ricci *et al.*, 2019). Prupe.8G057100 (mitochondrial transcription termination factor) is located ca. 80 kb from the selected haploblock (Fig. S8). Mitochondrial transcription termination factors have been reported to be related to abiotic stress response by controlling the expression level of nuclear genes (Quesada, 2016). Another region on chromosome 6 (ca. 15.2–15.3 Mb) also showed common introgression-sweep (Figs 5, S9, Table S6). Although this region also harbored no annotated genes in the peach reference genome as well as the described introgression-sweep region on chromosome 8, it may have been important in the evolution of *P. mume*.

Other than the introgression-sweep commonly underwent among the geographical cultivar groups, our exon capture sequencing would allow efficient identification of FNPs in introgression-sweep regions specific to each cultivar group. Japanese cultivars have the largest fractions of interspecific introgressions, and they also show several geographically unique introgression-sweep regions in nSL (Fig. 5b, Table S6). Of them, interleukin-1 receptor-associated kinase 1 (IRAK1) (Prupe.2G223200), which harbors the introgression signal of *P. salicina* on chromosome 2 (ca. 25 Mb), may act for pathogen recognition pathways (Jebanathirajah *et al.*, 2002; Dardick & Ronald, 2006). On chromosome 3 (ca. 3.6 Mb), 3-epi-6-deoxocathasterone 23-monooxygenase (Prupe.3G050900) was reported to be associated with brassinosteroid biosynthesis (Ohnishi *et al.*, 2006) and is potentially involved in the fruit-enlargement process.

Premnaspirodiene oxygenase (Prupe.4G237900) on chromosome 4 (ca. 15.5 Mb) is a kind of cytochrome P450, which participates in terpene biosynthesis (Weitzel & Simonsen, 2015). Terpenes are the largest class of plant-derived compounds that have numerous potential applications across food, beverage, pharmaceutical, cosmetic, and agriculture industries (Boutanaev *et al.*, 2015). These results suggest that Japanese cultivars might import genetic factors from other species to satisfy the preference of people in Japan.

Inferring the evolution of *P. mume*

The results so far propose an evolutionary model for the establishment of the current *P. mume* populations, where frequent interspecific introgressions with natural/artificial selection have played important roles (Fig. 7). Several hybridization events might have occurred in the differentiation of these three species of the subgenus *Prunus* (*P. mume*, *P. armeniaca*, and *P. salicina*). During the formation of the *P. mume* common ancestor, important introgressions related to unknown phenotypic changes on chromosomes 6 and 8 (Figs 5, 6, S9) were positively selected. After the establishment of the original *P. mume*, three core populations were further differentiated to be adapted to China, Japan, and Taiwan, in which independent introgressions from *P. armeniaca* or *P. salicina* and positive selection on some regions might have contributed to the establishment of each population (e.g., chromosomes 2, 3, and 4 in Japanese cultivars, Fig. 5b). Recently, the *P. mume* Japanese cultivars differentiated to form sub-populations, such as ornamental, fruit, and small-fruit clusters, based on human-preference-associated selection (Fig. 4b, Table S7).

The emphasis of the present study was the finding that there are many naturally/artificially selected regions derived from interspecific introgressions, and that they have considerably contributed to the establishment of current *P. mume* populations. Interspecific introgressions are thought to also be important in the evolution of other annual or perennial crops, such as rice (Choi & Purugganan, 2018), wheat (He *et al.*, 2019), maize (Hufford *et al.*, 2013; Brandenburg *et al.*, 2017), apple (Cornille *et al.*, 2012), and olive (Diez *et al.*, 2015; Gros-Balthazard *et al.*, 2019). These reports pointed out the importance of crop-wild introgressions to transfer wild beneficial alleles into domesticates and vice versa, which enabled them to rapidly expand and adapt to new climatic and agricultural conditions. Notwithstanding, especially in woody crops, the genomic landscape of interspecific introgressions and their actual contribution to evolution remains poorly understood. Our findings shed light on the complicated nature of adaptive evolution with interspecific introgressions among domesticated tree crops.

Acknowledgements

We are deeply grateful to researchers and technicians at the Japanese Apricot Laboratory, Wakayama Fruit Tree Experiment Station for maintaining the *Prunus* cultivars. We used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. This work was supported by JSPS KAKENHI Grant Number JP18K14449 to K. N. The English language of this paper was edited by Editage (<https://www.editage.com/>).

Author contributions

547

548 K. N. and T. A. planned and designed the research. K. N., Y. K. and R. I. carried out the
549 targeted resequencing experiments. K. N. analyzed the data. T. A. developed
550 bioinformatic approaches. K. N., T. A. and T. I. wrote and revised the manuscript.

551

552 **Data availability statement**

553

554 Raw fastq reads for 129 *Prunus* cultivars resequenced in this study were deposited in the
555 Sequence Read Archive (SRA) under DRA accession number DRA009691 (Bioproject
556 accession number PRJDB9365). Sources for all downloaded data are referred to in the
557 Supplementary information.

558

559 **References**

560

561 **Akagi T, Hanada T, Yaegaki H, Gradziel TM, Tao R. 2016.** Genome-wide view of
562 genetic diversity reveals paths of selection and cultivar differentiation in peach
563 domestication. *DNA Research* **23**: 271–282.

564 **Alexander DH, Novembre J, Lange K. 2009.** Fast model-based estimation of ancestry
565 in unrelated individuals. *Genome Research* **19**: 1655–1664.

566 **Allaby RG, Ware RL, Kistler L. 2019.** A re-evaluation of the domestication bottleneck
567 from archaeogenomic evidence. *Evolutionary Applications* **12**: 29–37.

568 **Baack EJ, Rieseberg LH. 2007.** A genomic view of introgression and hybrid speciation.
569 *Current Opinion in Genetics & Development* **17**: 513–518.

570 **Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure**

571 **J. 2011.** Exome sequencing as a tool for Mendelian disease gene discovery. *Nature*
572 *Reviews Genetics* **12**: 745–755.

573 **Bortiri E, Oh S-H, Jiang J, Baggett S, Granger A, Weeks C, Buckingham M, Potter**
574 **D, Parfitt DE. 2001.** Phylogeny and systematics of *Prunus* (Rosaceae) as
575 determined by sequence analysis of ITS and the chloroplast trnL-trnF spacer DNA.
576 *Systematic Botany* **26**: 797–807.

577 **Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ, Osbourn A. 2015.**
578 Investigation of terpene diversification across multiple sequenced plant genomes.
579 *Proceedings of the National Academy of Sciences, USA* **112**: E81–E88.

580 **Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007.**
581 TASSEL: software for association mapping of complex traits in diverse samples.
582 *Bioinformatics* **23**: 2633–2635.

583 **Brandenburg J-T, Mary-Huard T, Rigail G, Hearne SJ, Corti H, Joets J, Vitte C,**
584 **Charcosset A, Nicolas SD, Tenaillon MI. 2017.** Independent introductions and
585 admixtures have contributed to adaptation of European maize and its American
586 counterparts. *PLOS Genetics* **13**: e1006666.

587 **Brantley W. 2004.** About pluots. *Gastronomica* **4**: 84–89.

588 **Browning BL, Zhou Y, Browning SR. 2018.** A one-penny imputed genome from next-
589 generation reference panels. *American Journal of Human Genetics* **103**: 338–348.

590 **Cheng Y, Zhou W, El Sheery NI, Peters C, Li M, Wang X, Huang J. 2011.**
591 Characterization of the Arabidopsis glycerophosphodiester phosphodiesterase
592 (GDPD) family reveals a role of the plastid-localized AtGDPD1 in maintaining
593 cellular phosphate homeostasis under phosphate starvation. *The Plant Journal* **66**:
594 781–795.

595 **Choi JY, Purugganan MD. 2018.** Multiple origin but single domestication led to *Oryza*
596 *sativa*. *G3: Genes, Genomes, Genetics* **8**: 797–803.

597 **Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden**
598 **DM. 2012.** A program for annotating and predicting the effects of single
599 nucleotide polymorphisms, SnpEff. *Fly* **6**: 80–92.

600 **Civán P, Brown TA. 2018.** Role of genetic introgression during the evolution of
601 cultivated rice (*Oryza sativa* L.). *BMC Evolutionary Biology* **18**: 57.

602 **Clark RM, Linton E, Messing J, Doebley JF. 2004.** Pattern of diversity in the genomic
603 region near the maize domestication gene *tb1*. *Proceedings of the National*
604 *Academy of Sciences, USA* **101**: 700–707.

605 **Cornille A, Gladieux P, Smulders MJM, Roldán-Ruiz I, Laurens F, Le Cam B,**
606 **Nersesyan A, Clavel J, Olonova M, Feugey L, et al. 2012.** New insight into the
607 history of domesticated apple: secondary contribution of the European wild apple
608 to the genome of cultivated varieties. *PLoS Genetics* **8**: e1002703.

609 **Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE,**
610 **Lunter G, Marth GT, Sherry ST, et al. 2011.** The variant call format and
611 VCFtools. *Bioinformatics* **27**: 2156–2158.

612 **Dardick C, Ronald P. 2006.** Plant and animal pathogen recognition receptors signal
613 through non-RD kinases. *PLoS Pathogens* **2**: e2.

614 **Diamond J. 2002.** Evolution, consequences and future of plant and animal domestication.
615 *Nature* **418**: 700–707.

616 **Diez CM, Trujillo I, Martinez-Urdiroz N, Barranco D, Rallo L, Marfil P, Gaut BS.**
617 **2015.** Olive domestication and diversification in the Mediterranean Basin. *New*
618 *Phytologist* **206**: 436–447.

619 **Ellis J, Dodds P, Pryor T. 2000.** Structure, function and evolution of plant disease
620 resistance genes. *Current Opinion in Plant Biology* **3**: 278–284.

621 **Faust M, Timon B, Surányi D, Nyujtó F, Gradziel TM. 2011.** *Origin and*
622 *Dissemination of Prunus Crops: Peach, Cherry, Apricot, Plum and Almond.*
623 Leuven, Belgium: ISHS.

624 **Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014.** On detecting
625 incomplete soft or hard selective sweeps using haplotype structure. *Molecular*
626 *Biology and Evolution* **31**: 1275–1291.

627 **Frantz LAF, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A,**
628 **Linderholm A, Mattiangeli V, Teasdale MD, Dimopoulos EA, et al. 2016.**
629 Genomic and archaeological evidence suggest a dual origin of domestic dogs.
630 *Science* **352**: 1228–1231.

631 **Gaut BS, Díez CM, Morrell PL. 2015.** Genomics and the contrasting dynamics of
632 annual and perennial domestication. *Trends in Genetics* **31**: 709–719.

633 **Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T,**
634 **Giannoukos G, Fisher S, Russ C, et al. 2009.** Solution hybrid selection with
635 ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature*
636 *Biotechnology* **27**: 182–189.

637 **Gros-Balthazard M, Besnard G, Sarah G, Holtz Y, Leclercq J, Santoni S, Wegmann**
638 **D, Glémin S, Khadari B. 2019.** Evolutionary transcriptomics reveals the origins
639 of olives and the genomic changes associated with their domestication. *The Plant*
640 *Journal* **100**: 143–157.

641 **Harrison RG, Larson EL. 2014.** Hybridization, introgression, and the nature of species
642 boundaries. *Journal of Heredity* **105**: 795–809.

643 **Hayashi K, Shimazu K, Yaegaki H, Yamaguchi M, Iketani H, Yamamoto T. 2008.**
644 Genetic diversity in fruiting and flower-ornamental Japanese apricot (*Prunus*
645 *mume*) germplasms assessed by SSR markers. *Breeding Science* **58**: 401–410.

646 **He F, Pasam R, Shi F, Kant S, Keeble-Gagnere G, Kay P, Forrest K, Fritz A, Hucl P,**
647 **Wiebe K, et al. 2019.** Exome sequencing highlights the role of wild-relative
648 introgression in shaping the adaptive landscape of the wheat genome. *Nature*
649 *Genetics* **51**: 896–904.

650 **Horiuchi S, Yoshida M, Kariya H, Nakamura T, Hasebe H, Suzaki T, Sakitani T.**
651 **1996.** *Nihonnoume Sekainoume*. Tokyo, Japan: Yokendo (Japan).

652 **Hudson RR. 2002.** Generating samples under a Wright-Fisher neutral model of genetic
653 variation. *Bioinformatics* **18**: 337–338.

654 **Hufford MB, Lubinsky P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra**
655 **J. 2013.** The genomic signature of crop-wild introgression in maize. *PLoS*
656 *Genetics* **9**: e1003477.

657 **Ishii T, Numaguchi K, Miura K, Yoshida K, Thanh PT, Htun TM, Yamasaki M,**
658 **Komeda N, Matsumoto T, Terauchi R, et al. 2013.** *OsLGI* regulates a closed
659 panicle trait in domesticated rice. *Nature Genetics* **45**: 462–465.

660 **Jakobsson M, Rosenberg NA. 2007.** CLUMPP: a cluster matching and permutation
661 program for dealing with label switching and multimodality in analysis of
662 population structure. *Bioinformatics* **23**: 1801–1806.

663 **Jebanathirajah JA, Peri S, Pandey A. 2002.** Toll and interleukin-1 receptor (TIR)
664 domain-containing proteins in plants: a genomic perspective. *Trends in Plant*
665 *Science* **7**: 388–391.

666 **Jost L. 2008.** G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17**:

667 4015–4026.

668 **Kaur P, Gaikwad K. 2017.** From genomes to GENE-omes: exome sequencing concept
669 and applications in crop improvement. *Frontiers in Plant Science* **8**: 2164.

670 **Kistler L, Maezumi SY, Souza JG de, Przelomska NAS, Costa FM, Smith O, Loisele**
671 **H, Ramos-Madriral J, Wales N, Ribeiro ER, et al. 2018.** Multiproxy evidence
672 highlights a complex evolutionary legacy of maize in South America. *Science*
673 **362**: 1309–1313.

674 **Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock**
675 **GM, Wilson RK, Ding L. 2009.** VarScan: variant detection in massively parallel
676 sequencing of individual and pooled samples. *Bioinformatics* **25**: 2283–2285.

677 **Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis**
678 **ER, Ding L, Wilson RK. 2012.** VarScan 2: somatic mutation and copy number
679 alteration discovery in cancer by exome sequencing. *Genome Research* **22**: 568–
680 576.

681 **Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M. 2006.** An SNP
682 caused loss of seed shattering during rice domestication. *Science* **312**: 1392–1396.

683 **Kosova G, Pickrell JK, Kelley JL, McArdle PF, Shuldiner AR, Abney M, Ober C.**
684 **2010.** The CFTR Met 470 allele is associated with lower birth rates in fertile men
685 from a population isolate. *PLoS Genetics* **6**: e1000974.

686 **Kyotani H, Yoshida M, Yamaguchi M, Ishizawa Y, Kozono T, Nshida T, Kanato K.**
687 **1988.** Breeding of plum-mume parental lines ‘PM-1-1’ and ‘PM-1-4’,
688 interspecific hybrids of Japanese plum (*Prunus salicina* Lindl.) and mume (*P.*
689 *mume* Sieb. et Zucc.). *Bulletin of the Fruit Tree Research Station* **15**: 1–10.

690 **Lee W, Ahn S, Taye M, Sung S, Lee H-J, Cho S, Kim, H. 2016.** Detecting positive

691 selection of Korean Native Goat populations using next-generation sequencing.
692 *Molecules and Cells* **39**: 862–868.

693 **Lee T-H, Guo H, Wang X, Kim C, Paterson AH. 2014.** SNPhylo: a pipeline to construct
694 a phylogenetic tree from huge SNP data. *BMC Genomics* **15**: 162.

695 **Li H. 2013.** Aligning sequence reads, clone sequences and assembly contigs with BWA-
696 MEM. *arXiv*:1303.3997 [q-bio].

697 **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,**
698 **Durbin R. 2009.** The sequence alignment/map format and SAMtools.
699 *Bioinformatics* **25**: 2078–2079.

700 **Mariette S, Wong Jun Tai F, Roch G, Barre A, Chague A, Decroocq S, Groppi A,**
701 **Laizet Y, Lambert P, Tricon D, et al. 2016.** Genome-wide association links
702 candidate genes to resistance to *Plum Pox Virus* in apricot (*Prunus armeniaca*).
703 *New Phytologist* **209**: 773–784.

704 **Mascher M, Schuenemann VJ, Davidovich U, Marom N, Himmelbach A, Hübner S,**
705 **Korol A, David M, Reiter E, Riehl S, et al. 2016.** Genomic analysis of 6,000-
706 year-old cultivated grain illuminates the domestication history of barley. *Nature*
707 *Genetics* **48**: 1089–1093.

708 **Mega K, Tomita E, Kitamura S, Saito S, Mizukami S. 1988.** Ume. In: Aoba T, ed. *The*
709 *grand dictionary of horticulture*. Tokyo, Japan: Shogakukan, 289–300 (Japan).

710 **Molnár-Láng M, Ceoloni C, Doležel J. 2015.** *Alien Introgression in Wheat:*
711 *Cytogenetics, Molecular Biology, and Genomics*. New York, USA: Springer.

712 **Morimoto T, Kitamura Y, Numaguchi K, Akagi T, Tao R. 2019.** Characterization of
713 post-mating interspecific cross-compatibility in *Prunus* (Rosaceae). *Scientia*
714 *Horticulturae* **246**: 693–699.

715 **Nadachowska-Brzyska K, Burri R, Ellegren H. 2019.** Footprints of adaptive evolution
716 revealed by whole Z chromosomes haplotypes in flycatchers. *Molecular Ecology*
717 **28:** 2290–2304.

718 **Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S,**
719 **Lazaridis I, Nakatsuka N, Olalde I, Lipson M, et al. 2019.** The formation of
720 human populations in South and Central Asia. *Science* **365:** eaat7487.

721 **Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005.**
722 Genomic scans for selective sweeps using SNP data. *Genome Research* **15:** 1566–
723 1575.

724 **Numaguchi K, Ishio S, Kitamura Y, Nakamura K, Ishikawa R, Ishii T. 2019.**
725 Microsatellite marker development and population structure analysis in Japanese
726 apricot (*Prunus mume* Sieb. et Zucc.). *The Horticulture Journal* **88:** 222–231.

727 **Ohnishi T, Szatmari A-M, Watanabe B, Fujita S, Bancos S, Koncz C, Lafos M,**
728 **Shibata K, Yokota T, Sakata K, et al. 2006.** C-23 hydroxylation by *Arabidopsis*
729 CYP90C1 and CYP90D1 reveals a novel shortcut in brassinosteroid biosynthesis.
730 *Plant Cell* **18:** 3275–3288.

731 **Pankin A, Altmüller J, Becker C, von Korff M. 2018.** Targeted resequencing reveals
732 genomic signatures of barley domestication. *New Phytologist* **218:** 1247–1259.

733 **Patterson N, Price AL, Reich D. 2006.** Population structure and eigenanalysis. *PLoS*
734 *Genetics* **2:** e190.

735 **Pavlidis P, Živković D, Stamatakis A, Alachiotis N. 2013.** SweeD: Likelihood-based
736 detection of selective sweeps in thousands of genomes. *Molecular Biology and*
737 *Evolution* **30:** 2224–2234.

738 **Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M,**

739 **Prettenhofer P, Weiss R, Dubourg V, et al. 2011.** Scikit-learn: machine learning
740 in Python. *Journal of Machine Learning Research* **12**: 2825–2830.

741 **Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J,**
742 **Sklar P, de Bakker PIW, Daly MJ, et al. 2007.** PLINK: a tool set for whole-
743 genome association and population-based linkage analyses. *American Journal of*
744 *Human Genetics* **81**: 559–575.

745 **Purugganan MD, Fuller DQ. 2009.** The nature of selection during plant domestication.
746 *Nature* **457**: 843–848.

747 **Quesada V. 2016.** The roles of mitochondrial transcription termination factors (MTERFs)
748 in plants. *Physiologia Plantarum* **157**: 389–399.

749 **Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M,**
750 **Mejía-Guerra MK, Colomé-Tatché M, et al. 2019.** Widespread long-range cis-
751 regulatory elements in the maize genome. *Nature Plants* **5**: 1237–1249.

752 **Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne**
753 **EH, McCarroll SA, Gaudet R, et al. 2007.** Genome-wide detection and
754 characterization of positive selection in human populations. *Nature* **449**: 913–918.

755 **Shimada T, Haji T, Yamaguchi M, Takeda T, Nomura K, Yoshida M. 1994.**
756 Classification of mume (*Prunus mume* Sieb. et Zucc.) by RAPD assay. *Journal of*
757 *the Japanese Society for Horticultural Science* **63**: 543– 551 (Japan).

758 **Smith O, Nicholson WV, Kistler L, Mace E, Clapham A, Rose P, Stevens C, Ware R,**
759 **Samavedam S, Barker G, et al. 2019.** A domestication history of dynamic
760 adaptation and genomic deterioration in Sorghum. *Nature Plants* **5**: 369–379.

761 **Stephan W. 2019.** Selective Sweeps. *Genetics* **211**: 5–13.

762 **Suarez - Gonzalez A, Hefer CA, Lexer C, Douglas CJ, Cronk QCB. 2018.**

763 Introgression from *Populus balsamifera* underlies adaptively significant variation
764 and range boundaries in *P. trichocarpa*. *New Phytologist* **217**: 416–427.

765 **Szpiech ZA, Hernandez RD. 2014.** selscan: an efficient multithreaded program to
766 perform EHH-based scans for positive selection. *Molecular Biology and*
767 *Evolution* **31**: 2824–2827.

768 **Verde I, Jenkins J, Dondini L, Micali S, Pagliarani G, Vendramin E, Paris R,**
769 **Aramini V, Gazza L, Rossini L, et al. 2017.** The Peach v2.0 release: high-
770 resolution linkage mapping and deep resequencing improve chromosome-scale
771 assembly and contiguity. *BMC Genomics* **18**: 225.

772 **Vij S, Tyagi AK. 2008.** A20/AN1 zinc-finger domain-containing proteins in plants and
773 animals represent common elements in stress response. *Functional & Integrative*
774 *Genomics* **8**: 301–307.

775 **Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006.** A map of recent positive
776 selection in the human genome. *PLOS Biology* **4**: e72.

777 **Weitzel C, Simonsen HT. 2015.** Cytochrome P450-enzymes involved in the biosynthesis
778 of mono- and sesquiterpenes. *Phytochemistry Reviews* **14**: 7–24.

779 **Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X, Ruiz M,**
780 **Scalabrin S, Terol J, et al. 2014.** Sequencing of diverse mandarin, pummelo and
781 orange genomes reveals complex history of admixture during citrus domestication.
782 *Nature Biotechnology* **32**: 656–662.

783 **Wu GA, Terol J, Ibanez V, López-García A, Pérez-Román E, Borredá C, Domingo**
784 **C, Tadeo FR, Carbonell-Caballero J, Alonso R, et al. 2018.** Genomics of the
785 origin and evolution of Citrus. *Nature* **554**: 311–316.

786 **Yaegaki H, Yamaguchi M, Haji T, Suesada Y, Miyake M, Kihara T, Suzuki K, Uchida**

787 **M. 2012.** New Japanese apricot cultivar ‘Tsuyuakane’. *Bulletin of the NARO*
788 *Institute of Fruit Tree Science* **13**: 1–6 (Japan).

789 **Yamaguchi M, Yaegaki H, Suesada Y, Haji T, Miyake M. 2018.** Study on Interspecific
790 Hybridization among the Subgenera *Amygdalus*, *Prunophora*, and *Cerasus*.
791 *Horticultural Research (Japan)* **17**: 147–159 (Japan).

792 **Yu Y, Fu J, Xu Y, Zhang J, Ren F, Zhao H, Tian S, Guo W, Tu X, Zhao J, et al. 2018.**
793 Genome re-sequencing reveals the evolutionary history of peach fruit edibility.
794 *Nature Communications* **9**: 5404.

795 **Zeder MA. 2015.** Core questions in domestication research. *Proceedings of the National*
796 *Academy of Sciences, USA* **112**: 3191–3198.

797 **Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, Tao Y, Wang J, Yuan Z, Fan**
798 **G, et al. 2012.** The genome of *Prunus mume*. *Nature Communications* **3**: 1318.

799 **Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. 2019.** PopLDdecay: a fast and
800 effective tool for linkage disequilibrium decay analysis based on variant call
801 format files. *Bioinformatics* **35**: 1786–1788.

802 **Zhang Q, Zhang H, Sun L, Fan G, Ye M, Jiang L, Liu X, Ma K, Shi C, Bao F, et al.**
803 **2018.** The genetic architecture of floral traits in the woody plant *Prunus mume*.
804 *Nature Communications* **9**: 1702.

805

Figure legends

Fig. 1 Population structure analysis in *Prunus mume*. **(a)** Principal component analysis (PCA) of all the 208 *Prunus* cultivars. **(b)** Proportion of ancestry for all the 208 *Prunus* cultivars from $K = 2-6$ inferred with ADMIXTURE. Proportions of the membership to each cluster are shown with the lengths of the colored bar (y-axis). CVE: cross validation error. F: fruit cultivars, FS: small-fruit cultivars, O: ornamental cultivars, AM: putative hybrids with *P. armeniaca*, SM: putative hybrids with *P. salicina*.

Fig. 2 Maximum likelihood phylogenetic tree inferred with all the 208 *Prunus* cultivars. Values beside the nodes indicate bootstrap values generated with 1,000 replications. *P. mume* cultivars clustered with admixed cultivars are shown with navy stars. Dotted lines within Japanese clusters indicate the characteristic clusters for fruit (F), small-fruit (FS), and ornamental (O) cultivars. Green stars indicate Japanese cultivars in Chinese clusters. Chi: Chinese cultivars, Jap: Japanese cultivars, Tai: Taiwanese cultivars, F: fruit cultivars, FS: small-fruit cultivars, O: ornamental cultivars, AM: putative hybrids with *P. armeniaca*, SM: putative hybrids with *P. salicina*, Pa: *P. armeniaca*, Ps: *P. salicina* and Pp: *P. persica*.

Fig. 3 Patterns of linkage disequilibrium decay in Chinese, Japanese, and Taiwanese cultivars of *Prunus mume*.

Fig. 4 Identification of selective sweeps in Chinese, Japanese, and Taiwanese cultivars of *Prunus mume* based on, **(a)** single-population nSL and **(b)** dual-population XP-EHH analyses. P values calculated with normalized nSL values are shown in $-\log_{10}$ scale.

Detailed information of SNPs with $-\log_{10}P > 4$ (dotted line) is summarized in Tables S6 (nSL) and S7 (XP-EHH).

Fig. 5 Examination of conformity between putative interspecific introgression and selective sweep signals. Genomic locations of Bin-Admixture signals were compared with those of nSL peaks. Bin-Admixture (1-Mb-binned) analyses with *Prunus armeniaca* (Pa, red signals) and *P. salicina* (Ps, blue signals), and nSL scans were performed in, (a) Chinese, (b) Japanese, and (c) Taiwanese cultivars of *P. mume*. The degree of introgression is indicated by the color scales to the right (0: highly introgressed–1: not introgressed). Potential introgression-sweep regions are highlighted by green dotted rectangles.

Fig. 6 Consistent region for interspecific introgression and selective sweep on chromosome 8. (a) Bin-PCA, Bin-Admixture, and Jost's *D* patterns in Chinese, Japanese, and Taiwanese cultivars, in comparison to *Prunus armeniaca* (Pa, upper panels) and *P. salicina* (Ps, lower panels). The degree of introgression in Bin-Admixture is indicated by the color scales to the right (0: highly introgressed–1: not introgressed). (b) Transitions of nSL for an index of selective sweep. *P* values calculated with normalized nSL values were shown in $-\log_{10}$ scale. (c) Close up of the genomic fragment (ca. 6.0–8.5Mb) showing overlapped nSL peaks and potential interspecific-introgression in Japanese cultivars of *P. mume*. We further divided this region into three sub-fragments (Regions 1–3) according to the pattern of nSL plots to assess their phylogenetic relationships. (d) Neighbor-joining phylogenetic trees with the whole SNPs in chromosome 8 and with the SNPs in Regions 1–3. The tree for Region 2 showed a topology inconsistent with the

whole chromosome 8 and the flanking regions (Regions 1 and 3), and was also inconsistent with the estimated speciation pattern of the subgenus *Prunus*. Alleles that underwent potential selective sweeps were indicated with a green solid line.

Fig. 7 Tentative process of evolution in *Prunus mume*. The three subgenus *Prunus* species, Japanese apricot (*P. mume*), apricot (*P. armeniaca*), and Japanese plum (*P. salicina*), may have diverged from their common ancestor, experiencing mutual hybridizations. Important introgressions commonly detected on chromosomes 6 and 8 may have been selected during the formation of the *P. mume* common ancestor. *P. mume* may have further experienced independent introgression and selection, resulting in differentiation based on geographical separation and human preference.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Principal component analysis (PCA) of Japanese and Taiwanese cultivars of *Prunus mume*.

Fig. S2 Pairwise identity by descent (IBD) proportions in *Prunus mume* cultivars.

Fig. S3 Patterns of linkage disequilibrium decay among Japanese cultivars of *Prunus mume*: fruit (F), small-fruit (FS) and ornamental (O) cultivars.

Fig. S4 Identification of selective sweeps in Chinese, Japanese, and Taiwanese cultivars of *Prunus mume* based on site frequency spectrum (SFS)-based SweeD (composite likelihood ratio, CLR) analysis.

Fig. S5 Identification of selective sweeps in fruit, small-fruit, and ornamental cultivars of Japanese *Prunus mume* based on site frequency spectrum (SFS)-based SweeD (composite likelihood ratio, CLR) analysis.

Fig. S6 Chromosomal patterns of genetic differentiation among, (a) Chinese, (b) Japanese, and (c) Taiwanese cultivars of *Prunus mume* and *P. armeniaca*.

Fig. S7 Chromosomal patterns of genetic differentiation among, (a) Chinese, (b) Japanese,

and (c) Taiwanese cultivars of *Prunus mume* and *P. salicina*.

Fig. S8 Positions of annotated genes adjacent to the candidate region (red) on chromosome 8.

Fig. S9 Neighbor-joining phylogenetic trees with the single nucleotide polymorphisms (SNPs) in (a) 15.2–15.3, in its (b) upstream and (c) downstream 1-Mb regions and (d) with the whole SNPs in chromosome 6.

Table S1 Japanese and Taiwanese cultivars and other *Prunus* species used in the present study.

Table S2 Chinese cultivars (after Zhang *et al.*, 2018) used in the present study.

Table S3 Percentage of genomic locations of single-nucleotide polymorphisms (SNPs) derived from targeted resequencing.

Table S4 Pairwise F_{ST} among *Prunus* cultivars.

Table S5 Pairwise F_{ST} among Japanese *Prunus mume* cultivars.

Table S6 The strongest candidates for selective sweep based on the nSL test for selection.

Table S7 The strongest candidates for selective sweep based on the XP-EHH test for

913 selection.

Fig. 1

(a)

(b)

Fig. 3

Fig. 4

Fig. 5

Fig. 6

Fig. 7