

Research article

Genetic basis of lineage-specific evolution of fruit traits in hexaploid persimmon

Ayano Horiuchi¹, Kanae Masuda¹, Kenta Shirasawa², Noriyuki Onoue³, Ryusuke Matsuzaki³, Ryutaro Tao⁴, Yasutaka Kubo¹, Koichiro Ushijima¹, and Takashi Akagi^{1,5,*}

¹Graduate School of Environmental and Life Science, Okayama University, Okayama 700-8530, Japan,

²Kazusa DNA Research Institute, Chiba 292-0818, Japan,

³Institute of Fruit Tree and Tea Science, NARO, Hiroshima 739-2494, Japan,

⁴Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan, and

⁵Japan Science and Technology Agency (JST), PRESTO, Saitama 332-0012, Japan

*To whom correspondence should be addressed. Tel. +81-86-251-8337. Fax. +81-86-251-8337. Email: takashia@okayama-u.ac.jp

Abstract

Frequent polyploidization events in plants have led to the establishment of many lineage-specific traits representing each species. Little is known about the genetic bases for these specific traits in polyploids, presumably due to plant genomic complexity and their difficulties in applying genetic approaches. Hexaploid Oriental persimmon (*Diospyros kaki*) has evolved specific fruit characteristics, including wide variations in fruit shapes and astringency. In this study, using whole-genome diploidized/quantitative genotypes from ddRAD-Seq data of 173 persimmon cultivars, we examined their population structures and potential correlations between their structural transitions and variations in nine fruit traits. The population structures of persimmon cultivars were highly randomized and not substantially correlated with the representative fruit traits focused on in this study, except for fruit astringency. With genome-wide association analytic tools considering polyploid alleles, we identified the loci associated with the nine fruit traits; we mainly focused on fruit-shape variations, which have been numerically characterized by principal component analysis of elliptic Fourier descriptors. The genomic regions that putatively underwent selective sweep exhibited no overlap with the loci associated with these persimmon-specific fruit traits. These insights will contribute to understanding the genetic mechanisms by which fruit traits are independently established, possibly due to polyploidization events.

Key words: fruit shape, astringency, polyploid, population structure, GWAS

1. Introduction

In contrast to animals, plant taxa have historically experienced frequent polyploidization events. Ancient paleogenome doubling (or paleo-polyploidization) often triggered the establishment of novel traits in plants, in lineage-specific manners,¹ such as water adaptation in eelgrass (*Zostera marina*).² More recent polyploidizations have also contributed to ecological adaptations such as in the genus *Fragaria*,^{3,4} and crop domestication or improvement.^{5–7} Incremental yield increases,⁸ conversion to self-pollinating systems,^{9–11} and other traits thought to be beneficial for cultivation have been reported to involve recent polyploidization.⁵ Regarding fruit characters, ancient and recent polyploidization events have also played important roles in promoting lineage-specific evolution. Fruit-ripening pathway in tomato,¹² acquisition of a specific pungent odorant in durian,¹³ and oil composition specific to olive fruits¹⁴ were thought to be triggered by lineage-specific paleo-polyploidization events. Recent polyploidization events have often been involved in fruit development, classically represented by increases in fruit size (De Langhe et al.¹⁵ for banana; Wu et al.¹⁶ for kiwifruit). However, their

genetic mechanisms are not well understood due to genomic complexity and difficulties in polyploids.⁵

Cultivated Oriental persimmon (*Diospyros kaki*), a major fruit crop in East Asia, is autohexaploid with hexasomic inheritance ($2n = 90$).¹⁷ The genus *Diospyros* also includes other polyploid species, such as ebony (*D. ebenum*, hexaploid with $2n = 90$) and American persimmon (*D. virginiana*, tetra- or hexaploid with $2n = 60, 90$),¹⁸ which are mainly utilized as timber or edible fruits. The establishment of hexaploid *D. kaki* has given rise to many types of specific fruit traits, including variation in astringency^{19,20} and diverse fruit shapes.^{21,22} Regarding astringency, the fruit of wild relatives close to *D. kaki* stably accumulates high levels of proanthocyanidins (or condensed tannins), which results in strongly astringent fruit. The loss of fruit astringency is specific to cultivated Oriental persimmon and is caused by two independent genetic factors.^{23,19} One is dependent on volatile products (alcohol and acetaldehyde) emissions from the seeds, which insolubilizes tannin polymers for reducing astringency (defined as ‘volatile dependent group’ or ‘non-pollination-constant non-astringent (non-PCNA) type’). The other factor depends on a reduction in the ability to accumulate proanthocyanidins, which is regulated

Received 8 March 2023; Revised 6 June 2023; Accepted 13 June 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Kazusa DNA Research Institute.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

by a single hypothetical locus named *ASTRINGENCY* (defined as ‘volatile independent group’ or ‘PCNA type’).¹⁹ In terms of fruit-shape variation, Oriental persimmon cultivars exhibit wide multidimensional diversity in their fruit shapes, which is not observed in close wild relatives.²¹ Persimmon fruit-shape variations have been numerically characterized by principal component analysis (fruit-shape PCA) using elliptic Fourier descriptors in SHAPE program.^{24,21} The SHAPE program was applied not only for evaluating fruit shapes but also for comparative evolutionary studies on leaf development among grape cultivars.²⁵ Potentially, this programme is useful in elucidating the unclear evolutionary process for fruit-shape variations in hexaploid persimmon combined with genetic information. The physiological mechanism that determines fruit-shape variation involves persimmon-specific regulatory pathways using fruit-shape PCA and transcriptome analysis, which are independent of genes involved in determining tomato fruit shape.²² However, the genetic basis for the acquisition of fruit-shape variations in hexaploid persimmon is unclear. These Oriental-persimmon-specific traits are distributed widely in the current populations, while little is known about the transitions of population structures, or the breeding processes linked to these specific fruit traits. Insights into these processes would contribute to the understanding of the genetic and evolutionary mechanisms, possibly associated with polyploidization events, that have led to the establishment of lineage-specific fruit traits.

Assessing the genetic architecture of traits with multiple alleles (and often mixed inheritance patterns) is challenging in polyploid species.²⁶ Inheritance modes or chromosomal behaviours in polyploids are distinct from those in diploids. Genotyping in polyploids is complicated due to the possibility of there being more than two alleles at each locus and the existence of multiple heterozygous states. In the autohexaploidy inheritance mode of Oriental persimmon,^{17,27} we can define five heterozygous states: AAAAAa (pentaplex), AAAAaa (quadruplex), AAAaaa (triplex), AAaaaa (duplex), and Aaaaaa (simplex). Recent progress in large-scale sequencing technologies and the development of analytical tools for polyploids, such as updog (for quantitative genotyping in polyploids,²⁸ StAMPP (for calculation of genetic distances with quantitative allelic states,²⁹ and GWASpoly (for genome-wide association studies (GWAS) considering allele dosages,³⁰ has allowed the study of genetics and genomics to interpret the evolution of variations in non-model polyploid crops. Using these advanced techniques, we attempted to study the genetic basis and evolutionary processes involved in the formation of Oriental-persimmon-specific fruit traits involved in recent polyploidization events, mainly from the perspective of population genetics in complicated hexaploid genomes.

2. Materials and methods

2.1. Plant materials and phenotypes examined

Genomic DNA was extracted from young leaves of 173 persimmon cultivars sampled at the experimental orchards of Kyoto University, Kyoto, Japan and the Grape and Persimmon Research Station, National Institute for Fruit and Tea Science, Hiroshima, Japan (Supplementary Table S1) using the CTAB method. For phenotypes, we targeted nine fruit traits: four fruit-shape indexes (two major principal components in²¹ and two fruit groove indexes, as described later), PCNA- and

non-PCNA-type astringency loss,¹⁹ fruit size, fruit-ripening time, and peel colour, for which hexaploid persimmon cultivars exhibit notable diversity. For fruit shapes, we applied two principal components (fruit-shape PC1 and PC4) with elliptical Fourier descriptors in the SHAPE program²⁴ to fruit longitudinal sections (Fig. 1), which had previously been characterized as the main variations among the cultivars.²¹ The fruit-shape PC1 scores were strongly correlated with the ratio of length to diameter and fruit-shape PC4 was associated with fruit flatness.²¹ We also used the presence of grooves and the measured groove depths in the fruit peripheral regions. Fruit size, ripening time, peel colour, and the two astringency types were annotated by quantitative categories according to a previous report (Fruit Tree Experiment Station of Hiroshima Prefecture, 1979) and the NIAS Genebank database (https://www.gene.affrc.go.jp/databases-plant_search_char.php). The fruit astringency in hexaploid persimmon was classified into four types based on their astringency-loss ability, depending on the presence of seeds or astringency: pollination-constant astringency (PCA), pollination-variant astringency (PVA), pollination-variant non-astringency (PVNA), and pollination-constant non-astringency (PCNA).²³ PCA and PCNA-type fruits are always astringency and non-astringency, respectively, regardless of the presence of seeds. For PCNA-type astringency loss, four astringency types were classified into two categories, including PCNA and non-PCNA (PCA, PVA, and PVNA). For non-PCNA type astringency loss, quantitative values of astringency-loss ability were modified to 0, 1, and 2 with PCA, PVA, and PVNA, respectively.¹⁹

2.2. Quantitative genotyping

ddRAD-Seq libraries were prepared with ~200 ng gDNA, as described previously.²⁷ The libraries were sequenced on an Illumina HiSeq4000 platform, and PE100 reads were generated at the Vincent J. Coates Genomics Sequencing Laboratory, University of California Berkeley. The data from 173 persimmon cultivars were mapped to closely relative species *Diospyros lotus*^{31,32} whole-chromosome pseudomolecule sequences³³ as the reference using the Burrows–Wheeler Aligner³⁴ with default settings. The resultant sam files were converted into vcf files with Samtools³⁴ and VarScan v2.3.7.³⁵ The vcf files were filtered with the following thresholds for GWAS analysis (SNP nos = 113,181): minor allele frequency (>0.01), min DP (>50), and max missing (=0.9). The files were further filtered with LD values ($r^2 < 0.3$) for population structure analyses (SNP nos = 16,956) using PLINK.³⁶ The quantitative allele composition (0–6 or nulliplex–hexaplex) was determined by using updog, a genotyping programme for polyploids²⁸ and filtered by post-probability of genotyping (propmix < 0.1).

2.3. Population structure analysis

For ADMIXTURE analysis and evolutionary tree construction, we called diploidized genotypes with a custom Python-based programme to set the hetero-tolerance to 5% (or 0/1 for 5–95% alternative alleles), as described previously.²⁷ We converted diploidized vcf files to bed files with PLINK version 1.07 (Purcell *et al.*³⁶) for use with ADMIXTURE ver. 1.3.0.³⁷ The *K* (number of ancestral populations) value was initially set from 2 to 10 to examine the most likely cluster numbers based on delta-*K* (ΔK) analysis.³⁸ Evolutionary topology was calculated by the NJ method using MEGA X³⁹;

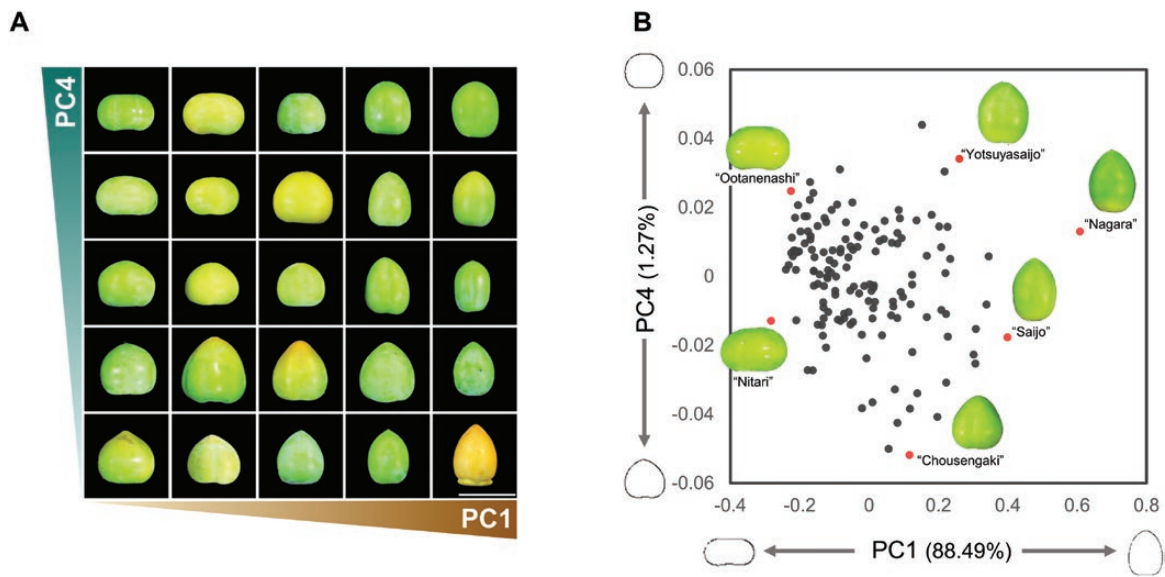


Figure 1. Fruit-shape diversity in persimmon cultivars. (A) Persimmon fruit-shape variations in the longitudinal direction, with various fruit-shape PC1/PC4 indexes. The length of the scale bar is 10 cm. (B) PC1 and PC4 in the previous principal component analysis for 153 persimmon cultivars (modified from ref. ²⁶). The percentage of variance explained by each principal component is provided in parentheses on each axis.

100 bootstrap replicates per branch); concatenated SNPs data were converted from vcf to phylip formats according to the modified method of <https://github.com/edgardomortiz/vcf2phylip>. Principal component analysis (PCA) was conducted using 16,956 SNPs with an error tolerance rate <10% with `pcaMethods` in R.⁴⁰ Pearson linear correlations were calculated between PC1-PC20 and the nine fruit traits. The distance matrix of the genetic diversity, considering the hexasomic allele state, was calculated with `StAMPP`.²⁹

2.4. Polyploidy GWAS including allele-dosage contribution

The quantitative genotype data (nulliplex–hexaplex) in 150 of 173 persimmon cultivars, of which confident/sufficient phenotypic data were found in the legacy databases, were applied to `GWASpoly` Version 2.³⁰ `GWASpoly` was performed based on a linear mixed model with population structures using the K model considering a random polygenic effect. `GWASpoly` can consider various allele dosages with two types of models (Fig. 2): (i) an additive model, in which the contribution is assumed to increase with allele dosage, and (ii) a dominant model, in which the contribution is considered binary. For a dominant model, we considered simplex–triplex allele states for the threshold (1-dom for simplex, 2-dom for duplex, and 3-dom for triplex) (Fig. 2). After filtering using `updog`, a total of 113,181 SNPs were used for GWAS analysis. The boundaries for significance were set by a Bonferroni-corrected threshold of 0.1 ($-\log(P\text{-value}) > 6.26\text{--}6.29$). Local LD values were calculated for loci surrounding the detected GWAS peaks to confine the genetic regions with potential links to the peaks.

2.5. Test for site-frequency spectrum-based selective sweeps

The selective sweep was estimated using a method based on the site-frequency spectrum (SFS) for 113,494 SNPs considering MAF and missing rates for 173 cultivars. Major methods to detect selective sweep would be categorized into SFS-based and extended haplotype-based.⁴¹ In polyploid

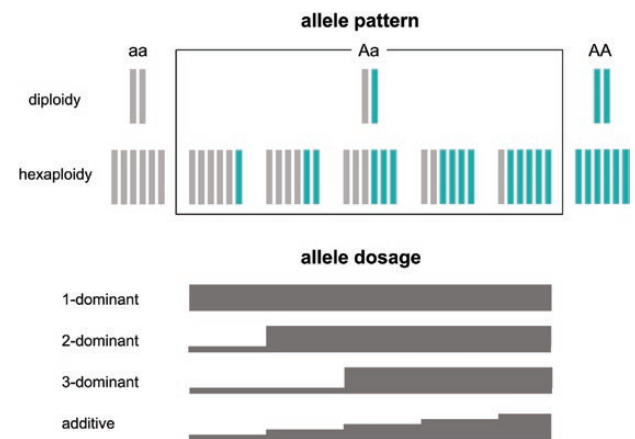


Figure 2. Definition of genotypes in hexaploidy allelic state. Five allelic states in hexaploid persimmon, AAAAAa (pentaplex), AAAAaa (quadruplex), AAAaaa (triplex), AAaaaa (duplex), and Aaaaaa (simplex), are uniformly defined as ‘heterozygous’ in diploidized genotypes. `GWASpoly` can set thresholds of allele dosages in the dominant model and determine the quantitative effects of allele dosages in the additive model.

species, detection of selective sweep would be limited with SFS-based, due to the difficulty in haplotype expectation with binary genotypic data. The SFS-based method was applied in 100-kb steps with a 400-kb sliding window using pooled heterozygosity (H_p), according to a previous report.⁴¹ The window contains only the positions of the three or fewer SNPs removed. The average expected heterozygosity was given for each window as follows:

$$H_p = \frac{1}{S} \sum_{m=1}^S \left(\frac{2N(m)_{maj}N(m)_{min}}{\{N(m)_{maj} + N(m)_{min}\}^2} \right)$$

S is the number of SNP positions in the window, and $N(m)_{maj}$ and $N(m)_{min}$ are the respective numbers of major and minor alleles at the m th SNP locus in the window.

The individual H_p values were then Z -transformed as follows: $ZH_p = (H_p - \mu H_p) / \sigma H_p$, as previously reported for pooled heterozygosity in selective sweep analysis in chicken.⁴¹

3. Results

3.1. Population structure of persimmon cultivars

We attempted to uncover population structures in hexaploid persimmon using four methods considering quantitative allelic states: principal component analysis (PCA), ADMIXTURE, evolutionary topology, and genetic-distance matrixes with StAMPP.²⁹ Both of PCA and ADMIXTURE analyses suggested weak clusterization coordinated with genotypes of astringency/non-astringency traits in fruit (Fig. 3A and B, Supplementary Fig. S1). This result is consistent with the PCNA-type (or strictly J-PCNA¹⁹) astringency-loss group originating in a very narrow genetic background, such as ‘Gosho’ and related varieties (Yamada and²⁰), and the PCNA phenotype required recessive homozygosity at the single *ASTRINGENCY* locus.¹⁹ However, in the PCA with genetic data, genome PC1 explained only 4.8% of the variance, suggesting random distribution during cultivar differentiation rather than subpopulations in peach.⁴² Genetic-distance matrixes from StAMPP analysis, considering quantitative genotypes in hexasomic allele compositions, exhibited substantially low genetic correlations among the persimmon cultivars, which was consistent with the results of PCA and supported the random differentiation of persimmon cultivars (Fig. 3C).

Correlation tests between principal component (PC) values and fruit traits exhibited a clear correlation between PCNA-type astringency loss and PC1 ($r^2 = 0.52$; Fig. 4, Supplementary Fig. S2A), which was consistent with their co-segregation in the described structural analysis using ADMIXTURE or evolutionary topology (Figs. 3 and 4). However, there were no substantial correlations ($r^2 < 0.35$) between the PC1–PC20 values and the other fruit phenotypes examined (Fig. 4). Only weak correlations ($r^2 > 0.1$) were detected between non-PCNA-type astringency loss and the genome PC2 ($r^2 = 0.32$), fruit-shape PC1 and the genome PC7 ($r^2 = 0.10$), presence of a groove and the genome PC10 ($r^2 = 0.12$), and fruit size and the genome PC1 ($r^2 = 0.12$) (Supplementary Fig. S2B–E). These results suggest that, except for PCNA-type astringency loss, notably diversified fruit traits in persimmon cultivars were independent of genome-wide structure and that there was a variety of differentiation paths.

3.2. Genomic regions associated with persimmon-specific fruit traits

We conducted GWASpoly to detect associations between genomic regions and the nine traits examined using two types of models (Fig. 2). Our results exhibited significant peaks in fruit shapes and astringency-loss traits in both the dominant and additive models (Fig. 5).

For the fruit-shape-related traits, we detected significant associations with fruit-shape PC1, fruit-shape PC4, and depth of groove, although they are independent of each other. This suggested that persimmon fruit-shape variation is complex and determined by multiple genetic factors. For fruit-shape PC1, the highest peak appeared on chromosome 2 in both the additive and 1-dominant models (Fig. 5A). Cultivars with narrow fruit shapes (fruit-shape PC1 > 0.30), except

‘Fudegaki’ (fruit-shape PC1 = 0.40), were heterozygous for the genotype at these loci (Supplementary Fig. S3A). A previous physiological assessment suggested that the narrow (or high fruit-shape PC1) fruit shape in ‘Fudegaki’ is due to a unique mechanism: a reduction in the number of locules.²¹ The fruit-shape PC1 value tended to be increased in proportion to allele dosage (Supplementary Fig. S3A), suggesting that the responsible gene (or mutation) can act additively. Linkage disequilibrium was maintained over 1.6 Mbp in the region surrounding the chromosome 2 peak (Supplementary Fig. S4A). Chromosome 12 exhibited the second highest peak associated with fruit-shape PC1, although this was slightly less significant in the Bonferroni threshold analysis (Fig. 5A). LD was potentially conserved over 3 Mbp surrounding the peak (Supplementary Fig. S4B). For fruit-shape PC4, which is thought to reflect a blossom end shape diversity,²¹ a peak was observed on chromosome 3 but only in the dominant model (Fig. 5B). For groove depth, significant peaks appeared on chromosomes 3 and 8 in both the dominant and additive models (Fig. 5D), although the peak on chromosome 3 did not overlap the peak associated with fruit-shape PC4 (Fig. 5B). Chromosomes 2 and 9 also exhibited significant peaks in the additive and dominant models, respectively (Fig. 5D). Although the peak on chromosome 2 for groove depth (Chr2:490,014) was located near a peak for fruit-shape PC1 (Chr2: 1,565,911), they were 1 Mbp apart and exhibited no correlation in genotypes ($r = 0.07$).

For the two types of astringency-loss traits, the association test for non-PCNA-type astringency loss exhibited a very clear peak on chromosome 15 in both the additive and dominant models (Fig. 5F). A high portion of individuals with astringency loss (or PVNA/PVA traits) tended to be heterozygous for the peak (Supplementary Fig. S5), suggesting that astringency loss is a dominant mutation that establishes a new function: the production of ethanol in seeds to insolubilize high levels of proanthocyanidins in fruit flesh. This is also supported because close wild relatives of hexaploid *D. kaki*, *D. lotus*, and *D. oleifera* cannot insolubilize proanthocyanidins in fruit. The association test for PCNA-type astringency loss exhibited multiple small peaks scattered throughout the genomes in both the additive and dominant models. Of them, the most promising peak would be on chromosome 5, since an *ASTRINGENCY* locus-linked genomic marker overlapped on the peak in chromosome 5 (*ca.* 29Mb, referring⁴³). The lack of a single, clear peak on chromosome 5 might be due to three possible reasons, (i) the clusterized population structure in the PCNA cultivars, as shown in Fig. 3, (ii) high diversities of functional *ASTRINGENCY* alleles,¹⁷ and (iii) potential multiple haplotypes of the mutated *ASTRINGENCY* alleles (or *ast*) among the cultivars.¹⁷⁴⁴

3.3. Selective sweep in persimmon cultivar differentiation

SFS-based detection of selective sweep identified statistically significant multiple peaks on chromosomes 3, 4, and 9 ($P < 0.01$, Fig. 6). None of them overlapped the peak regions detected by GWAS. This suggests that the loci responsible for hexaploid persimmon-specific fruit traits were not established via strong selections (or bottlenecks) on recent mutations but via natural population transitions. It is worth noting that perennial or tree crops often maintain favoured alleles heterozygously during clonal propagation.⁴⁴ Extended haplotype homozygosity (EHH)-based methods^{45,46} would better fit

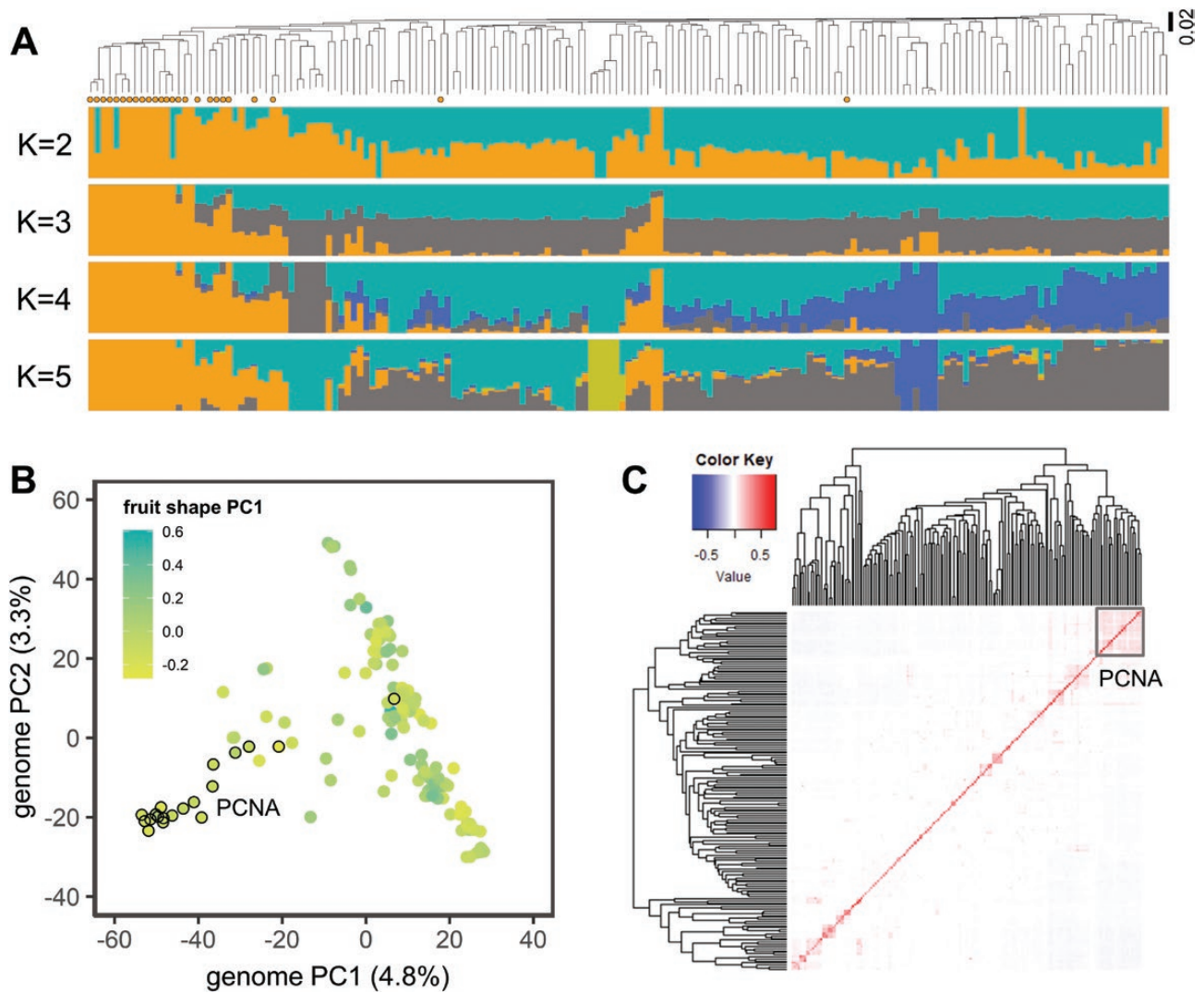


Figure 3. Population structures of persimmon cultivars. (A) Phylogenetic analysis and proportion of ancestry from $K = 2-5$ inferred using ADMIXTURE in 173 hexaploid persimmon cultivars. The orange circles indicated PCNA cultivars. PCNA-type astringency-loss varieties mostly formed a separate cluster. The cultivar names were listed in [Supplementary Table S1](#). (B) Distribution of the first two principal components (genome PC1 and PC2) in the PCA. Fruit-shape PC1 values²¹ are shown with a colour gradient. No substantial correlation was detected between population structure and fruit-shape PC1. (C) Genetic distance matrix in 173 hexaploid persimmon cultivars created with StAMPP using quantitative genotypes.

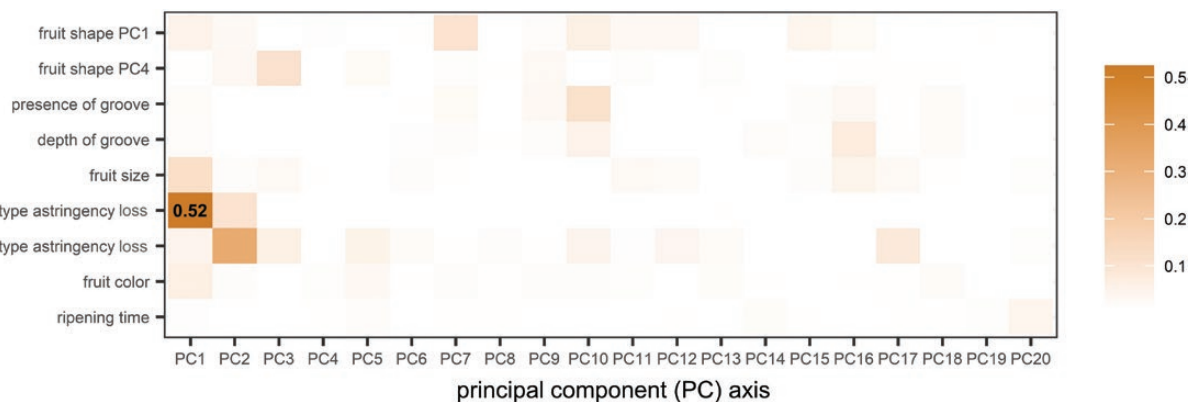


Figure 4. Correlation between population structures and each fruit trait. Pearson's product-moment correlation (r^2 values) between persimmon fruit traits and the principal components (genome PC1–PC20) of genome-wide SNPs. The PCNA-type astringency-loss trait was substantially correlated with the first principal component, which is consistent with empirical knowledge about persimmon breeding.²⁰

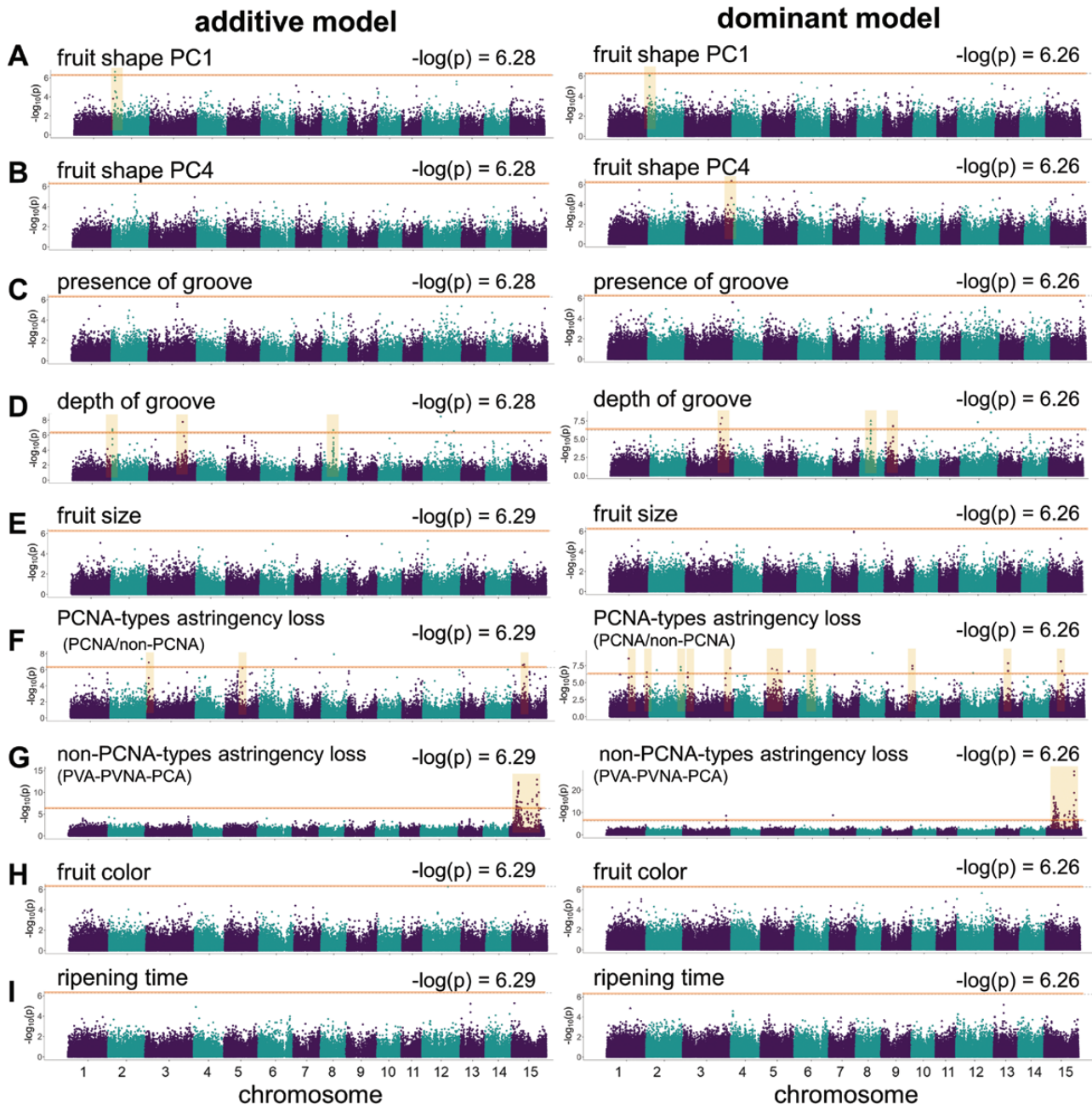


Figure 5. GWAS for representative fruit traits in hexaploid persimmon cultivars. GWAS with additive (left) and dominant (right) models for nine fruit traits in hexaploid persimmon cultivars: fruit-shape PC1 and PC4 values (A and B), presence of groove (C), depth of groove (D), fruit size (E), PCNA-type astringency loss (F), non-PCNA-type astringency loss (G), fruit colour (H), and ripening time (I). Potential peak regions are highlighted in pale orange. The thick orange lines indicate $-\log(P\text{-value}) > 6.26\text{--}6.29$, which represents the Bonferroni-corrected threshold of 0.1.

this situation than SFS-based methods. However, it would be difficult to properly define haploblocks (or LD decay), as indicated in the method section, in autohexaploid persimmon using only binary SNPs. Thus, further analyses would be required to detect selective sweeps involving persimmon-specific fruit traits.

4. Discussion

4.1. Fruit-shape diversification after hexaploidization

Compared with hexaploid persimmon cultivars, its close diploid wild (or semi-cultivated) relatives, including *D.*

oleifera and *D. lotus*, exhibit much less morphological diversity in fruits.⁴⁷⁴⁸²¹ This leads to two hypotheses about how fruit-shape diversification was triggered: (i) in pathways arising from domestication (or cultivation) or (ii) via hexaploidization. Our results indicated that the population structure of hexaploid persimmon cultivars is little clustered, except PCNA group (Fig. 3), and there were no principal component axes significantly correlated with the fruit-shape indexes except for fruit astringency traits (Fig. 4). These findings suggested that most of the persimmon cultivars were randomly evolved rather than forming subpopulations (or subclusters), where fruit shapes were also evolved independently; this contrasts with the

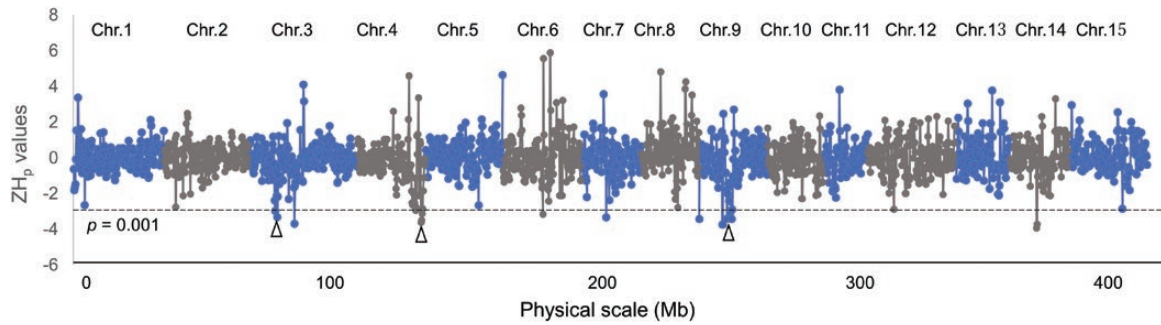


Figure 6. Detection of potential selective sweeps based on a site-frequency spectrum index. Transitions of standardized log-transformed values of the pooled heterozygosity (ZH_p^{41}) in 400-kb bins with 200-kb steps are given. The dotted line indicates the threshold, with $p < 0.001$. The peaks exhibiting a gradual reduction in ZH_p , which was lower than the threshold, are marked with white arrowheads as potential sweep regions.

morphological evolution of grapevine leaves, which is correlated with the transition of population structures.²⁵ On the other hand, there have been no direct clues or molecular mechanism to explain the hypothesis (ii). Our results so far suggested that the fruit-shape variations of hexaploid persimmon are likely to be controlled by multiple independent genes. Our GWAS analysis detected a locus related to fruit-shape PC1 on chromosome 2 (Fig. 5A), whereas this locus explained only some of the varieties with oblong fruit shapes. In other words, this variety would have lineage-specific variations. Thus, at least for this case, hexaploidization might not be a direct effect but satisfy potential requirements for the current fruit-shape diversity, while local selections of different fruit shapes might contribute more to the current diversity. Furthermore, other fruit-shape traits, such as the depth of lateral grooves, were associated with chromosome 8, which was independent of fruit-shape PC1 (Fig. 5A–D). Together, these results imply that various genetic factors have accumulated multidimensionally to determine fruit shape among a wide variety of hexaploid persimmon cultivars.

Several examples of the facilitation of phenotypic variations via polyploidization have been reported; for example, studies of *Fragaria* have indicated that polyploids show greater environmental adaptation than diploids.⁴ The results in *Fragaria* support the ‘jack-and-master’ strategy, where non-directional phenotypic changes appear via polyploidization with the potential to be used in adaptation to new environments. *Allium* has also shown higher speciation rates with more trait diversification via intraspecific polyploidy than in diploid or low-ploid populations.⁴⁹ Hexaploid persimmon cultivars could be another example of this phenomenon. Considering the phenotypic variations in fruit shape, which are not expressed in close diploids and not correlated with paths of variety differentiation in hexaploids, the ‘jack-and-master’ hypothesis may also be applicable to persimmon, although further experimental validation is necessary. However, the significance of fruit-shape variations in adaptation may be difficult to define, in contrast with other fruit traits, such as astringency traits (or accumulation of second metabolites) or fruit size. Although the SFS-based selective sweep did not overlap with the fruit-shape-associated loci in this study, artificial selections (or bottlenecks) via domestication (or local selections of specific varieties) may be partially involved in the development of the current fruit-shape variations, in addition to the effects of hexaploidization.

4.2. Identification of candidate genes determining fruit shape in persimmon

As described, the fruit-shape variations in hexaploid persimmon are thought to be regulated by multiple genetic factors and to have evolved independently.²² Our GWAS identified the highest peaks on chromosome 2 in both the dominant and additive models (Fig. 5A), and only ‘Saijo’ and its close varieties with high fruit-shape PC1 values showed heterozygous genotypes (Supplementary Fig. S3). Twenty-nine genes were located within 1.6 Mbp near the peak of chromosome 2, including the gene encoding auxin-inducible in root cultures-like protein, which is involved in auxin production in roots,⁵⁰ and that encoding the HAUS augmin-like complex subunit, which is involved in the formation of spindle threads during cell division.^{51,52} Since the fruit-shape variation in persimmon cultivars depends on diversity in cell division (or proliferation) in the mesocarp,²¹ these genes may be good candidates as determinants of the specificity of fruit shape in the ‘Saijo’ variety. Chromosome 12 also exhibited an association peak for fruit-shape PC1, independent of the peak on chromosome 2 (Fig. 5A). On chromosome 12, 244 genes, including *CYCD*, which is involved in the cell cycle, were localized to regions surrounding the peaks with significant LD values. Our results successfully identified the complicated genetic bases for fruit-shape variations in hexaploid persimmon. Additional fine-expression profiling covering the entire process of fruit development in specific genotypes identified in this study will narrow down the candidate genes. It is also worth noting that this study referred to the genome information of a close diploid relative, *D. lotus*, which shows no substantial fruit-shape variations and has no genes/mutations corresponding the fruit-shape diversity observed in hexaploid *D. kaki* cultivars. Hence, it would be much preferable to use the genome information of the hexaploid *D. kaki* itself when discussing the possibility of fruit-shape diversification through hexaploidization.

Acknowledgements

We thank Candace Webb, Ph.D., from Edanz (<https://jp.edanz.com/ac>) for editing a draft of this manuscript. This work was supported by PRESTO from Japan Science and Technology Agency (JST) [JPMJPR20D1] to T.A. and Grant-in-Aid for Scientific Research (B) from JSPS [22H02339] to T.A. and for Transformative Research Areas (A) from JSPS [22H05172 and 22H05173] to T.A., and [22H05181] to K.S.

Author contributions

T.A. conceived the study. A.H., K.M., and T.A. designed experiments. A.H., K.M., and K.S. conducted the experiments. A.H., K.M., and T.A. analysed the data. N.O., R.M., R.T., Y.K., K.U., and T.A. contributed to plant resources and facilities. A.H., and T.A. drafted the manuscript. All authors approved the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Data availability

The sequence reads are available from the DDBJ Sequence Read Archive (DRA) under the accession number DRA015334.

Supplementary Data

Supplementary data are available at DNARES online.

References

1. Van de Peer, Y., Mizrachi, E., and Marchal, K. 2017, The evolutionary significance of polyploidy, *Nat. Rev. Genet.*, **18**, 411–424.
2. Olsen, J.L., Rouzé, P., Verhelst, B., et al. 2016, The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea, *Nature*, **530**, 331–335.
3. Liston, A., Cronn, R., and Ashman, T.L. 2014, *Fragaria*: a genus with deep historical roots and ripe for evolutionary and ecological insights, *Am. J. Bot.*, **101**, 1686–1699.
4. Wei, N., Cronn, R., Liston, A., and Ashman, T.L. 2019, Functional trait divergence and trait plasticity confer polyploid advantage in heterogeneous environments, *New Phytol.*, **221**, 2286–2297.
5. Akagi, T., Jung, K., Masuda, K., and Shimizu, K.K. 2022, Polyploidy before and after domestication of crop species, *Curr. Opin Plant Biol.*, **69**, 102255.
6. Salman-Minkov, A., Sabath, N., and Mayrose, I. 2016, Whole-genome duplication as a key factor in crop domestication, *Nat. Plants*, **2**, 1–4.
7. Heslop-Harrison, J.S., and Schwarzacher, T. 2007, Domestication, genomics and the future for banana, *Ann. Bot.*, **100**, 1073–1084.
8. Hauck, N.R., Yamane, H., Tao, R., and Iezzoni, A.F. 2006, Accumulation of nonfunctional S-haplotypes results in the breakdown of gametophytic self-incompatibility in tetraploid *Prunus*, *Genetics*, **172**, 1191–1198.
9. Muñoz-Sanz, J.V., Zuriaga, E., Cruz-García, F., McClure, B., and Romero, C. 2020, Self-(in) compatibility systems: target traits for crop-production, plant breeding, and biotechnology, *Front. Plant Sci.*, **11**, 195.
10. Masuda, K., Ikeda, Y., Matsuura, T., et al. 2022, Reinvention of hermaphroditism via activation of a *RADIALIS*-like gene in hexaploid persimmon, *Nat. Plants*, **8**, 217–224.
11. Tomato Genome Consortium, 2012, The tomato genome sequence provides insights into fleshy fruit evolution, *Nature*, **485**, 635–641.
12. Teh, B.T., Lim, K., Yong, C.H., et al. 2017, The draft genome of tropical fruit durian (*Durio zibethinus*). *Nat. Genet.*, **49**, 1633–1641.
13. Unver, T., Wu, Z., Sterck, L., et al. 2017, Genome of wild olive and the evolution of oil biosynthesis. *Proc. Nat. Acad. Sci. USA*, **114**, 9413–9422.
14. De Langhe, E., Vrydaghs, L., De Maret, P., Perrier, X., Denham, T. 2009, Why bananas matter: an introduction to the history of banana domestication, *Ethnobot. Res. Appl.*, **7**, 165–177.
15. Wu, R.M., Walton, E.F., Richardson, A.C., Wood, M., Hellens, R.P., and Varkonyi-Gasic, E. 2012, Conservation and divergence of four kiwifruit *SVP*-like MADS-box genes suggest distinct roles in kiwifruit bud dormancy and flowering, *J. Exp. Bot.*, **63**, 797–807.
16. Akagi, T., Tao, R., Tsujimoto, T., Kono, A., and Yonemori, K. 2012, Fine genotyping of a highly polymorphic *ASTRINGENCY*-linked locus reveals variable hexasomic inheritance in persimmon (*Diospyros kaki* Thunb.) cultivars, *Tree Genet. Genomes*, **8**, 195–204.
17. Tamura, M., Tao, R., Yonemori, K., Utsunomiya, N., and Sugiura, A. 1998, Ploidy level and genome size of several *Diospyros* species, *J. Japan. Soc. Hort. Sci.*, **67**, 306–312.
18. Akagi, T., Katayama-Ikegami, A., and Yonemori, K. 2011, Proanthocyanidin biosynthesis of persimmon (*Diospyros kaki* Thunb.) fruit, *Sci. Hortic.*, **130**, 373–380.
19. Sato, A., and Yamada, M. 2016, Persimmon breeding in Japan for pollination-constant non-astringent (PCNA) type with marker-assisted selection, *Breed. Sci.*, **66**, 60–68.
20. Maeda, H., Akagi, T., and Tao, R. 2018, Quantitative characterization of fruit shape and its differentiation pattern in diverse persimmon (*Diospyros kaki*) cultivars, *Sci. Hortic.*, **228**, 41–48.
21. Maeda, H., Akagi, T., Onoue, N., Kono, A., and Tao, R. 2019, Evolution of lineage-specific gene networks underlying the considerable fruit shape diversity in persimmon, *Plant Cell Physiol.*, **60**, 2464–2477.
22. Sugiura, A., and Tomana, T. 1983, Relationships of ethanol production by seeds of different types of Japanese persimmons and their tannin content, *HortScience*, **18**, 319–321.
23. Iwata, H., and Ukai, Y. 2002, SHAPE: a computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors, *J. Hered.*, **93**, 384–385.
24. Chitwood, D.H., Ranjan, A., Martinez, C.C., et al. 2014, A modern ampelography: a genetic basis for leaf shape and venation patterning in grape, *Plant Physiol.*, **164**, 259–272.
25. Dufresne, F., Stift, M., Vergilino, R., and Mable, B.K. 2014, Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools, *Mol. Ecol.*, **23**, 40–69.
26. Masuda, K., Yamamoto, E., Shirasawa, K., et al. Genome-wide study on the polysomic genetic factors conferring plasticity of flower sexuality in hexaploid persimmon. *DNA Res.*, **27**, dsaa012.
27. Gerard, D., Ferrão, L.F.V., Garcia, A.A.F., and Stephens, M. 2018, Genotyping polyploids from messy sequencing data, *Genetics*, **210**, 789–807.
28. Pembleton, L.W., Cogan, N.O., and Forster, J.W. 2013, StAMPP: An R package for calculation of genetic differentiation and structure of mixed - ploidy level populations, *Mol. Ecol. Resour.*, **13**, 946–952.
29. Rosyara, U.R., De Jong, W.S., Douches, D.S., and Endelman, J.B. 2016, Software for genome - wide association studies in autopolyploids and its application to potato, *Plant Genome*, **9**, 08.
30. Yonemori, K., Honsho, C., Kanzaki, S., Ino, H., Ikegami, A., Kitajima, A., Sugiura, A., and Parfitt, D. E. 2008, Sequence analyses of the ITS regions and the *mat K* gene for determining phylogenetic relationships of *Diospyros kaki* (persimmon) with other wild *Diospyros* (Ebenaceae) species. *Tree Genet. Genom.*, **4**: 149–158.
31. Nishiyama, S., Onoue, N., Kono, A., Sato, A., Ushijima, K., Yamane, H., Tao, R., and Yonemori, K. 2018, Comparative mapping of the *ASTRINGENCY* locus controlling fruit astringency in hexaploid persimmon (*Diospyros kaki* Thunb.) with the diploid *D. lotus* reference genome. *Hortic. J.*, **87**: 315–323.
32. Akagi, T., Shirasawa, K., Nagasaki, H., et al. 2020, The persimmon genome reveals clues to the evolution of a lineage-specific sex determination system in plants, *PLoS Genet.*, **16**, e1008566.
33. Li, H., and Durbin, R. 2009, Fast and accurate short read alignment with Burrows–Wheeler transform, *Bioinformatics*, **25**, 1754–1760.
34. Koboldt, D.C., Zhang, Q., Larson, D.E., et al. 2012, VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing, *Genome Res.*, **22**, 568–576.
35. Purcell, S., Neale, B., Todd-Brown, K., et al. 2007, PLINK: a tool set for whole-genome association and population-based linkage analyses, *Am. J. Hum. Genet.*, **81**, 559–575.

36. Alexander, DH., Novembre, J., Lange, K., 2009, Fast model-based estimation of ancestry in unrelated individuals, *Genome Res.*, **19**, 1655–1664.
37. Evanno, G., Regnaut, S., and Goudet, J. 2005, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, *Mol. Ecol.*, **14**, 2611–2620.
38. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018, MEGA X: molecular evolutionary genetics analysis across computing platforms, *Mol. Biol. Evol.*, **35**, 1547.
39. Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. 2007, pcaMethods—a bioconductor package providing PCA methods for incomplete data, *Bioinformatics*, **23**, 1164–1167.
40. Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A. G. 2007. Recent and ongoing selection in the human genome. *Nat. Rev. Genet.*, **8**, 857–868.
41. Rubin, CJ., Zody, MC., Eriksson, J., et al. 2010, Whole-genome resequencing reveals loci under selection during chicken domestication, *Nature*, **464**, 587–591.
42. Akagi, T., Hanada, T., Yaegaki, H., Gradziel, T.M., and Tao, R. 2016, Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication. *DNA Res.*, **23**, 271–282.
43. Akagi, T., Takeda, Y., Yonemori, K., et al. 2010, Quantitative genotyping for the astringency locus in hexaploid persimmon cultivars using quantitative real-time PCR. *J. Am. Soc. Hortic.*, **135**, 59–66.
44. Kono, A., Kobayashi, S., Onoue, N., and Sato, A. 2016, Characterization of a highly polymorphic region closely linked to the *AST* locus and its potential use in breeding of hexaploid persimmon (*Diospyros kaki* Thunb.), *Mol. Breed.*, **36**, 1–13.
45. Numaguchi, K., Akagi, T., Kitamura, Y., Ishikawa, R., and Ishii, T. 2020, Interspecific introgression and natural selection in the evolution of Japanese apricot (*Prunus mume*), *Plant J.*, **104**, 1551–1567.
46. Sabeti, PC., Varilly, P., Fry, B., et al. 2007, Genome-wide detection and characterization of positive selection in human populations, *Nature*, **449**, 913–918.
47. Sabeti, PC., Reich, DE., Higgins, JM., et al. 2002, Detecting recent positive selection in the human genome from haplotype structure, *Nature*, **419**, 832–837.
48. Utsunomiya, N., Subhadrabandhu, S., Yonemori, K., et al. 1998, *Diospyros* species in Thailand: their distribution, fruit morphology and uses, *Econ. Bot.*, **52**, 343–351.
49. Yang, Y., Jing, ZB., Ruan, XF., and Cheng, JM. 2015, Development of simple sequence repeat markers in persimmon (*Diospyros L.*) and their potential use in related species, *Genet. Mol. Res.*, **14**, 609.
50. Han, TS., Zheng, QJ., Onstein, RE., et al. 2020, Polyploidy promotes species diversification of *Allium* through ecological shifts, *New Phytol.*, **225**, 571–583.
51. Gibson, SW., and Todd, CD. 2015, Arabidopsis *AIR12* influences root development. *Physiol. Mol. Biol. Plants*, **21**, 479–489.
52. Hotta, T., Kong, Z., Ho, CMK., et al. 2012, Characterization of the *Arabidopsis* Augmin complex uncovers its critical function in the assembly of the acentrosomal spindle and phragmoplast microtubule arrays, *Plant Cell*, **24**, 1494–1509.