

## Research Article

# Propagation path of a flowering cherry (*Cerasus × yedoensis*) cultivar ‘Somei-Yoshino’ traced by somatic mutations

Kenta Shirasawa<sup>1,\*</sup>, Tomoya Esumi<sup>2</sup>, Akihiro Itai<sup>3</sup>, Katsunori Hatakeyama<sup>4</sup>, Tadashi Takashina<sup>5</sup>, Takuji Yakuwa<sup>6</sup>, Katsuhiko Sumitomo<sup>7</sup>, Takeshi Kurokura<sup>8</sup>, Eigo Fukai<sup>9</sup>, Keiichi Sato<sup>10</sup>, Takehiko Shimada<sup>11</sup>, Katsuhiko Shiratake<sup>12</sup>, Munetaka Hosokawa<sup>13</sup>, Yuki Monden<sup>14</sup>, Makoto Kusaba<sup>15</sup>, Hidetoshi Ikegami<sup>16</sup>, and Sachiko Isobe<sup>1</sup>

<sup>1</sup>Department of Frontier Research and Development, Kazusa DNA Research Institute, Kisarazu, Japan

<sup>2</sup>Academic Assembly Institute of Agricultural and Life Sciences, Shimane University, Matsue, Japan

<sup>3</sup>Department of Agricultural and Life Science, Kyoto Prefectural University, Kyoto, Japan

<sup>4</sup>Faculty of Agriculture, Iwate University, Morioka, Japan

<sup>5</sup>Horticultural Research Institute, Yamagata Integrated Agricultural Research Center, Sagae, Japan

<sup>6</sup>Yamagata Nishi High School, Yamagata, Japan

<sup>7</sup>Institute of Vegetable and Floriculture Science, NARO, Tsukuba, Japan

<sup>8</sup>Faculty of Agriculture, Utsunomiya University, Utsunomiya, Japan

<sup>9</sup>Graduate School of Science and Technology, Niigata University, Niigata, Japan

<sup>10</sup>Yamanashi Kofu Minami High School, Kofu, Japan

<sup>11</sup>Institute of Fruit Tree and Tea Science, NARO, Shizuoka, Japan

<sup>12</sup>Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan

<sup>13</sup>Department of Agricultural Science, Kindai University, Nara, Japan

<sup>14</sup>Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan

<sup>15</sup>Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Japan

<sup>16</sup>Resident, Fukuoka, Japan

\*To whom correspondence should be addressed: Tel.: +81 438 52 3935. Fax: +81 438 52 3934. E-mail: [shirasaw@kazusa.or.jp](mailto:shirasaw@kazusa.or.jp)

## Abstract

In the long history of human relations with flowering cherry trees in Japan, ‘Somei-Yoshino’ occupies an exceptional position among a variety of flowering trees: it is a self-incompatible interspecific hybrid but has been enthusiastically planted by grafting throughout Japan, due most likely to its flamboyant appearance upon full bloom. Thus, ‘Somei-Yoshino’ gives us a rare opportunity to trace and investigate the occurrence and distribution of somatic mutations within a single plant species through analysis of the genomes of the clonally propagated trees grown under a variety of geographical and artificial environments. In the studies presented here, a total of 46 samples of ‘Somei-Yoshino’ trees were collected and their genomes were analysed. We identified 684 single nucleotide mutations, of which 71 were present in more than two samples. Clustering analysis of the mutations indicated that the 46 samples were classified into eight groups, four of which included 36 of the 46 samples analysed. Interestingly, all the four tree samples collected in Ueno Park of Tokyo were members of the four groups mentioned above. Based on comparative analysis of their mutations, one of the four trees growing in Ueno Park was concluded to be the closest to the original ancestor. We propose that somatic mutations may be used as tracers to establish the ancestral relationship amongst clonally propagated individuals.

**Key words:** clone, flowering cherry, genome sequence, somatic mutation, Somei-Yoshino

## Introduction

More than 200 natural and domesticated cultivars of flowering cherry trees are said to exist in Japan. Of them, the most popular cultivar is named ‘Somei-Yoshino’, which is a self-incompatible interspecific hybrid between *Cerasus spachiana* and *C. speciosa*<sup>1,2</sup> produced some 200 years ago.<sup>3</sup> In the long history of human relations with these flowering

cherry trees in Japan, ‘Somei-Yoshino’ is known to have frequently been chosen to be planted in parks, gardens, river banks, and various other places, largely because of its flowering characteristics: it may often be observed that almost an entire tree of ‘Somei-Yoshino’ is covered by thousands of tiny white flowers at the time of full bloom as if they inform visitors of the arrival of spring. The south to north migrating

Received 21 February 2024; Revised 1 August 2024; Accepted 6 August 2024

© The Author(s) 2024. 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 License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

zone of blooming ‘Somei-Yoshino’ trees along the Japanese archipelago is often referred to as ‘the flowering cherry front’ which is daily forecasted during the season.<sup>4,5</sup> Thus, ‘Somei-Yoshino’ is a best-known example of flowering trees favoured by people for planting almost everywhere throughout Japan.

‘Somei-Yoshino’ possesses a highly heterozygous genome, partly because it is an interspecific hybrid.<sup>1,2</sup> As it has self-incompatibility as well as the high-heterozygous genome due to the interspecific hybrid, ‘Somei-Yoshino’ has been propagated mainly by grafting.<sup>6</sup> Whether ‘Somei-Yoshino’ emerged naturally or it was artificially created remains unknown except that its origin was somehow related to a village named Somei in Tokyo and its existence was known at the end of the Edo period some 150 years ago.<sup>3</sup> Studies suggest that the original tree of ‘Somei-Yoshino’ was planted either in Koishikawa Botanical Garden (Tokyo, Japan),<sup>7</sup> Ueno Park (Tokyo, Japan)<sup>8</sup>, or Kaiseizan Park (Fukushima, Japan).<sup>9</sup> Despite the note of Nakamura *et al.*<sup>8</sup> who, based on the sequence analysis of *PolA1* gene, suggested that there are sufficient genetic resources to create ‘Somei-Yoshino’ in Ueno Park, details are still unknown. The origin of ‘Somei-Yoshino’ thus remains an interesting issue for researchers as well as people who are fascinated with ‘Somei-Yoshino’ and other cherry species.

Somatic mutations sometimes occur in different organs of an organism and are rarely eliminated, leading to chimerism.<sup>10–12</sup> Some of the mutations in genes affect their functions, resulting in phenotypic alterations.<sup>13</sup> Bud sports may be found in branches of trees as a result of somatic mutations.<sup>13,14</sup> Since bud sports are genetically stable, grafting or cutting is used to propagate the mutants into new cultivars.<sup>14</sup> Furthermore, silent somatic mutations, which do not affect gene functions nor cause phenotypic variations, also exist and are stably inherited. Therefore, somatic mutations may be used to trace the history of a clonally propagated plants such as ‘Somei-Yoshino’.

Considering these, we thought that ‘Somei-Yoshino’ would give us a rare opportunity to trace and investigate the occurrence and distribution of somatic mutations through analysis of the genomes of the clonally propagated ‘Somei-Yoshino’ trees planted in different environments under a variety of geographical and artificial conditions such as heavy snow- and/or rain-fall, different ambient temperatures and humidities, interactions with other plants, birds, animals, insects etc. The clonality of ‘Somei-Yoshino’ has been investigated by DNA fingerprinting using microsatellite markers.<sup>1,6</sup> In addition, somatic mutations within a single ‘Somei-Yoshino’ tree were studied using techniques such as temperature gradient gel electrophoresis (TGGE)<sup>15</sup> and double-digest restriction site-associated DNA sequencing (ddRAD-Seq).<sup>16</sup> These studies indicated that the genetic identity of ‘Somei-Yoshino’ is high, although quite a few somatic mutations were detected. The genome ( $2n = 16$ ) of ‘Somei-Yoshino’ (Tree ID #136 in Ueno Park) has been sequenced at the haplotype-phased chromosome level and 95,076 genes were identified in the 16 chromosome sequences spanning 690.1 Mb in total.<sup>2</sup> Here, we report the whole-genome sequence analysis of somatic mutations in the 46 ‘Somei-Yoshino’ tree samples to characterize and establish their mutual relations.

## Materials and methods

### Plant materials

Leaves were collected from 46 ‘Somei-Yoshino’ trees grown in 19 prefectures of Japan (Supplementary Table S1); these 46

trees included the four trees, tree IDs of #133, #134, #136, and #138 planted in Ueno Park, Tokyo, Japan, which geographical positions are available in Nakamura *et al.*<sup>8</sup> Genomic DNA was extracted from the leaves using the FavorPrep Plant Genomic DNA Extraction Mini Kit (Favorgen, Ping-Tung, Taiwan).

### DNA sequencing

Genomic DNA libraries were prepared with a PCR-free method using the Swift 2S Turbo Flexible DNA Library Kit (Swift Biosciences, Ann Arbor, MI, USA), and converted into a DNA nanoball sequencing library with the MGI Easy Universal Library Conversion Kit (MGI Tech, Shenzhen, China). The library was sequenced on the DNBSEQ G400RS (MGI Tech) instrument in paired-end, 150 bp mode. The sequence data of tree #136, with accession numbers DRR169775 (Sample name of SyTKY0 in this study) and DRR169776 (SyTKY1), were obtained from a DNA database.

### Detection and analysis of somatic mutations

Low-quality bases (quality score < 10) and adaptor sequences (AGATCGGAAGAGC) were trimmed with PRINSEQ (version 0.20.4)<sup>17</sup> and *fastx\_clipper*, respectively, in FASTX-Toolkit (version 0.0.14) ([https://github.com/agordon/fastx\\_toolkit](https://github.com/agordon/fastx_toolkit)), and the remaining high-quality reads were mapped on to the haplotype-resolved chromosome-level genome sequence of ‘Somei-Yoshino’ (CYE\_r3.1.pseudomolecule, obtained from DBcherry: <https://cherry.kazusa.or.jp>)<sup>2</sup> with Bowtie 2 (version 2.3.5.1).<sup>18</sup> The resultant SAM files were converted into BAM files with the *view* command of SAMtools (version 0.1.19).<sup>19</sup> The gVCF files were generated from the BAM files using the *mpileup* command (-Ou -a DP,AD,INFO/AD) and *call* command (-Oz -m -g 0,10) in BCFtools (version 1.9),<sup>19</sup> and normalized using the *norm* command of BCFtools. The individual gVCF files were merged into a single VCF file using the *merge* command of BCFtools. High-confidence, biallelic, homozygous single nucleotide variants (SNVs) were selected with VCFtools (version 0.1.12b).<sup>20</sup> Genome positions of the SNVs were visualized with MapChart (version 2.2).<sup>21</sup> Copy number variations (CNVs) were detected with *cnv-seq* (version 0.2.7),<sup>22</sup> and the effect of sequence variations on gene function was predicted with SNPeff (version 4.2).<sup>23</sup>

### Cultivar identification

To identify cultivars of the tested 46 samples, the ddRAD-Seq data of 139 lines were downloaded from a DNA database (GenBank accession numbers: DRR169804–DRR169942). High-quality reads, selected as described above, were mapped on to the genome sequence of ‘Somei-Yoshino’ (CYE\_r3.1.pseudomolecule), and SNVs were detected as described previously.<sup>2</sup> SNVs identified in the 46 samples, based on the analysis of ddRAD-Seq reads and gVCF files, were combined to be subject to a principal component analysis (PCA). A phylogenetic tree based on 100 bootstrap replicates was created with SNPhylo (version 20140701),<sup>24</sup> in which *Padus grayana* Uwamizu-zakura (Cerasus\_2-58) was employed as an outgroup, and visualized with iTOL (version 6.9.1).<sup>25</sup> To estimate the optimal number of clusters, the cumulative explained variance was calculated with Tassel 5.<sup>26</sup>

### SNV analysis of ancestral *Cerasus*

To identify the possible ancestral alleles of SNVs, whole-genome sequence reads were obtained from a public DNA

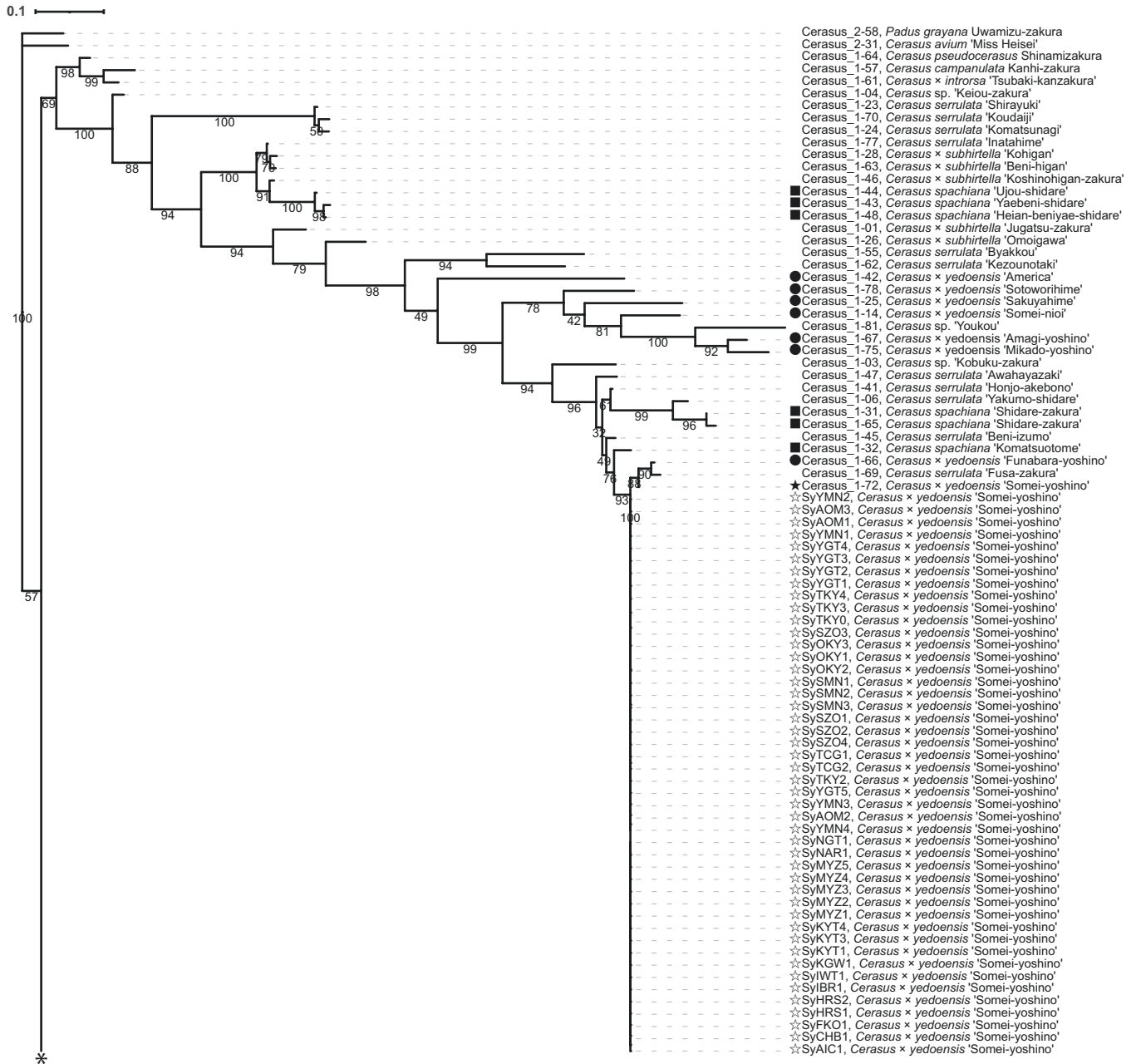
database (GenBank accession numbers: DRR169795–DRR169803 and SRR6957274)<sup>2</sup> for 10 lines: *C. campanulata* (Kanhi-zakura [Cerasus\_1-57]); *C. serrulata* ('Azumanishiki' [Cerasus\_2-51], 'Gyoikou' [Cerasus\_2-19], 'Ichihara-toranowo' [Cerasus\_2-27], and 'Senrikou' [Cerasus\_1-37]); *C. spachiana* ('Yaebeni-shidare' [Cerasus\_1-43]); *C. speciosa* (Ohshima-zakura [Cerasus\_1-71]); *C. jamasakura* (Yamazakura); *C. × nudiflora* ('Eishu-zakura'); and *Cerasus* sp. ('Oshidori-fujizakura' [Cerasus\_2-25]). SNV detection was performed as above. Major alleles across the 10 lines were

presumed to be ancestor-type alleles, and the ancestor-type alleles in each of the 46 'Somei-Yoshino' samples were counted.

## Results

### Cultivar identification

To confirm the 46 samples used in this study were indeed 'Somei-Yoshino' clones, a clustering analysis based on genotypes of the 46 'Somei-Yoshino' samples and 139 *Cerasus* lines. Whole-genome sequence reads obtained from 46



**Figure 1.** Phylogenetic tree of flowering cherry lines. Phylogenetic tree of 46 'Somei-Yoshino' clones from this study and 139 cherry lines from our previous study.<sup>2</sup> Black and white stars indicate the 'Somei-Yoshino' tree (Cerasus\_72-1) used in our previous study<sup>2</sup> and 46 clones used in this study, respectively. Black circles show synthetic hybrids, *Cerasus* × *yedoensis*, generated by crossing *C. spachiana* and *C. speciosa*.<sup>3</sup> Black and white squares indicate probable parental lines of *C. spachiana* and *C. speciosa*, respectively. Unlabeled lines are non-relatives of 'Somei-Yoshino'. Numbers on branches indicate bootstrap values based on 100 replicates. Asterisk indicates the connecting point of the branch.

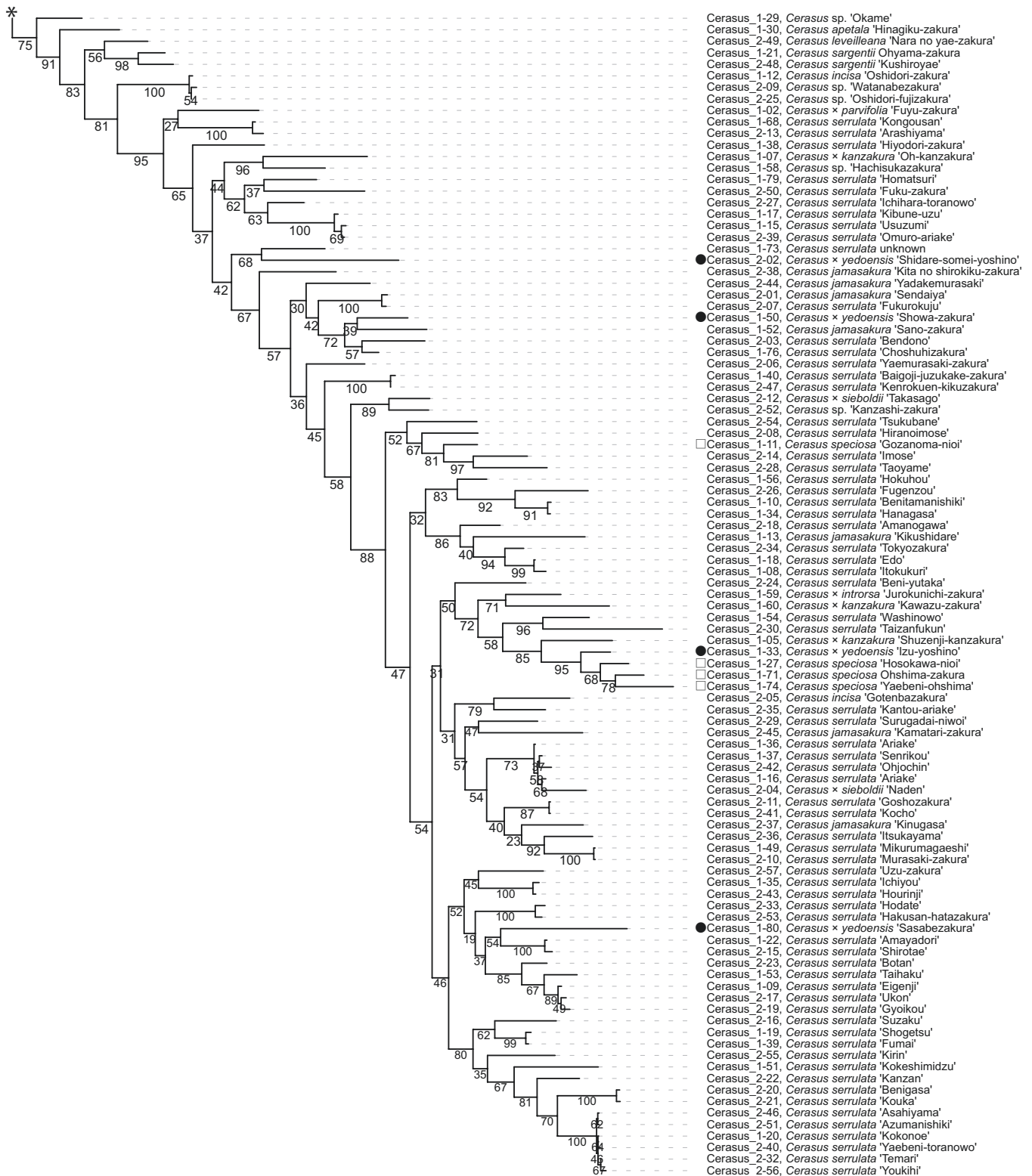


Figure 1. Continued

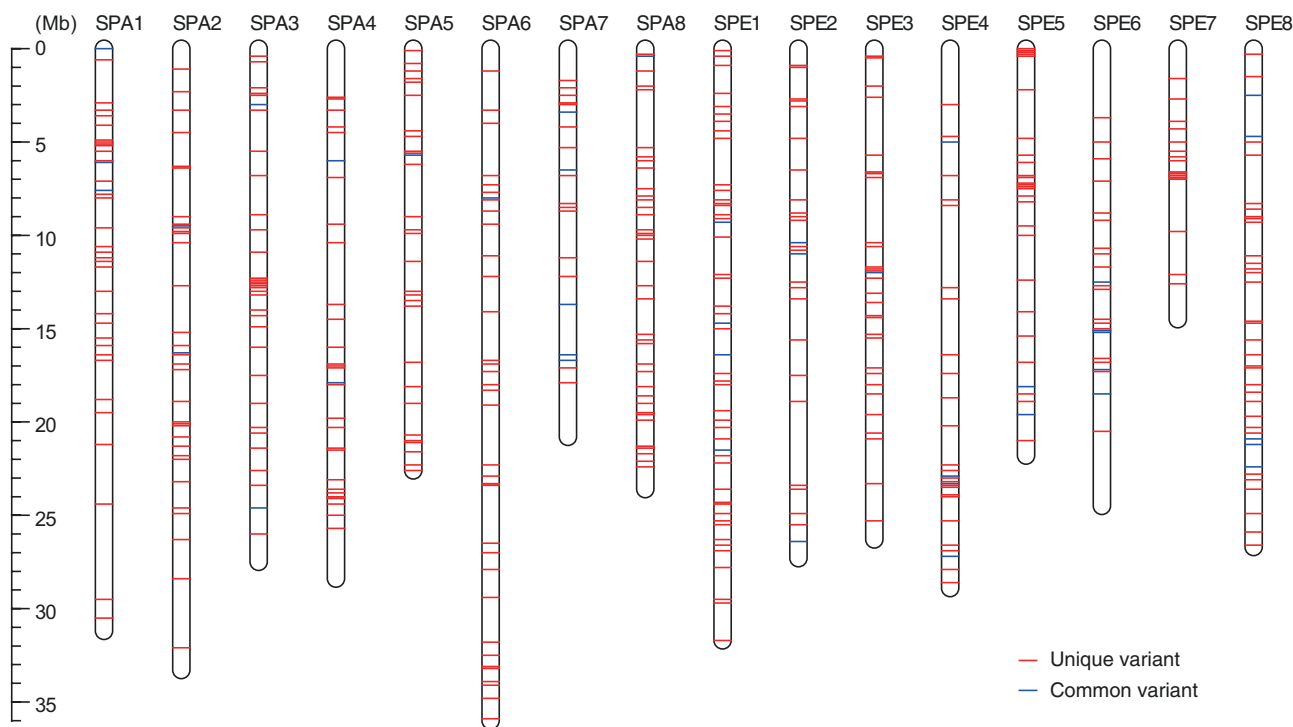
samples (mean depth of 34× genome coverage: [Supplementary Table S1](#)) and ddRAD-Seq reads obtained from 139 lines in our previous study,<sup>2</sup> including 'Somei-Yoshino' (Sample ID of Cerasus\_1-72 in the previous study), were aligned on to the reference genome sequence of 'Somei-Yoshino'. A total of 5,328 SNPs were detected across the 185 samples, and genetic distances between each pair of the 185 samples were calculated

to create a dendrogram (Fig. 1, [Supplementary Table S2](#)). As expected, the 46 samples and 'Somei-Yoshino' (Cerasus\_1-72) clustered together, while the other lines including probable sister (*Cerasus* × *yedoensis*) and parent (*C. spachiana* and *C. speciosa*) lines were distinguishable from 'Somei-Yoshino'. The result indicated that the 46 samples used in this study were genetically identified as 'Somei-Yoshino' clones.

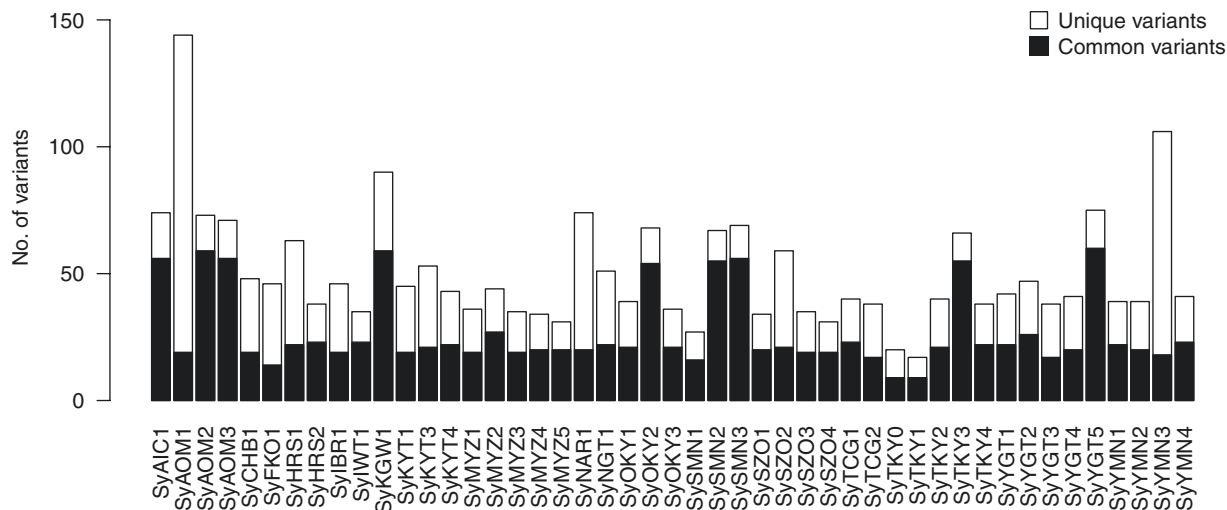
## Detection and characterization of somatic mutations

Based on the whole-genome sequence analysis of the 46 'Somei-Yoshino' samples, 80,334 sequence variant candidates were detected. First, we selected 35,757 biallelic SNVs since tri- and tetra-allelic SNVs are generally rare and probably due to mapping errors in and near repeated sequences.<sup>27</sup> Next, 1,942 sites were selected, on which only reads from haplotype-specific alleles were mapped on the haplotype-resolved phased genome sequence of 'Somei-Yoshino'. Then, 1,749 single nucleotide variants, whose genotypes were consistent among the biological replicates of SyTKY0 and SyTKY1, were retained. Finally, we applied two filtering criteria, namely, read depth ( $\leq 50$ ) and

quality ( $\geq 80$ ), to select 684 high-confidence SNVs, which were evenly distributed across the genome (Fig. 2, Supplementary Table S3). The number of variants across the 46 samples was 50.3, on average, with the maximum value of 144 in SyAOM1, followed by 106 in SyYMN3 and 90 in SyKGW1 (Fig. 3, Supplementary Tables S1 and S3). The 684 SNVs consisted of 285 C/G to T/A transitions (41.7%), 122 A/T to T/A transversions (17.8%), 120 A/T to G/C transversions (17.5%), 72 C/G to A/T transversions (10.5%), 51 C/G to G/C transversions (7.5%), and 34 A/T to C/G transversions (5.0%). The transition/transversion ratio was 1.45. Among the 684 variants, 88 variants (12.9%) were in gene bodies, whereas the remaining 596 variants (87.1%) were found in intergenic regions.



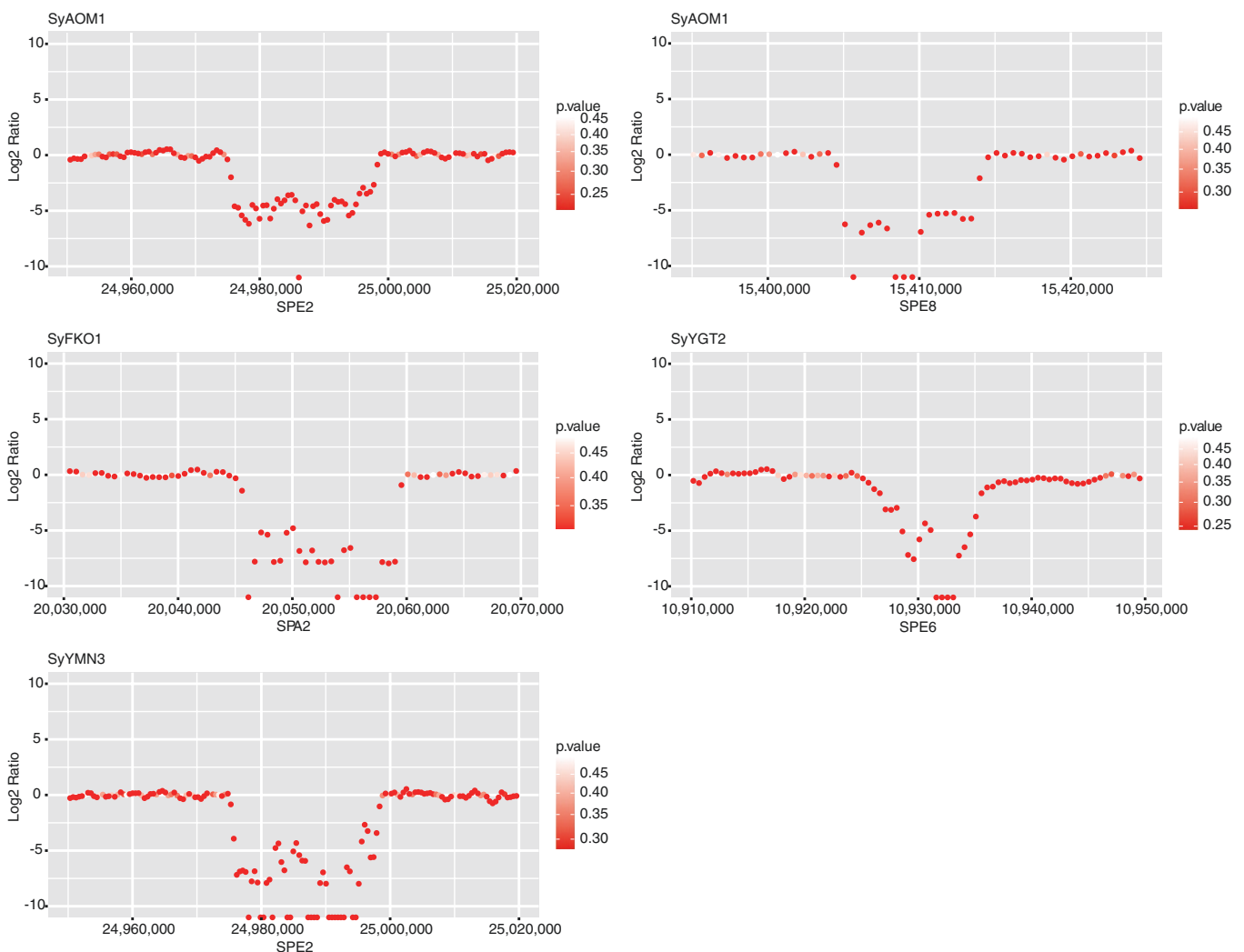
**Figure 2.** Genomic position of the 684 somatic mutations. SPA1–SPA8 and SPE1–SPE8 indicate the chromosomes of 'Somei-Yoshino'.



**Figure 3.** Numbers of variants in the 46 'Somei-Yoshino' clones. Black and white bars show numbers of common and unique variants, respectively.

**Table 1.** Somatic mutations highly affecting gene functions

Gene ID	Mutation	Description
CYE_r3.1SPA4_g020770	Splice acceptor variant and intron variant	[Y4729_ARATH] G-type lectin S-receptor-like serine/threonine protein kinase At4g27290
CYE_r3.1SPA5_g009120	Stop gained	[ARC_HUMAN] Activity-regulated cytoskeleton-associated protein
CYE_r3.1SPA5_g023100	Splice acceptor variant and intron variant	[SIS3_ARATH] E3 ubiquitin-protein ligase SIS3
CYE_r3.1SPE0_g037740	Splice donor variant and intron variant	[RNHX1_ARATH] Putative ribonuclease H protein At1g65750
CYE_r3.1SPE0_g067760	Stop gained	[POLX_TOBAC] Retrovirus-related Pol polyprotein from transposon TNT 1-94
CYE_r3.1SPE2_g031860	Gene deletion	[LBD29_ARATH] LOB domain-containing protein 29
CYE_r3.1SPE2_g031870	Gene deletion	[GRD2I_MOUSE] Delphinin
CYE_r3.1SPE2_g031880	Gene deletion	No hits
CYE_r3.1SPE5_g007060	Stop gained	[ELH1_APLCA] ELH
CYE_r3.1SPE6_g015220	Gene deletion	[DRL36_ARATH] Probable disease resistance protein At5g45510
CYE_r3.1SPE6_g015230	Gene deletion	[CITG2_SALTY] Probable 2-(5'-triphosphoribosyl)-3'-dephosphocoenzyme-A synthase 2
CYE_r3.1SPE8_g002010	Stop gained	[LKHA4_DEBHA] Leukotriene A-4 hydrolase homolog
CYE_r3.1SPE8_g017860	Gene deletion	[AB2A_ARATH] ABC transporter A family member 2



**Figure 4.** Copy number variations (CNVs) detected in the four ‘Somei-Yoshino’ clones. Clone and chromosome names are shown above and below the plots, respectively. Numbers on x-axes indicate chromosome positions (bp). The log<sub>2</sub> ratio on y-axes above zero indicates insertions in the tested lines with respect to that of TKY0 as a reference line. The *P*-values are calculated as a probability of a copy number ratio being divergent from 1:1 ratio by a random chance. Details are mentioned by Xie and Tammi (2009).

Of the 684 variants, 613 were unique to a single clone and 71 were common to at least two clones (Fig. 3, Supplementary Tables S1 and S3). Of the 88 variants identified in gene bodies, seven variants, all of which were unique to a single clone, were predicted to have a high impact on gene functions (Table 1, Supplementary Tables S3 and S4). These seven variants included four nonsense mutations (in *CYE\_r3.1SPA5\_g009120* in SyYMN2, *CYE\_r3.1SPE0\_g067760* in SyYGT3, *CYE\_r3.1SPE5\_g007060* in SyIBR1, and *CYE\_r3.1SPE8\_g002010* in SyYMN3), two mutations at splice acceptor sites (in *CYE\_r3.1SPA4\_g020770* in SyNGT1 and *CYE\_r3.1SPA5\_g023100* in SyYGT4), and one at a splice donor site (in *CYE\_r3.1SPE0\_g037740* in SyYMN3). Missense mutations were found in 23 genes (Supplementary Tables S3 and S4).

### CNVs among the ‘Somei-Yoshino’ clones

At least five CNVs were found in four clones, SyAOM1, SyFKO1, SyYGT2, and SyYMN3 (Fig. 4, Table 1, Supplementary Table S4). All five CNVs were deletion mutations with respect to SyTKY1 as a standard; SyAOM1 carried two CNVs, one on chromosome SPE2 (~25 kb deletion) and another on chromosome SPE8 (~10 kb deletion), while each of the three remaining clones, SyFKO1, SyYGT2, and SyYMN3, carried a single deletion on chromosomes SPA2 (~15 kb deletion), SPE6 (~10 kb deletion), and SPE2 (~25 kb deletion), respectively. Among these CNVs, the deletions on chromosome SPE2 in SyAOM1 and SyYMN3 might be identical. The deletions encompassed three genes on chromosome SPE2 (*CYE\_r3.1SPE2\_g031860*, *CYE\_r3.1SPE2\_g031870*, and *CYE\_r3.1SPE2\_g031880*), one gene on SPE8 (*CYE\_r3.1SPE8\_g017860*), and two genes on SPE6 (*CYE\_r3.1SPE6\_g015220* and *CYE\_r3.1SPE6\_g015230*), while the deletion on SPA2 contained no gene.

### Clustering analysis of ‘Somei-Yoshino’ clones

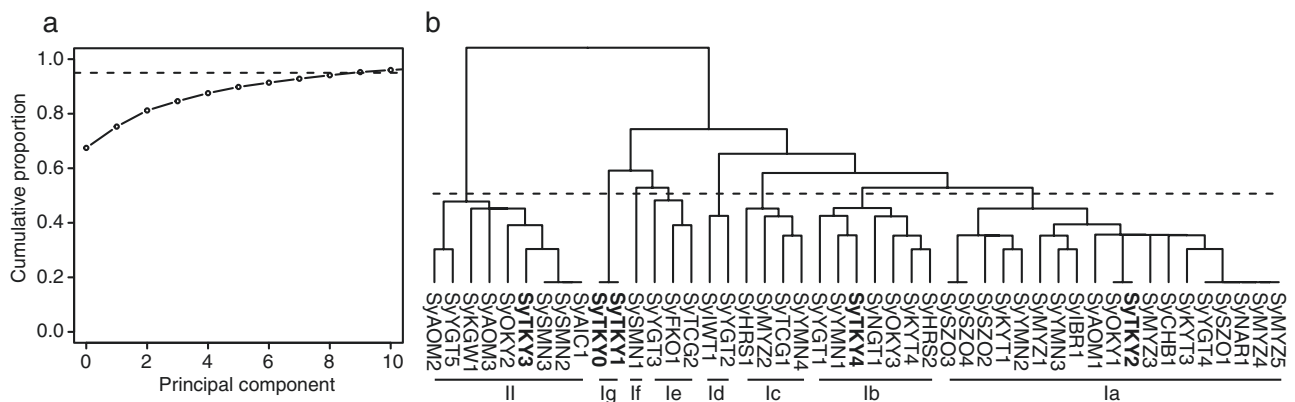
The cumulative explained variance was calculated with the 71 common variants (Fig. 5a). When the threshold value was set to 0.95, a cluster number of 8 was the smallest value of the cluster with a cumulative explained variance ratio. Therefore, the 46 samples were divided into eight groups, including two major groups (I and II) with seven subgroups (Ia to Ig) (Fig. 5b). The numbers of clones in each group were as follows: 19 (Ia), 7 (Ib), 4 (Ic), 2 (Id), 3 (Ie), 1 (If), 1 (Ig), and 9 (II). These

clusters showed no correlation with the sample collection site. Each of the four clones collected from Ueno Park (Tokyo, Japan), SyTKY1–SyTKY4, grouped into four different clusters (Ia, Ib, Ie, and II), which contained 38 of the 46 clones tested.

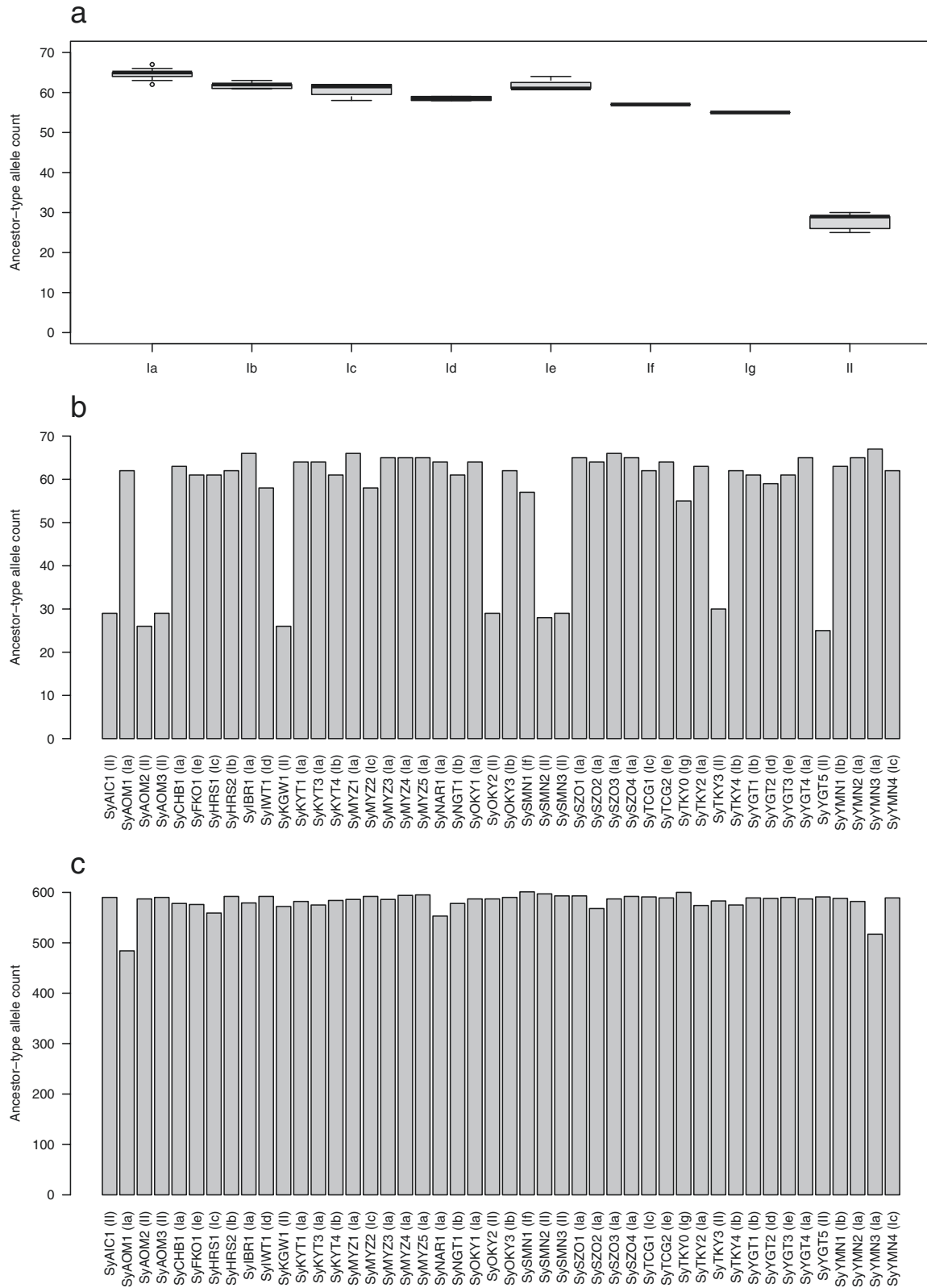
To identify the group close to the potential ancestral type of ‘Somei-Yoshino’, genotypes of the 71 common variants in 46 samples were compared with those of 10 *Cerasus* lines belonging to at least six species, *C. campanulate* (Kanhi-zakura), *C. serrulate* (‘Azumanishiki’, ‘Gyoikou’, ‘Ichihara-toranowo’, and ‘Senrikou’), *C. spachiana* (‘Yaebeni-shidare’), *C. speciosa* (Ohshima-zakura), *C. jamasakura* (Yama-zakura), and *C. × nudiflora* (‘Eishu-zakura’) in addition to *Cerasus* sp. (‘Oshidori-fujizakura’). Since no variations were observed at the 71 sites of the 10 lines, the alleles possessed by the 10 lines were presumed to be the ancestor-type alleles of *Cerasus*. The mean values of the ancestor-type allele count in the eight groups were 64.6 (Ia), 61.7 (Ib), 60.8 (Ic), 58.5 (Id), 62.0 (Ie), 57.0 (If), 55.0 (Ig), and 27.9 (II) (Fig. 6a). At the sample level, the ancestor-type allele counts of the 71 common variants ranged from 25 (SyYGT5) to 67 (SyYMN3) (Fig. 6b) and those of the 613 unique variants were from 484 (SyAOM1) to 601 (SySMN1) (Fig. 6c).

### Discussion

A total of 684 somatic mutations were detected across 46 ‘Somei-Yoshino’ trees (Fig. 2, Supplementary Table S3). Because this mutation frequency was quite lower than that found among the flowering cherry accessions, the 46 trees were identified as clones of ‘Somei-Yoshino’ (Fig. 1, Supplementary Table S2). Of the 684 mutations, 71 were shared by multiple clones (Fig. 3, Supplementary Table S4). Since somatic mutations are seldom reversible, each of these 71 mutations would have a common origin and would be distributed across multiple lines via clonal propagation. Based on this analysis, the 46 ‘Somei-Yoshino’ clones were clustered into eight groups, Ia–Ig and II (Fig. 5). Interestingly, four trees collected from Ueno Park were classified into four different groups (Fig. 5b). Furthermore, 34 clones in addition to the four Ueno trees were included in the four groups (Fig. 5b). The ‘Somei-Yoshino’ clones tested in this study were collected from different locations across Japan, which might imply that



**Figure 5.** Clusters of 46 ‘Somei-Yoshino’ clones. (a) Cumulative explained variance in the principal component analysis. A horizontal dotted line indicates 0.95 of cumulative explained variance. (b) Dendrogram of 46 ‘Somei-Yoshino’ clones. Boldface indicates four trees planted in Ueno Park (Tokyo, Japan). SyTKY0 and SyTKY1 represent different sequence datasets obtained from the same tree (#136). Group names (Ia–Ie and II) are shown below the dendrogram.



**Figure 6.** Ancestral alleles of 'Somei-Yoshino' clones. (a) Ancestral allele count in 71 common mutations for the eight groups of 'Somei-Yoshino' clones. (b) Ancestral allele count of 'Somei-Yoshino' clones in 71 common mutations at the sample level. (c) Ancestral allele count of 'Somei-Yoshino' clones in 613 unique mutations at the sample level.

'Somei-Yoshino' trees in Japan might be mainly clustered into the four groups. Among the four trees planted in Ueno Park, SyTKY2 (tree ID #133), a member of group Ia, possessed the highest number of ancestor-type alleles among the four trees

in Ueno Park (Fig. 6). This result suggested that SyTKY2 (tree ID #133) was the closest to the ancestral type. Although the actual origin of the clonally propagated 'Somei-Yoshino' was still unclear, we obtained a key set of somatic mutations to



identify the origin. We hypothesize that the original tree could be a chimera composed of the somatic mutations found in the four groups. Since there are many candidates for the origin of 'Somei-Yoshino' in Japan<sup>28</sup> as well as in Koishikawa Botanical Garden (Tokyo, Japan)<sup>7</sup> and Kaiseizan Park (Fukushima, Japan),<sup>9</sup> the origin could be discovered by finding the chimera.

The detected mutations consisted of 405 transitions and 279 transversions, with the transition/transversion ratio of 1.45. Among the different mutation types, the C/G to T/A transitions were the most prominent, the proportion of which (41.7%) was comparable with that of somatic mutations in popular<sup>12</sup> and ethyl methanesulfonate (EMS)-induced artificial mutations in tomato.<sup>29</sup> In addition, large-scale deletions (10–25 kb) were also found as somatic mutations, which have been found in not only chemical mutagenesis but also physical mutagenesis studies.<sup>27,29</sup> Owing to these mutations, the functions of at least 13 and 23 genes might be severely and partially lost in the clones, respectively (Table 1, Supplementary Tables S3 and S4). Out of them, seven genes were reported as disease resistance-related genes, suggesting that these mutations could change disease resistance levels even in clonally propagated 'Somei-Yoshino'. In several vegetatively propagated crops, mainly fruit trees, bud sports caused by somatic mutations were reported and used as new cultivars.<sup>14</sup> For example, in grapes, a transposable element was reported as an inducer of a bud sport, in which the berry skin color was changed from black to white.<sup>13</sup> However, few reports on phenotypic variations are available in 'Somei-Yoshino'. Even though no deleterious mutations were found in genes involved in the flower opening mechanism,<sup>5</sup> further investigation would be required to clarify whether the phenotypic variations are caused by genetic factors (somatic mutations) and/or environmental conditions.

The number of somatic mutations in the 46 trees was varied (Fig. 3, Supplementary Table S3). This variation in the number of somatic mutations was likely reflected by the number of unique variants rather than that of common variants, even though common variants were more frequent in group II than in group I (Fig. 3). The number of ancestor-type allele counts in the common variants might indicate the time of divergence from the ancestor, suggesting that the group II might be older than the other groups. On the other hand, the number of ancestor-type allele counts in the unique variants (Fig. 6c) might indicate the age of the clone after its propagation via cutting or grafting. It is believed that SyAOM1 is the oldest 'Somei-Yoshino' clone planted in 1888.<sup>28</sup> Although the SyAOM2 tree is thought to be as old as the SyAOM1 trees, the ancestor-type allele counts of SyAOM2 was quite different from that of SyAOM1 (Fig. 6). SySZO1 tree is believed to have been obtained from Washington DC, USA, where 'Somei-Yoshino' trees from Japan were planted in 1912, but the number of ancestor-type allele counts in SySZO1 was similar to that in other clones (Fig. 6). In addition, although SyHRS2 is thought to have survived the atomic bomb attack in 1945 during World War II, the number of ancestor-type allele counts in this clone is not high (Fig. 6). Overall, the relationship between the ancestor-type allele counts and the age of 'Somei-Yoshino' was unclear.

In summary, we identified and characterized somatic mutations in 'Somei-Yoshino' clones collected from all over Japan. Since somatic mutations occur in different organs of an individual and are rarely reversed, leading to chimerism. Even if clonally propagated offspring basically inherit the

mutations from the ancestors, their genotypes could be genetically divergent depending on the branches used for the propagation. Conversely, the somatic mutations could be used as tracers to find out the original tree. The somatic mutations found in this study could be key to identifying the origin of 'Somei-Yoshino', which has not been found to date. Furthermore, this somatic mutation-based tracing method could be used in agriculture to ensure the quality control of vegetatively propagated crops such as orange, apple, grape, strawberry, sweet potato, and tea, to protect the rights of breeders.

## Acknowledgements

We thank Prof. O. Arakawa (Hirosaki University, Japan) and two anonymous contributors (A.T. and M.T.) for collecting leaf samples; Hirosaki Park (Aomori, Japan), Takamatsu Park (Iwate, Japan), Kajo Park (Yamagata, Japan), Ueno Park (Tokyo, Japan), Kokubo Park (Yamanashi, Japan), Kumano Shrine (Yamanashi, Japan), Okayama Prefectural Multipurpose Grounds (Okayama, Japan), Toryo Park (Kagawa, Japan), Maiduru Park (Fukuoka, Japan), and Mochio Park (Miyazaki, Japan) for providing leaf samples; Prof. K. Okada (Ryukoku University, Japan) and Dr T. Yukawa (National Museum of Nature and Science, Japan) for insightful discussions; Ms H. Umemura (The Yokohama Nursery Co., Ltd, Japan) for providing literatures about 'Somei-Yoshino' planted in Washington DC; and Y. Kishida, C. Minami, K. Ozawa, H. Tsuruoka, and A. Watanabe (Kazusa DNA Research Institute) for technical assistance.

## Funding

This study was supported in part by JSPS KAKENHI (grant numbers 22H05172, 22H05181, 22K05613, and 22K19192) and the Kazusa DNA Research Institute Foundation.

## Conflict of interest

None declared.

## Data availability

The sequence reads are available from the DDBJ Sequence Read Archive (DRA) under the accession numbers DRR493485–DRR493529 of BioProject PRJDB16216.

## Supplementary Data

Supplementary data are available at *DNARES* online.

## References

- Innan, H., Terauchi, R., Miyashita, N.T., and Tsunewaki, K. 1995, DNA fingerprinting study on the intraspecific variation and the origin of *Prunus yedoensis* (Someiyoshino), *Jpn. J. Genet.*, 70, 185–96.
- Shirasawa, K., Esumi, T., Hirakawa, H., et al. 2019, Phased genome sequence of an interspecific hybrid flowering cherry, 'Somei-Yoshino' (*Cerasus* × *yedoensis*), *DNA Res.*, 26, 379–89.
- Takenaka, Y. 1963, The origin of the yoshino cherry tree, *J. Hered.*, 54, 207–11.

4. Aono, Y. and Murakami, N. 2017, A simplified method to estimate cherry blossom phenology considering temperature during endodormancy process, *Clim Biosphere*, **17**, 25–33.
5. Shirasawa, K., Esumi, T., Itai, A., and Isobe, S. 2022, Cherry blossom forecast based on transcriptome of floral organs approaching blooming in the flowering cherry (*Cerasus × yedoensis*) cultivar ‘Somei-Yoshino’, *Front. Plant Sci.*, **13**, 802203.
6. Iketani, H., Ohta, S., Kawahara, T., et al. 2007, Analyses of clonal status in ‘Somei-yoshino’ and confirmation of genealogical record in other cultivars of *Prunus × yedoensis* by microsatellite markers, *Breed. Sci.*, **57**, 1–6.
7. Nakai, T. 1935, The oldest specimen of *Prunus yedoensis* in the botanic garden of Tokyo Imperial University, *J. Jpn. Bot.*, **11**, 341–6.
8. Nakamura, I., Takahashi, H., and Ohta, S. 2015, Origin of *Prunus × yedoensis* Somei-yoshino based on sequence analysis of PolA1 gene, *Adv Horticult. Sci.*, **29**, 17–23.
9. Sampei, Y. 2018, A verification of the age about old Tokyo cherry at Kaiseizan Park in Koriyama City, *Tree and Forest Health*, **22**, 195–8.
10. Wang, L., Ji, Y., Hu, Y., et al. 2019, The architecture of intra-organism mutation rate variation in plants., *PLoS Biol.*, **17**, e3000191.
11. Orr, A.J., Padovan, A., Kainer, D., et al. 2020, A phylogenomic approach reveals a low somatic mutation rate in a long-lived plant, *Proc. Biol. Sci.*, **287**, 20192364.
12. Hofmeister, B.T., Denkena, J., Colomé-Tatché, M., et al. 2020, A genome assembly and the somatic genetic and epigenetic mutation rate in a wild long-lived perennial *Populus trichocarpa*, *Genome Biol.*, **21**, 259.
13. Kobayashi, S., Goto-Yamamoto, N., and Hirochika, H. 2004, Retrotransposon-induced mutations in grape skin color, *Science*, **304**, 982.
14. Sanada, T., Kotobuki, K., Nishida, T., Fujita, H., and Ikeda, F. 1993, A new Japanese pear cultivar ‘gold nijisseiki’, resistant mutant to black spot disease of Japanese pear, *Ikushugaku Zasshi.*, **43**, 455–61.
15. Diwan, D., Komazaki, S., Suzuki, M., et al. 2014, Systematic genome sequence differences among leaf cells within individual trees, *BMC Genomics.*, **15**, 142.
16. Ueno, S., Hasegawa, Y., Kato, S., et al. 2023, Rapid survey of de novo mutations in naturally growing tree species following the March 2011 disaster in Fukushima: the effect of low-dose-rate radiation, *Environ. Int.*, **174**, 107893.
17. Schmieder, R. and Edwards, R. 2011, Quality control and preprocessing of metagenomic datasets, *Bioinformatics*, **27**, 863–4.
18. Langmead, B. and Salzberg, S.L. 2012, Fast gapped-read alignment with Bowtie 2, *Nat. Methods.*, **9**, 357–9.
19. Li, H. 2011, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, *Bioinformatics*, **27**, 2987–93.
20. Danecek, P., Auton, A., Abecasis, G., et al.; 1000 Genomes Project Analysis Group. 2011, The variant call format and VCFtools, *Bioinformatics*, **27**, 2156–8.
21. Voorrips, R.E. 2002, MapChart: software for the graphical presentation of linkage maps and QTLs, *J. Hered.*, **93**, 77–8.
22. Xie, C. and Tammi, M.T. 2009, CNV-seq, a new method to detect copy number variation using high-throughput sequencing, *BMC Bioinf.*, **10**, 80.
23. Cingolani, P., Platts, A., Wang, L.L., et al. 2012, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3, *Fly*, **6**, 80–92.
24. Lee, T.-H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. 2014, SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data, *BMC Genomics*, **15**, 162.
25. Letunic, I. and Bork, P. 2021, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Res.*, **49**, W293–6.
26. Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. 2007, TASSEL: software for association mapping of complex traits in diverse samples, *Bioinformatics*, **23**, 2633–5.
27. Ichida, H., Morita, R., Shirakawa, Y., Hayashi, Y., and Abe, T. 2019, Targeted exome sequencing of unselected heavy-ion beam-irradiated populations reveals less-biased mutation characteristics in the rice genome, *Plant J.*, **98**, 301–14.
28. Iwasaki, F. 1989, Bibliographic studies on the origin of the flowering cherry, Somei-yoshino (*Prunus yedoensis* Matsumura), *Bull Agr For Res U Tsukuba.*, **1**, 85–103.
29. Shirasawa, K., Hirakawa, H., Nunome, T., Tabata, S., and Isobe, S. 2016, Genome-wide survey of artificial mutations induced by ethyl methanesulfonate and gamma rays in tomato, *Plant Biotechnol. J.*, **14**, 51–60.