

# Engineering homoeologs provide a fine scale for quantitative traits in polyploid

Eun Song Lee<sup>1,†</sup>, Jung Heo<sup>1,2,†</sup>, Woo Young Bang<sup>3,†</sup>, Kapeel M. Chougule<sup>4,†</sup>, Nomar Espinosa Waminal<sup>5,6</sup>, Nguyen Thi Hong<sup>6</sup>, Min Ji Kim<sup>1</sup>, Hong Kwan Beak<sup>1</sup>, Yong Jun Kim<sup>1</sup>, Ryza A. Priatama<sup>1,7</sup>, Ji In Jang<sup>1,2</sup>, Kang Il Cha<sup>1</sup>, Seung Han Son<sup>1</sup>, Sujeevan Rajendran<sup>1</sup>, Young-Kug Choo<sup>1</sup>, Jong Hyang Bae<sup>8</sup>, Chul Min Kim<sup>8</sup>, Young Koung Lee<sup>7</sup>, Sangsu Bae<sup>9</sup>, Jonathan D. G. Jones<sup>10</sup>, Kee Hoon Sohn<sup>11</sup>, Jiyoung Lee<sup>12</sup>, Hyun Hee Kim<sup>6</sup>, Jong Chan Hong<sup>2</sup>, Doreen Ware<sup>4,13,\*</sup>, Keunhwa Kim<sup>1,2,\*</sup> and Soon Ju Park<sup>1,2,\*</sup> 

<sup>1</sup>Division of Biological Sciences, Wonkwang University, Iksan, Korea

<sup>2</sup>Division of Applied Life Science (BK21 four) and Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Gyeongsang National University, Jinju, Korea

<sup>3</sup>Biological and Genetic Resources Assessment Division, National Institute of Biological Resources, Incheon, Korea

<sup>4</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

<sup>5</sup>Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

<sup>6</sup>BioScience Institute, Department of Chemistry & Life Science, Sahmyook University, Seoul, Korea

<sup>7</sup>Institute of Plasma Technology, Korea Institute of Fusion Energy, Gunsan-si, Korea

<sup>8</sup>Division of Horticulture Industry, Wonkwang University, Iksan, Korea

<sup>9</sup>Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, South Korea

<sup>10</sup>The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Norwich, UK

<sup>11</sup>Department of Agricultural Biotechnology, Plant Immunity Research Center, Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

<sup>12</sup>Korean Collection for Type Cultures (KCTC), Biological Resource Center, Korea Research Institute of Bioscience and Biotechnology, Jeongeup, Korea

<sup>13</sup>U.S. Department of Agriculture-Agricultural Research Service, NEA Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA

Received 12 May 2023;

revised 23 June 2023;

accepted 20 July 2023.

\*Correspondence (Tel +1-516-367-6979; fax +1-516-367-6851; email [ware@cshl.edu](mailto:ware@cshl.edu) (DW); Tel +82-55-772-1356; fax +82-55-772-2563; email [keunhwa@gnu.ac.kr](mailto:keunhwa@gnu.ac.kr) (KK); Tel +82-55-772-1356; fax +82-55-772-2563; email [sjpark75@gnu.ac.kr](mailto:sjpark75@gnu.ac.kr) (SJP))

†These authors contributed equally to this work.

**Keywords:** polyploid, homoeologous gene editing, quantitative traits, agricultural improvement.

## Summary

Numerous staple crops exhibit polyploidy and are difficult to genetically modify. However, recent advances in genome sequencing and editing have enabled polyploid genome engineering. The hexaploid black nightshade species *Solanum nigrum* has immense potential as a beneficial food supplement. We assembled its genome at the scaffold level. After functional annotations, we identified homoeologous gene sets, with similar sequence and expression profiles, based on comparative analyses of orthologous genes with close diploid relatives *Solanum americanum* and *S. lycopersicum*. Using CRISPR-Cas9-mediated mutagenesis, we generated various mutation combinations in homoeologous genes. Multiple mutants showed quantitative phenotypic changes based on the genotype, resulting in a broad-spectrum effect on the quantitative traits of hexaploid *S. nigrum*. Furthermore, we successfully improved the fruit productivity of Boranong, an orphan cultivar of *S. nigrum* suggesting that engineering homoeologous genes could be useful for agricultural improvement of polyploid crops.

## Introduction

Polyploidy is commonly observed in flowering plants and can contribute to plant speciation (Leitch and Leitch, 2008; Wendel, 2000). Staple crops such as wheat, potatoes and rapeseed are polyploid (Comai, 2005). Despite the disadvantages conferred by chromosomal instabilities, the resulting gene dosage changes provide gene redundancy and protect from deleterious mutations (Comai, 2005). However, polyploid genome sequencing and assembly is still challenging due to its genome size and the presence of duplicated sequences. Improvements in current sequencing technology and genome assembly tools make polyploid genome analysis more feasible (Weeks, 2017). In addition, using the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 system, we could perform targeted mutagenesis, which also aids in the genome engineering of polyploids.

*Solanum nigrum* is a naturally occurring hexaploid black nightshade belonging to the Solanaceae family; its fruit contains beneficial metabolites, including anthocyanins, and has great potential as a human health supplement (Edmonds and Chweya, 1997; Jabamalairaj *et al.*, 2019). *Solanum nigrum* is evolutionarily close to tomato (*Solanum lycopersicum*; Edmonds and Chweya, 1997); indeed, it has small berry fruits and can tolerate environmental stress, like ancient wild tomato species (Blanca *et al.*, 2015). As the genetic changes during tomato domestication have been extensively studied, *de novo* domestication to improve and optimize agricultural traits such as fruit yield could be achieved by genetic modification of wild species in the Solanaceae family (Zsögön *et al.*, 2018). Moreover, specific genes involved in quantitative traits (QT) regulation have been identified to improve tomatoes (Rodríguez-Leal *et al.*, 2017; Soyk *et al.*, 2020). These genes can generate genetic dosage effects that induce QTs through three

possible mechanisms: (i) conventional dosage by the knockout allele showing hybrid vigour in heterozygotes, (ii) expressional dosage of the gene by promoter editing and (iii) functional dosage by paralogous gene compensation. For example, *SINGLE FLOWER TRUSS (SFT)* was the first identified single overdominant gene, where the heterozygous loss-of-function alleles (*sft*+/+) increase fruit yield compared to the wild type or the *sft* knockout mutant (Krieger *et al.*, 2010). Additionally, heterozygous *COMPOUND INFLORESCENCE (S)* gene (*s*+) also results in increased yield (Lippman *et al.*, 2008; Park *et al.*, 2012; Soyk *et al.*, 2017) and promoter editing of *S* and *SELF PRUNING (SP)* generates diverse *cis*-regulatory alleles that provide beneficial quantitative variation (Rodríguez-Leal *et al.*, 2017). Recently, *CLV3/EMBRYO-SURROUNDING REGION-RELATED 9 (CLE9)*, a paralog of tomato *CLAVATA3 (CLV3)*, was found to partially compensate for the *clv3* mutation in the reproductive organ development, suggesting a possible QT contribution (Kwon *et al.*, 2022). In diploid species, it is difficult to generate a broad spectrum of QT by single gene editing due to limited genetic resources. However, in polyploid species, it is possible given the increased copy numbers of QT-related genes as homoeologous genes. We hypothesized we could generate a fine-scale quantitative variation by modulating genetic dosages through gene editing. Taking advantage of CRISPR-Cas9 system, we were able to fine-tune the genetic and/or functional

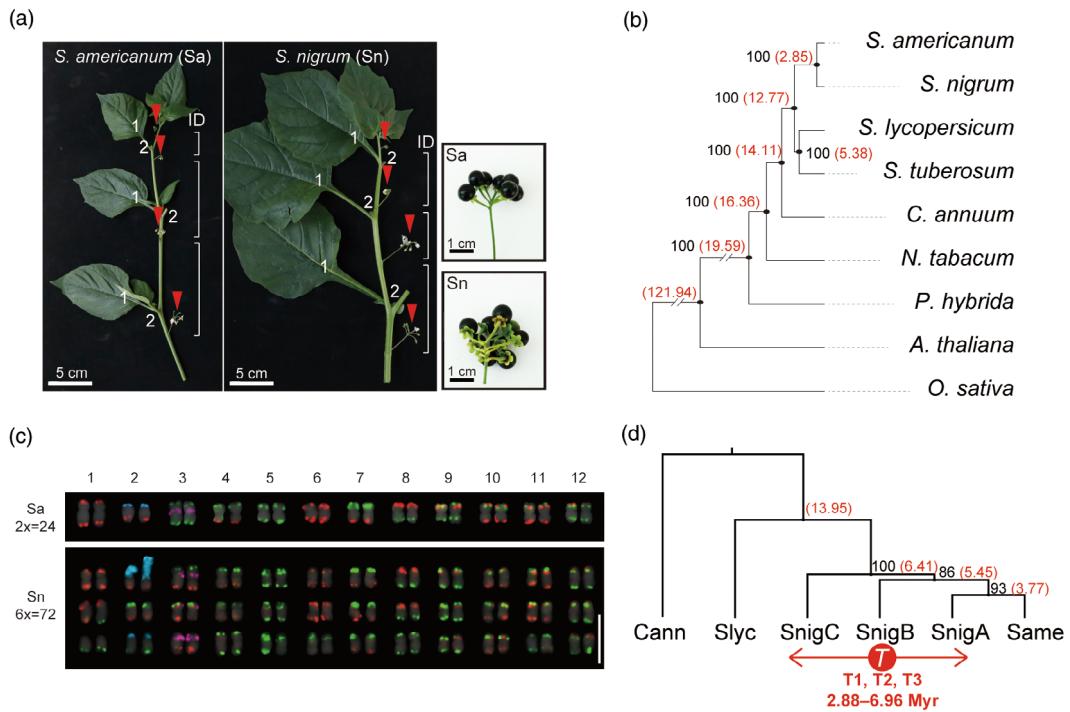
dosage of the target genes and select trait-optimized polyploid plant lines.

To engineer a polyploid genome for crop improvement, genomic information must be obtained. Although a molecular analysis of the fruit metabolite regulation is performed based on the transcriptome (Heo *et al.*, 2022), the whole genome of hexaploid *S. nigrum* was not previously defined. In this study, we sequenced and assembled the *S. nigrum* genome. After functional annotation of the genome information in comparison with the diploid relatives *Solanum americanum* and *S. lycopersicum*, we identified and analysed homoeologs in *S. nigrum*. Through homoeologs editing, we validated the effects of genetic dosage on QT and modified a regional crop cultivar of *S. nigrum*. Taken together, this study provides a broadly applicable path for engineering polyploid crops.

## Results

### Hexaploid *S. nigrum* and its closely related diploid *S. americanum*

With respect to shoot architecture, the hexaploid *S. nigrum* and diploid *S. americanum* (SP2273, Figure 1a) showed a similar indeterminate (ID) growth pattern with the same sympodial index (SI = 2). We intercrossed these species, and F1 hybrid plants produced fertile reproductive organs (Figure S1). Additionally,



**Figure 1** Evolutionary closeness between hexaploid *Solanum nigrum* and diploid *Solanum americanum*. (a) Sympodial shoot structures and ripe fruits (inset) of *S. nigrum* (accession NIBRGR0000189638) and *S. americanum* (accession SP2273), respectively. Red arrowheads indicate inflorescence. Bracket and number represent the interval between inflorescences and leaf numbers on each sympodial unit. ID, indeterminate growth. (b) Phylogenetic tree of chloroplast cDNA sequences from various Solanaceae species. The black number denotes the bootstrap of each node. The red number indicates the estimated divergence time (million years ago, myr). *Oryza sativa* was used as an outgroup control. (c) Fish karyotypes of *S. americanum* and *S. nigrum*. The numbers above indicate the chromosome number. Green and red signals indicate sequence-specific barcode-probing regions based on the *S. americanum* chromosome. Purple and blue signals indicate 5S rDNA and 45S rDNA, respectively. Scale bar, 10 µm. (d) Evolution tree constructed using synteny block analysis in Solanaceae family. The black number denotes the bootstrap of each node, and the red number indicates the estimated divergence time (myr) based on Ks values on the blocks. T means genome triplication event generating three genomic contents in *S. nigrum*.

interspecific crossing of *S. nigrum* with another ecotype *S. americanum* (SP2275) generated fertile F1 hybrids (Figure S1), suggesting that the two species were sufficiently close. To further confirm the evolutionary proximity, we constructed a phylogenetic tree based on the chloroplast cDNA sequences of selected Solanaceae species (Table S1). As shown in Figure 1b, *S. nigrum* and *S. americanum* were the closest relatives and are more distant from tomato (*S. lycopersicum*) and potato (*S. tuberosum*).

Since the whole-genome sequencing of *S. americanum* was completed at the chromosomal level (Lin et al., 2022), we attempted to verify the chromosomal similarity using chromosome-scale oligo-FISH (Fluorescence *in situ* Hybridization) (Waminal et al., 2018). We designed barcode probes to identify all 12 *S. americanum* chromosomes and performed FISH karyotyping of *S. nigrum* using these probes. FISH signals were spread throughout the 72 chromosomes in *S. nigrum* and there were no blank chromosomes (Figure 1c and Figure S2), indicating the practical utility of *S. americanum*-derived chromosome barcode libraries for *S. nigrum*. Eleven out of the 12 chromosome segments of *S. americanum* appear to remain intact in at least one set of *S. nigrum* chromosomes, and the remaining segments showed some variation, possibly resulting from either genomic rearrangement or with closely related ancestral species chromosomes during evolution (Figure 1c). The chromosome compositions of the two species were calculated and counted, showing three-fold differences, consistent with the ploidy level (Table S2). Taken together, the hexaploid *S. nigrum* may have evolved from its diploid ancestor, *S. americanum* acting as a suitable diploid standard for the polyploid black nightshade species study.

#### Genome assembly and gene annotation in *S. nigrum*

To obtain genomic information of hexaploid *S. nigrum* (accession NIBRG0000189638), we generated a sequencing read assembly of high-quality PacBio long reads (30 $\times$ ) and Illumina short reads (90 $\times$ ; Table S3, Figure S3). To increase the contig size, the initial genome was scaffolded using high-quality 10 $\times$  Genomics linked reads (49 $\times$ ; Table S3). The final *S. nigrum* consists of 8292 scaffolds with an N50 of 2.8 Mb and a total size of 2.9 Gb (Table S4). To perform a comparative analysis between *S. americanum* and *S. nigrum*, we re-annotated the *S. americanum* gene using the same annotation pipeline with *S. nigrum*. Combining gene evidence from *ab initio* predictions, ESTs, transcripts, and protein alignments, a total of 36 920 and 101 945 gene models in *S. americanum* and *S. nigrum*, respectively, were annotated with overall benchmarking universal single-copy orthologs (BUSCO) completeness score of 91.9% and 95.9% in *S. americanum* and *S. nigrum*, respectively (Table 1, Tables S5 and S6). The gene models were further validated using annotation edit distance (AED). The cumulative AED plots of both species showed high concordance between the gene prediction and transcript/protein evidence (Figure S4). Repetitive sequences comprised 79.1% and 80.0% of the *S. americanum* and *S. nigrum* assembly with long terminal repeat (LTR) retrotransposons as the most abundant class (Table S7). Moreover, 90% of the gene models of both species were functionally annotated using representative databases (Data S1). The complete process of genome sequencing, assembly, and functional annotation of the hexaploid *S. nigrum* is described in the supplementary information.

#### Genome evolution and triplication in *S. nigrum*

As we could not successfully establish the *S. nigrum* genome at the chromosomal level without chromosome conformation

**Table 1** Summary of genome assembly of *Solanum americanum* and *Solanum nigrum*.

	<i>S. americanum</i>	<i>S. nigrum</i>
Number of scaffolds	100	8292
Total length of scaffolds	1.0 Gb	2.9 Gb
N50 of scaffolds	86.1 Mb	2.8 Mb
Longest/shortest scaffold length	110.3 Mb/5.0 Kb	14.9 Mb/1.1 Kb
GC content	36.2%	36.3%
Number of genes	36 920	101 945
Average/total gene length	3.8 Kb/44.6 Mb	3.7 Kb/120.0 Mb
Average exon/intron length	268 bp/746 bp	266 bp/747 bp
Total length of repetitive sequences	0.8 Gb (79.1%)	2.3 Gb (80.0%)

capture, it was difficult to identify structural variations between the two species. Instead, we identified 225 synteny blocks from *S. nigrum* scaffolds by comparing the genomic sequences of the Solanaceae family to analyse the *S. nigrum* evolution. We calculated Ks values of the genes in the identified synteny blocks and estimated diversification timing (Figure 1d). We inferred that the *S. nigrum* genome underwent genome triplication after the diversification from potato and tomato (~13.95 Myr ago); possibly before *S. americanum* speciation was completed (range 2.88–6.96 Myr ago, Figure 1d and Table S8). We aligned all scaffolds of *S. nigrum* to each chromosome of *S. americanum* and verified that 1069 scaffolds of *S. nigrum* were aligned and covered most of the *S. americanum* genome (Figure S5). The read coverage depth of the aligned scaffolds on *S. americanum* chromosomes was nearly 3.0 (Figure S6). Together with the FISH results in Figure 1c, we can assume that most of the genome parts of *S. americanum* are conserved and triplicated in *S. nigrum* with some variation; thus, *S. nigrum* might have approximately three homoeologous genes orthologous to the genes in *S. americanum*.

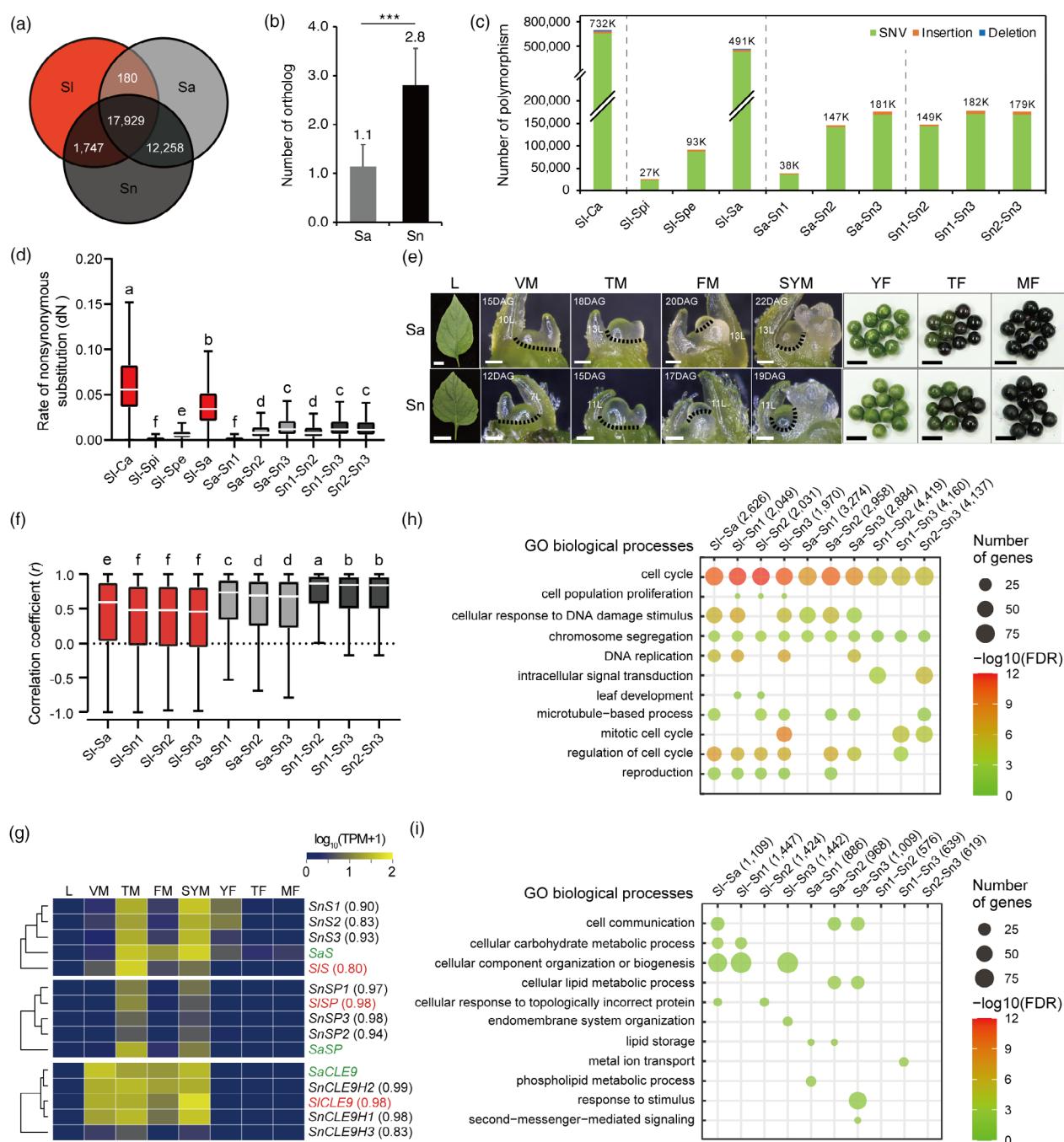
#### Identification and characterization of the orthologous gene groups in *S. americanum* and *S. nigrum*

To characterize homoeologous genes in *S. nigrum*, we defined orthologous gene groups using Ortholog MAtrix (OMA; Train et al., 2017) based on the tomato genome. A total of 17 929 hierarchical orthologous gene groups (HOGs) were shared gene groups among the three species, and 30 187 gene groups were found between *S. americanum* and *S. nigrum* (Figure 2a). Average gene numbers in HOGs were 1.1 and 2.8 in *S. americanum* and *S. nigrum*, respectively, which is consistent with the genome triplication of *S. nigrum* (Figure 2b and Figure S7). To investigate how well these orthologous genes were conserved, we analysed pairwise nucleotide sequence polymorphisms in various Solanaceae species (Figure 2c). Single nucleotide variations (SNVs) were mainly observed, and insertion and deletion events were rarely observed in any of the tested pairs. The number of polymorphic events was the lowest between *S. lycopersicum* and *S. pimpinellifolium*, the closest ancestor of tomato (Blanca et al., 2015), and the highest between *S. lycopersicum* and *C. annuum*, which reflects the evolutionary distance. As *S. nigrum* has orthologous gene triplicates, we mainly focused on the triads in *S. nigrum* and classified them into three homoeologous gene groups (Sn1, Sn2 and Sn3). Sn1 showed the least polymorphisms in the orthologs of *S. americanum* (Sa), implying that one set of homoeologous genes in *S. nigrum* might have been derived from *S. americanum*.

(Figure 2c). To determine the nucleotide polymorphism effects, we calculated the nonsynonymous substitutions (dN) rate in the protein sequences of each orthologous/homoeologous pair. Consistent with the polymorphism data, dN values were relatively low in homoeologous pairs of *S. nigrum* compared to those in inter-species pairs (Figure 2d). Although the dN/dS ratio could not be calculated, as the dS values were very low (near zero), the dN and dS plots further confirmed the evolutionary closeness of the intra-species protein functions (Figure S8).

We then investigated whether the expression patterns of orthologous genes were conserved. To test this hypothesis, we performed mRNA-Seq using tissue samples from various stages of *S. americanum* and *S. nigrum*. Despite differences in the actual

growth speed (Figure S9); the morphological phenotypes at the sampling stages were similar (Figure 2e). The mRNA-Seq reads were processed (Table S9) and validated using principal component analysis (PCA; Figure S10). We then calculated the correlation coefficients ( $r$ ) of spatiotemporal expression profiles of orthologous genes pairwise together with public tomato-gene expression data (Lemmon et al., 2016; Sato et al., 2012). The correlations of expression patterns in *S. nigrum* homoeologous gene pairs were higher than any other orthologous pairs (Figure 2f), suggesting that the expression regulation of triplicated homoeologous genes was highly conserved within the species. Interestingly, even though the number of polymorphisms in Sn1–3 homoeologous gene pairs was similar to that



**Figure 2** Characterization of the orthologous genes in *Solanum* species. (a) Venn diagram describing hierarchical orthologous groups (HOGs) among *Solanum lycopersicum* (Sl), *S. americanum* (Sa) and *S. nigrum* (Sn). (b) Average orthologous gene numbers included in common HOGs in *S. americanum* and *S. nigrum* (mean  $\pm$  SD). Statistical significance was obtained using a two-tailed Mann–Whitney *U* test ( $***P < 0.001$ ). (c) Pairwise comparison of nucleotide polymorphisms among pepper (Ca, *C. annum*), tomatoes (Sl, *S. lycopersicum*; Spi, *S. pimpinellifolium*; Spe, *S. pennellii*), *S. americanum* and *S. nigrum*. Sn1, Sn2 and Sn3 refer to each orthologous gene existing as a triad in *S. nigrum*. Pepper and wild tomato species were used as inter-genus and inter-species controls, respectively.  $n = 5257$  HOGs. (d) Pairwise comparison of amino acid differences through nonsynonymous substitution rate (dN) calculation among species used in (c).  $n = 7473$  HOGs. (e) Tissue samples from leaves, shoot apical meristems, and fruits for transcriptome profiling in *S. americanum* and *S. nigrum*. DAG, days after germination; FM, floral meristem; L, leaf; MF, mature fruit; SYM, sympodial shoot meristem; TF, transition fruit; TM, transition meristem; VM, vegetative meristem; YF, young fruit. Dotted lines indicate excision sites in meristem samples. Scale bar, 1 cm (leaf and fruits) and 100  $\mu$ m (shoot apical meristems). (f) Pairwise analyses of Pearson's correlation coefficients of gene expression patterns using eight tissue samples among the three species in (e). For each box plot, the lower and upper bounds of the box indicate the first (Q1) and third (Q3) quartiles, respectively, and the centre white line indicates the median. Outliers determined by Tukey's HSD are not shown. Different letters indicate statistical significance obtained using Kruskal–Wallis test with Dunn's multiple comparisons.  $n = 7473$  HOGs. (g) Heat map represents gene expression patterns of three selected individual genes in three species. Numbers in parentheses indicate correlation coefficients compared to an expression pattern in *S. americanum* (green colour). Red and black colours indicate genes from *S. lycopersicum* and *S. nigrum*, respectively. (h, i) Enriched GO terms of orthologous gene set with either highly correlated expressions ( $H$ ,  $r \geq 0.8$ ) or noncorrelated expressions ( $I$ ,  $-0.2 < r < 0.2$ ) from data in (f). Statistical significance for differences was obtained using Fisher's exact test with false discovery rate (FDR)  $<0.05$ . Numbers in parentheses indicate the numbers of genes in each set.

in Sa (Figure 2c,d), expression correlations were much higher in Sn homoeologous genes. To test whether this was due to algorithmic bias from high sequence similarity, we manually analysed the RNA-seq results using uniquely mapped reads. The expression profiles of the unique reads were comparable to the *in silico* results (Figure S11), therefore, strong correlations among homoeologous gene expression are characteristic of polyploid gene expression, probably because of conserved promoter sequences in the same plant (Figure S12). To further examine the expression correlations of individual genes of interest, we selected three orthologous genes to tomato *S*, *SP* and *CLE9* which exist as triads, and one orthologous to tomato *CLV3* which is a quartet in *S. nigrum*. Since these genes play essential roles in the development and maintenance of shoot meristems in tomato (Kwon et al., 2022; Park et al., 2014; Soyk et al., 2017), their expression patterns are highly conserved in *Solanum* species (Figure 2g). We validated the expression profiles of the genes of interest using qRT-PCR, followed by targeted sequencing, revealing similar expression patterns (Figure S11). The spatial expression of *S* and *SP* orthologs in *S. americanum* was similar to *S. lycopersicum*, as confirmed by RNA *in situ* hybridization (Figure S13). Taken together, these results suggest that essential genes involved in growth and development may be functionally conserved. These findings led us to investigate characteristics of highly correlated genes ( $r \geq 0.8$ ) and noncorrelated genes ( $-0.2 < r < 0.2$ ) through Gene Ontology (GO) analysis. GO terms related to the cell cycle and reproduction were enriched in highly correlated gene sets (Figure 2h), whereas no significant enrichment of GO terms was observed in a relatively small number of noncorrelated genes (Figure 2i).

#### CRISPR-Cas9-mediated gene editing enables quantitative genetic variation in hexaploid genome

Based on the finding that the genome is triplicated and that most of the orthologs exist as triads in hexaploids, we hypothesized that if we engineer genes determining quantitative traits for agricultural properties, we could generate more variable genotypes showing a broader range of agricultural phenotypes in hexaploids than in diploids. Given that *S*, *SP*, *CLV3* and *CLE9* genes are well characterized and are known to regulate fruit development and yield in *S. lycopersicum* (Kwon et al., 2022;

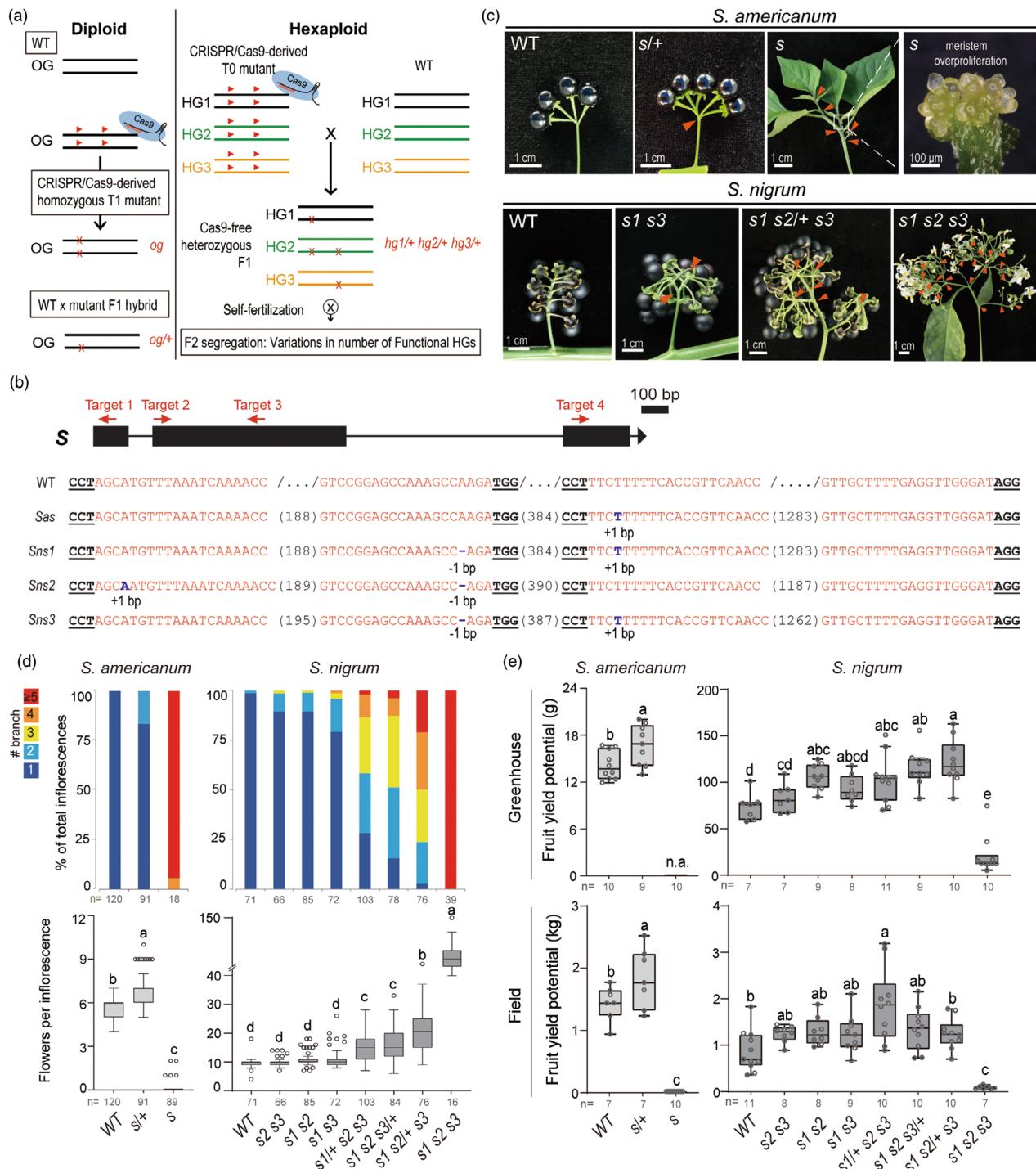
Park et al., 2014; Soyk et al., 2017), we focused on these genes in *S. americanum* and *S. nigrum*. According to the phylogenetic trees of these gene products, the black nightshade ortholog seemed to be diverse from the tomato orthologs, and one *S. nigrum* ortholog was closely related to *S. americanum* while the others were more diverged (Figure S14), further supporting the genome triplication timing after diversification. We performed CRISPR-Cas9-mediated mutagenesis using four sgRNAs targeting each gene and analysed the generated mutations in the  $T_0$  transgenic population. We thus observed more mutations in hexaploid *S. nigrum* than in diploids because of the increased ortholog (homoeologs), although the type of mutation was not significantly altered (Figure S15). Visible phenotypes driven by targeted mutagenesis were classified into three types: WT-like type 1, moderate type 2 and null mutant-like type 3 in the  $T_0$  populations. Unlike the diploid mutants showing mostly either type 1 or type 3 mutations, the hexaploid mutants predominantly showed type 1 and 2 mutations (Figure S16). These data suggest that polyploid plants have better compensatory mechanisms than diploid plants because of their increased copy numbers of functionally conserved orthologous genes. We also confirmed that each one of *S* or *SP* genes in *S. nigrum* triads could completely restore the triple knockout mutant phenotypes through complementation analyses (Figure S17). Therefore, we might be able to establish genetically optimized plants with desired agricultural traits by exploiting the genetic redundancy of polyploid crops.

#### A series of *s* mutants exhibit cumulative dosage effects in hexaploid *S. nigrum*

To generate a series of mutants with varying numbers of functional homoeologous genes, the selected CRISPR-Cas9-mediated knockout mutant was backcrossed with the wild type. We eliminated Cas9 transgenes, confirmed genotypes in heterozygous F1 plants and obtained a series of mutants among the F2 and F3 segregants (Figure 3a). *S* gene encodes a WUSCHEL-RELATED HOMEOBOX 9 (WOX9) transcription factor that regulates inflorescence meristem maturation, and the *s* null coding sequence mutant shows embryonic lethality in tomato (Hendelman et al., 2021; Park et al., 2012). In addition, the evergreen (evg/Phwox9) mutant developed vegetative meristems

instead of floral meristems and failed to generate flowers in petunia (Rebocho *et al.*, 2008). Similar to the *evg* mutant, we frequently observed mutants showing vegetative inflorescence phenotypes in  $T_0$  transgenic populations among the *s* mutants of *S. americanum* (Figure 3b,c). Magnified inflorescence meristems showed an over-proliferating meristem phenotype (Figure 3c),

similar to that of the tomato *s* promoter mutants found in a previous report (Rodríguez-Leal *et al.*, 2017), although they still could not develop any normal fertile flowers. To retain the *s* mutant allele in *S. americanum*, we selected one with a lower (~50%) mutation rate at  $T_0$  for backcrossing. In contrast, hexaploid *S. nigrum* has three *S* genes, and we selected two



**Figure 3** Functional gene copy numbers of *S* homoeologous genes determine inflorescence branching and fruit yield in *S. nigrum*. (a) Schematic diagram explaining CRISPR-Cas9-generated mutants by backcrossing in *S. americanum* and *S. nigrum*. OG, orthologous gene; HG, homoeologous gene. (b) CRISPR-Cas9-generated mutations in *S* gene loci. *S* gene structure and four sgRNA targeting sites are depicted in the upper diagram. sgRNA-targeted sequences are red-coloured, and each protospacer-adjacent motif (PAM) sequence is underlined. Parentheses number represents gap lengths between sgRNAs. Each mutation is denoted in blue colour. (c) Inflorescence branching of the *s* mutants in *S. americanum* and *S. nigrum*. Red arrowheads indicate branching points in the inflorescence. The magnified picture of emerging inflorescence in the *s* mutant shows meristem overproliferation phenotype. The *s1s2s3* triple mutant rarely develops mature black fruits from flowers. (d) Quantifications of inflorescence branching and flower numbers of the *s* mutants in *S. americanum* and *S. nigrum*. Each inflorescence branch number is counted and denoted using different colours in the stacked bar chart. Each inflorescence flower number is measured and depicted in the box plot. *n*, number of inflorescences. Genotypes are ordered to show gradual phenotypic effect changes. Open circles represent data values of outliers. (e) Estimated fruit yield potential of various *s* mutant combinations in *S. americanum* and *S. nigrum*. Total yield potential is calculated based on the average weight of each inflorescence and the total number of maturing inflorescences in each plant grown in the greenhouse (upper) or field (lower). *n*, number of plants. Open circles represent all data values. (d, e) Box plots show the 25th, 50th and 75th percentiles and different letters indicate statistically significant differences (ANOVA, Tukey's HSD,  $P < 0.05$ ).

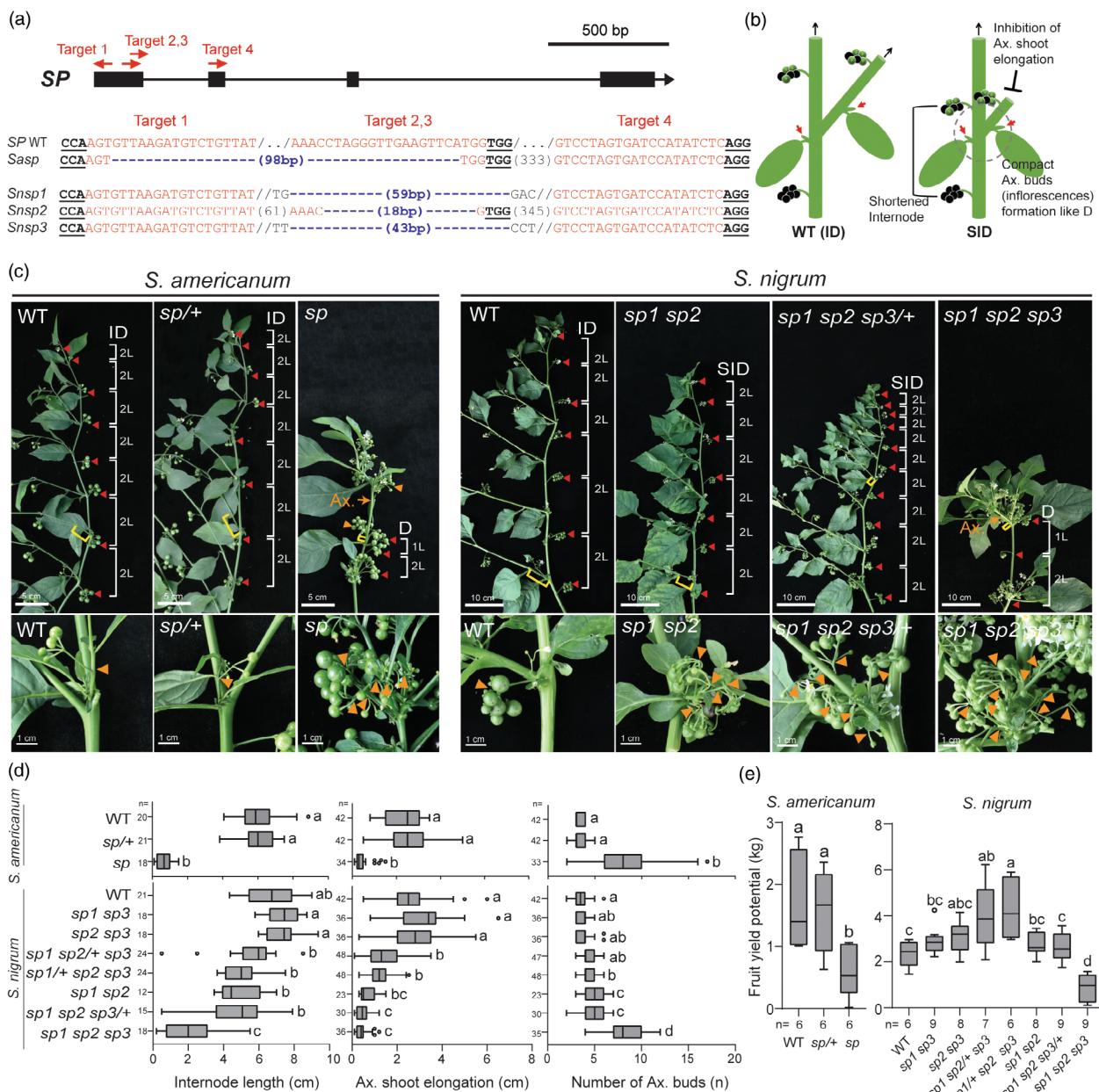
different allele sets of the *s1s2s3* triple mutant in the  $T_1$  generation (Figure 3b and Figure S18). Interestingly, we were able to harvest normal seeds from fruits that developed in the relatively weak evg-type *S. nigrum*. Severe *s1s2s3* triple mutants, in which all *S* protein homeodomains (HDs) were disrupted (Figure S19) showed some bract-like organs in the inflorescences (Figure 3c) but still developed normal flowers and fruits that produced viable seeds.

As the heterozygote *s*/+ shows an intermediate inflorescence branching phenotype, resulting in a moderately increased fruit yield in tomato (Soyk et al., 2017), the number of functional *S* gene copies might determine quantitative traits. Thus, we tested whether the three *S* homoeologous genes in the *S. nigrum* genome could have an advanced dosage effect with a broader range of quantitative traits than in diploids. While the *s* mutant showed extremely branched inflorescences, the heterozygote *s*/+ plant showed only a few branched inflorescences in *S. americanum*, resulting in slightly increased flower numbers compared to the WT (Figure 3c,d). In *S. nigrum*, however, we observed quantitative phenotypic variation between the WT and the *s1s2s3* triple mutant. Moderately increased branch numbers (2–3 branches) in each inflorescence were observed in the double mutants (e.g. *s1s3*), and further increases (3–5 branches) were observed in the higher-order mutants (e.g. *s1s2/+s3*). With an increase in inflorescence branches, flower numbers gradually increased in multiple-order *s* mutants of *S. nigrum* (Figure 3c,d). Since the flower number directly correlates with the number of fruits, we measured the consequent fruit yield in the *s* mutants. Under greenhouse conditions supplied with sufficient nutrients, the fruit yield potentials of six multiple-order *s* mutants were elevated as the flower numbers increased in *S. nigrum* (Figure 3e and Figure S20). Among them, the *s1s2/+s3* mutant was selected as the best yield-optimized genotype, and the fruit yield increased almost twice as much as that of the WT. The *s1s2s3* mutant displayed extremely branched inflorescences and increased flowers; however, we could harvest only a small number of fruits due to the low fruit-setting ratio (Figure 3e). We also measured fruit yields in the field, which were similar but slightly different from the greenhouse results, where *s1/+s2s3* produced the highest fruit yield (Figure 3e). Despite the increase in fruit quantity, the quality of fruits, such as 10-fruit weight and Brix did not significantly change in the best *s* mutants (Figure S21). These variations in phenotype and yield performance were not successfully implemented in diploid *S. americanum* because there was only a single *S* ortholog for

this trait (Figure 3d,e). In conclusion, we precisely generated inbred fruit yield-optimized *S. nigrum* by editing homoeologous *S* genes and screening for genotypes with varying functional *S* copy numbers.

#### Mutagenesis in *SP* homoeologous genes results in quantitative variation affecting fruit yield in hexaploid *S. nigrum*

The tomato *sp* knockout mutant exhibits determinate sympodial shoot growth (D) (Pnueli et al., 1998). The wild-type *SP* allele is predominantly functional; thus, tomato *sp*/+ shows a WT-like indeterminate (ID) phenotype in shoot structure, indicating no dosage sensitivity under heterozygote mutant (Thouet et al., 2008). To test whether this functional mechanism of the *SP* gene is similarly implemented in black nightshade species, we generated a variety of *sp* mutants in *S. americanum* and *S. nigrum* using the CRISPR-Cas9 system with the same scheme used to establish the *s* mutants (Figure 3a). All generated mutant alleles had a large deletion between sgRNA targets 1 and 3, possibly encoding nonfunctional *SP* proteins (Figure 4a and Figure S22). The *S. americanum* *sp* and *S. nigrum* *sp1sp2sp3* mutants displayed D-type shoot structures, and *sp*/+ heterozygote showed a WT-like ID phenotype in *S. americanum*, confirming the functional conservation of *SP* genes; these phenotypes were not allele-specific (Figure 4b,c and Figure S23). Intriguingly, however, we observed an intermediate shoot growth phenotype between ID and D in the *sp1sp2* double- and higher-order mutants of *S. nigrum* unlike in other diploid species (Figure 4b,c). These plants exhibited notable characteristics of shortened internode length, inhibited axillary shoot elongation and inflorescence bundle emergence composed of dense axillary buds; thus, we defined this intermediate shoot growth pattern as the semi-indeterminate (SID) type (Figure 4c). We carefully measured each characteristic of the SID phenotypes in a series of *sp* mutants of *S. nigrum* and identified a new scope for quantitative traits related to *SP*-mediated shoot growth (Figure 4b). In summary, the internode lengths and first internode lengths of the axillary shoots gradually decreased, and the axillary buds in each inflorescence bundle gradually increased as the functional *SP* gene copies decreased according to the mutant combinations (Figure 4d). The *sp1sp2* double mutant and *sp1sp2sp3*/+ mutant showed the most severe phenotypic effects, closest to D-like growth, suggesting that *SP3* might be the least functional among the three homoeologous genes. Indeed, the expression level of *SP3* was relatively low

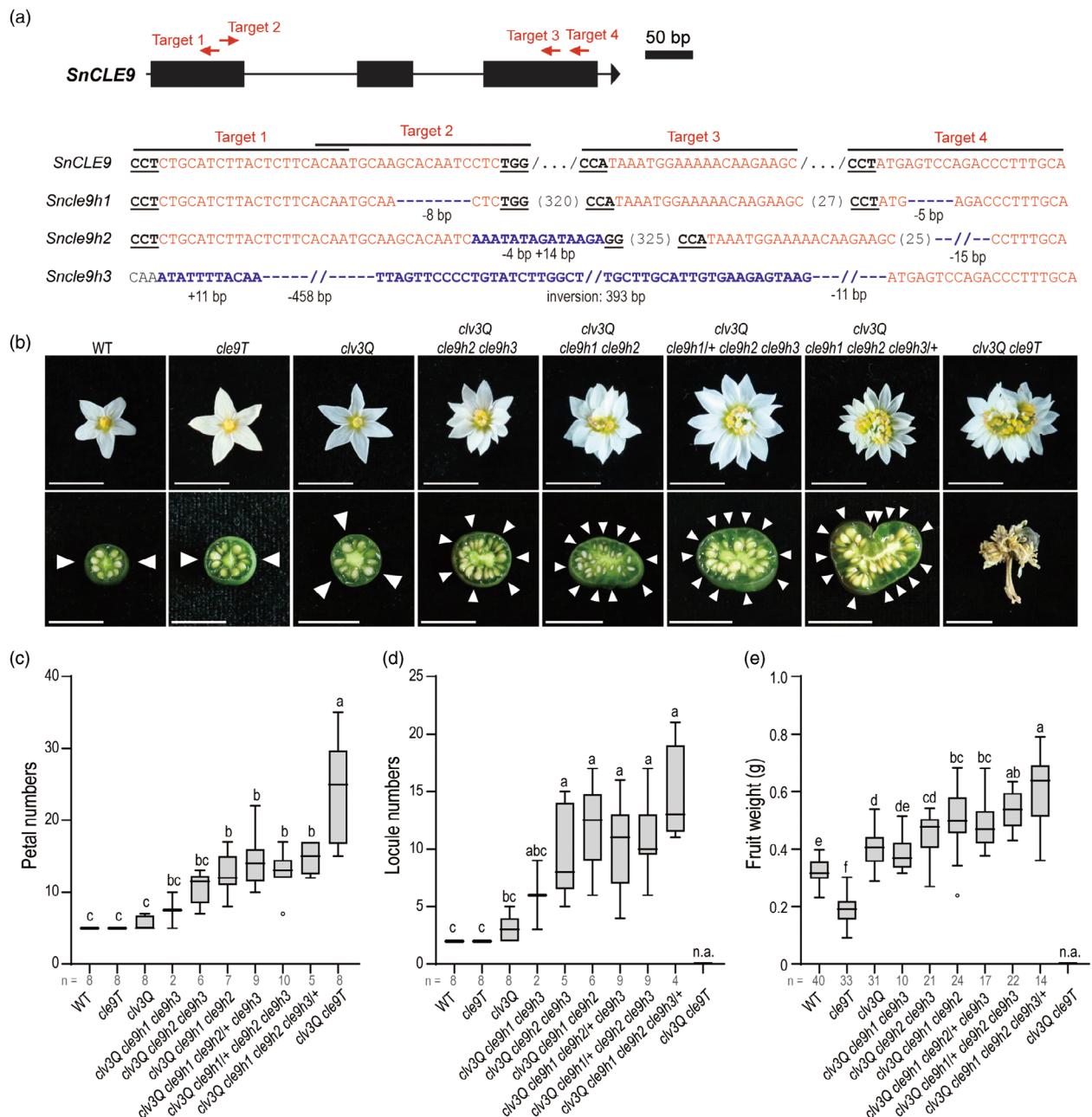


**Figure 4** Mutant series possessing various copy numbers of functional *SP* homoeologous genes exhibits altered shoot growth patterns affecting fruit yield in *S. nigrum*. (a) CRISPR-Cas9-generated mutations in *SP* gene loci. *SP* gene structure and four sgRNA targeting sites are depicted in the upper diagram. sgRNA and PAM sequences are highlighted in red and bold underlined, respectively. Each mutation is denoted in blue colour. (b) Shoot structure of SID (semi-indeterminate)-type *S. nigrum* *sp* mutants. Black arrows indicate stem growth directions, and red arrows indicate emerging axillary inflorescences. ID, indeterminate growth; D, determinate growth. (c) Shoot growth patterns (top) and axillary bud formations (bottom) of the *sp* mutants in *S. americanum* and *S. nigrum*. Red arrowheads indicate main stem inflorescences, and orange arrowheads indicate axillary stem inflorescences. Yellow brackets indicate elongated internode lengths of the axillary shoots. Sympodial indices are denoted with leaf numbers between two adjacent inflorescences. Axe. indicates an emerged axillary branch. (d) Quantification of the shoot growth patterns and axillary bud formations of the *sp* mutants in *S. americanum* and *S. nigrum*. Internode lengths (left), axillary shoot elongations (middle) and numbers of emerged axillary inflorescences (right) were measured in each genotype. Genotypes are ordered to show gradual changes in phenotypic effects. *n*, number of tested shoots. (e) Estimated fruit yield potential of various *sp* mutant combinations in *S. americanum* and *S. nigrum*. Total yield potentials were calculated based on the average weight of each inflorescence and the total numbers of maturing inflorescence in each plant. Genotype order is as same as (d). *n*, number of plants. (d, e) Box plots depict the 25th, 50th and 75th percentiles and different letters indicate statistically significant differences (ANOVA, Tukey's HSD,  $P < 0.05$ ). Open circles represent outlier data values.

(Figure S11), and the finding that the highly conserved D residue was changed to N specifically in SnSP3 (Figure S22), further supporting our hypothesis. Other phenotypes, such as sympodial indices and flowering time, were comparable to those of the WT

(Figure S24), suggesting that the SID plants undergo ID-like growth.

To investigate whether this variety of SID phenotypes influenced fruit yield, we measured the yield potential of sp



**Figure 5** Dosage effects on paralogous compensation by functional *CLE9* homoeologous genes in flower and fruit development in the *clv3Q* mutant of *S. nigrum*. (a) CRISPR-Cas9-generated mutations in *CLE9* homoeologous gene loci. *CLE9* gene structure and four sgRNA targeting sites are depicted in the upper diagram. sgRNA-targeted sequences in red, and each PAM sequence is underlined. Each mutation is denoted in blue colour. (b) Reproductive organ phenotype of the *cle9* mutants in *S. nigrum*. Morphology of flowers (upper) and cross-section of the immature fruit (lower) showing locule photographed. White arrowheads indicate each fruit locule. The *clv3Q cle9T* septuple mutant shows severely fasciated flower that fails to develop as a fruit. Scale bar, 1 cm. (c) Petal number quantification in *cle9* mutant flowers from *S. nigrum*. *n*, number of plants. (d) Locule number quantification of *cle9* mutant fruits in *S. nigrum*. *n*, number of plants. (e) Fruit weight quantification of the *cle9* mutants in *S. nigrum*. *n*, number of fruits. (c–e) Box plots depicting the 25th, 50th and 75th percentiles and different letters indicate statistically significant differences (ANOVA, Tukey's HSD,  $P < 0.05$ ). Genotypes were ordered to show gradual changes in phenotypic effects. Open circles represent outlier data values.

mutants in the field. Fruit yields of the SID-type *S. nigrum* sp mutants mostly increased compared to those of the WT and were maximized in the *sp1/+sp2sp3* and *sp1sp2/+sp3* genotypes (Figure 4e). The increase in fruit yield was mainly due to the increase in inflorescence numbers due to structural changes in each mutant, as plant height and fruit weights did not change (Figure S25). Although the D-type *sp1sp2sp3* mutant

displayed the most compact shoot structure and produced larger fruits, fruit yield was largely reduced owing to the early growth termination (Figure 4e and Figure S25). Thus, we conclude that mutations in SP could also cause dosage effects on plant compactness in hexaploid *S. nigrum*, and by screening various genotypes, a yield-maximizing plant could be obtained.

## Paralogous compensation regulations of CLV3-CLE9 determine the optimal fruit size of *S. nigrum*

*CLV3* gene encodes a small peptide that inhibits stem cell proliferation in the shoot apical meristem, and the conserved WUS (WUSCHEL)-CLV3 negative feedback regulation plays a central role in meristem maintenance in many plant species (Somssich *et al.*, 2016). Loss of tomato *S/CLV3* results in enlarged and fasciated meristematic organs, makes bigger fruits (Rodríguez-Leal *et al.*, 2017; Xu *et al.*, 2015) and diverse paralogous compensation mechanisms involving CLE9 have evolved in the Solanaceae family and influenced phenotypic variations (Kwon *et al.*, 2022). To investigate CLV3-CLE9-related paralogous compensation in hexaploid *S. nigrum*, we generated mutant plants using CRISPR. Four homoeologous genes of *CLV3* in *S. nigrum* were simultaneously targeted by the CRISPR-Cas9 system, and a transgene-free *clv3*-quadruple (*clv3Q*) mutant was obtained in the *T*<sub>2</sub> generation (Figure S26). Unlike tomatoes (Rodríguez-Leal *et al.*, 2017), the *clv3Q* mutant did not show fasciated flowers or fruit organs, suggesting a paralogous compensation mechanism in *S. nigrum*. Thus, we further edited three *CLE9* homoeologous genes in the *clv3Q* background, isolated the *clv3Q cle9T* (*cle9-triple*) septuple mutant and crossed it with the parental *clv3Q* mutant to generate mutants with varying functional *CLE9* homoeologous (*CLE9H*) gene copy numbers (Figures 3a and 5a). The *cle9h2* mutation showed severe truncation even at the transcription start site, resulting in a complete knockout allele, while *cle9h1* and *cle9h3* mutations produced prematurely stopped proteins of each *CLE9*, which were possibly nonfunctional (Figure S27). Similar to the tomato *clv3cle9* double mutant displaying severely fasciated stems and defects in floral organ development (Kwon *et al.*, 2022), the *clv3Qcle9T* mutant developed fasciated flowers with significantly increased petal numbers and failed to develop a normal fruit from the primary inflorescence in *S. nigrum* (Figure 5b,c). Another mutant with a different mutant allele showed a similar phenotype (Figure S28).

Similar to *S* and *SP*, the functional gene number of *CLE9* also generated genetic dosage effects on the reproductive organ development in *S. nigrum*. Quintuple mutants with only one *CLE9H* functional gene (e.g. *clv3Qcle9h1cle9h2*) showed significantly increased petal and locule numbers, and higher-order mutants (e.g. *clv3Qcle9h1cle9h2cle9h3/+*) displayed more

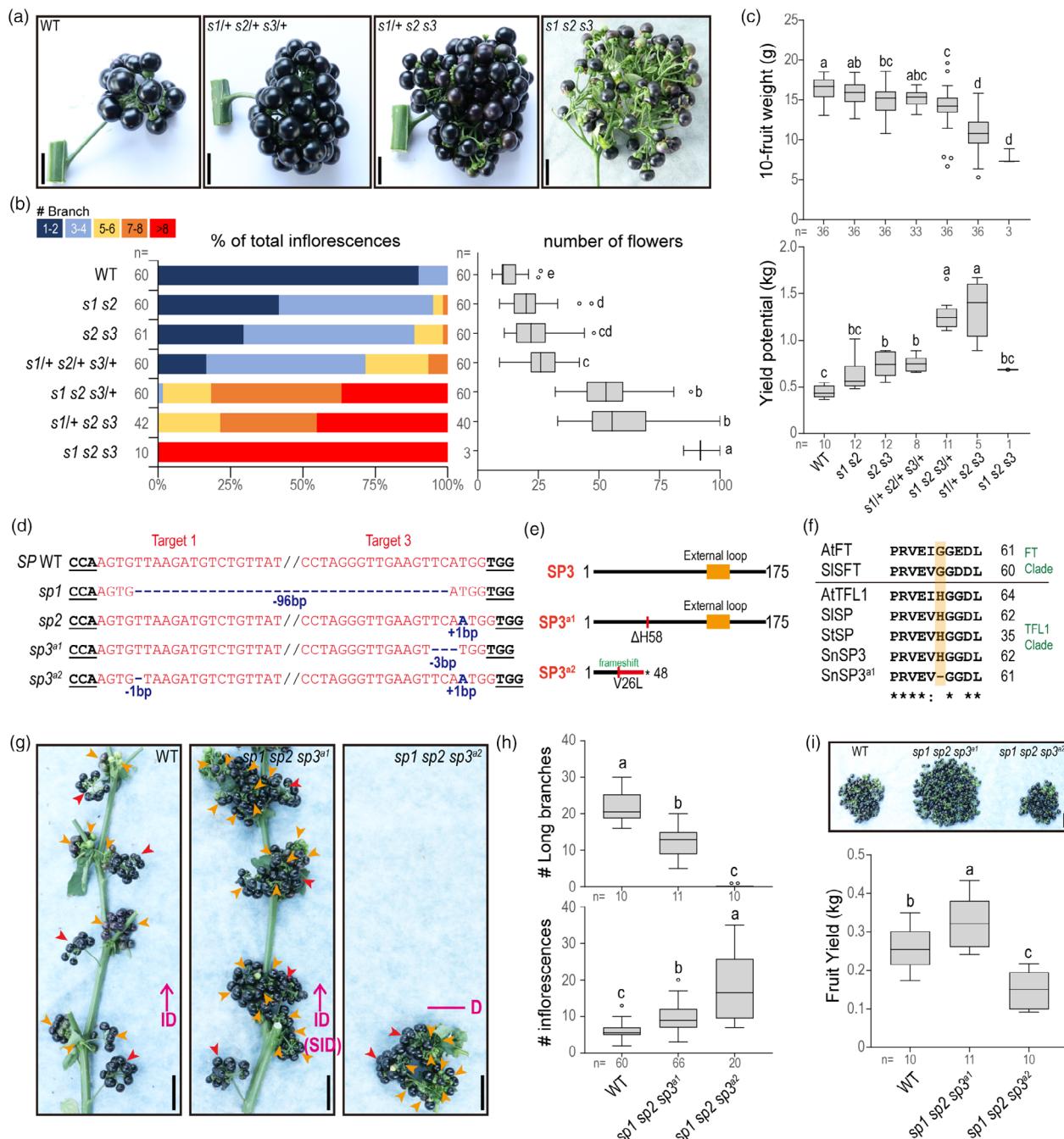
fascinated flowers and fruits with increased petals and locules compared to WT, *clv3Q* or *cle9T* (Figure 5b-d). Consequently, fruit weights increased in multiple mutants (Figure 5e), suggesting that we could optimize fruit size by modulating functional *CLV3* and *CLE9* genes. Therefore, engineering homoeologous genes related to paralogous compensation can generate a range of quantitative traits that can be optimized in polyploid organisms.

## Boranong, an orphan crop of *S. nigrum*, can be rapidly improved in fruit yield through CRISPR-Cas9-mediated mutagenesis

Boranong is a regional cultivar of *S. nigrum* generated by crossing the Korean native cultivar Naju and the Chinese cultivar MPR-0156 by Korean breeders (see Methods). Boranong is a large and firm fruit with high anthocyanin levels, which can be a suitable material for tinted cosmetics, in addition to its usage as a food supplement with benefits in the harvest. To improve the Boranong yield, we applied the same strategy to generate genotype variations in *s* and *sp* mutants by CRISPR-Cas9-mediated mutagenesis and subsequent backcrossing. The *s* mutations had deletions in the sgRNA target-2 region of all three *S* genes (Figure S29), and possibly null alleles. Similar to *S. nigrum* data (Figure 3), the higher-order *s* mutants of Boranong, including *s1/+s2/+s3/+* showed increased inflorescence branches, flower and fruit numbers in each inflorescence (Figure 6a,b). The *s1/+s2s3* mutant displayed grape-like fruits (Figure 6a) owing to increased inflorescence branches, and, together with the *s1s2s3/+* mutant, the total fruit yield of these mutants was significantly high (Figure 6c). Although fruit size decreased as inflorescence branching increased (Figure 6c), the number of fruits alone increased the yield potential of each plant. Despite the tremendous number of flowers in the *s1s2s3* triple mutant, the low fruit-setting ratio and smaller fruit size caused low productivity compared to the yield-maximized mutants.

The *sp* knockout mutants of Boranong showed distinct features compared to *S. nigrum* (Figure 4). The *sp1sp3* double mutant also exhibited a D phenotype similar to that of *sp1sp2sp3*. However, after the termination of the first flower pair, SID-like side shoots developed, resulting in compact fruit bundles (Figure S30). This caused *sp1sp3* plants to produce more fruit with a D-type structure. Additionally, we found the *sp3<sup>a1</sup>* mutation, a weak mutant allele in the *SP3* locus, with a 3-bp

**Figure 6** Fruit yield improvement by polyploid mutagenesis in *S. nigrum* cv. Boranong. (a) Inflorescence phenotypes in selected *s* multiple mutants in *S. nigrum* cv. Boranong. Scale bar, 2 cm. (b) Quantification of inflorescence branching and flower numbers of various *s* mutant combinations in Boranong. Branch numbers in each inflorescence were counted and grouped using different colours in the stacked bar chart (left). Flower numbers of each inflorescence measured (right). *n*, number of inflorescences. (c) Fruit yields of Boranong *s* mutants. The 10-fruit weight of each genotype was measured using fully matured black fruits (upper). Total yield potentials were calculated based on the average weight of each inflorescence and the total number of maturing inflorescences in each plant (lower). (d) CRISPR-Cas9-generated *SP* gene loci mutation in Boranong. sgRNA-targeted sequences are red-coloured, and each PAM sequence is underlined. Each mutation is denoted in blue colour. (e) Proteins encoded by *SP3* alleles. Single amino acid (H58) deletion occurred by *sp3<sup>a1</sup>* mutation, *SP3<sup>a2</sup>* severely impaired due to the frameshift mutation. The asterisk indicates a premature stop. (f) Peptide sequence alignment of *SP* proteins from various species. Histidine residue is conserved only in the TFL1 clade and missing in the *SP3<sup>a1</sup>* of Boranong (yellow box). (g) Sympodial shoot structures and fruit formations of the *sp* allele-specific mutants in Boranong. Red arrowheads indicate main stem inflorescence, and orange arrowheads indicate axillary bud inflorescence. D, determinate growth; ID, indeterminate growth; Scale bar, 5 cm. (h) Quantification of shoot branching and inflorescence development in axillary buds. Numbers of the long axillary branches ( $\geq 15$  cm) measured in each plant (upper). Emerged inflorescence numbers of each fruit bundle (lower). (i) Fruit yield of the *sp* allele-specific mutants in Boranong. The weight of fully matured black fruits are measured in each plant at the same stage. Representative image of collected black fruits from a single plant (upper inbox; Scale bar, 5 cm). *n*, number of plants. (b, c, h and i) Box plots depict the 25th, 50th and 75th percentiles and different letters indicate statistically significant differences (ANOVA, Tukey's HSD,  $P < 0.05$ ). Open circles represent outlier data values.



deletion which encodes 1-a.a.-deleted SP3 protein (Figure 6d,e). *SP* is homologous to the *Arabidopsis* antiflorigen *TERMINAL FLOWER1* (*TFL1*) and antagonistic to the florigen *FLOWERING LOCUS T* (*FT*). The H58 residue is conserved in the *TFL1* proteins of dicots, but not in *FT* proteins (Figure 6f), suggesting that this residue might be an additional important amino acid for the function of *TFL1*s, but not *FT*s. The *sp1sp2sp3<sup>a1</sup>* mutant only expresses *SP3ΔH58* protein as a functional *SP*, resulting in compact SID shoot structures (Figure 6g,h). As side shoot elongation was strongly inhibited in the *sp1sp2sp3<sup>a1</sup>*, the numbers of long side branches ( $\geq 15$  cm) were greatly reduced compared to WT, which produced large fruit bundles alongside the main stems (Figure 6g,h). This is literally the 'self-pruning' effect, thus growing this mutant has the benefits of reducing

labour and dense greenhouse cultivation. In addition, similar to other SID-type *sp* mutants of *S. nigrum*, the *sp1sp2sp3<sup>a1</sup>* mutant showed increased fruit yield under greenhouse conditions (Figure 6i). As ID- or SID-type plants continue to produce new fruits over time, we harvested fruits at the same time after growing the plants with proper pruning. In summary, using gene editing, we engineered *SP* homologous genes of *Boranong* and *sp1sp2sp3<sup>a1</sup>* mutants resulting in useful agricultural traits in a relatively short period.

## Discussion

In this study, we compared the hexaploid *S. nigrum* with its diploid relatives *S. americanum* and *S. lycopersicum* using

genome engineering and crop optimization. We assembled the ~3 Gb genome of hexaploid *S. nigrum* at the scaffold level with 2.8 Mb of N50 and defined homoeologous genes based on comparisons of orthologs with tomato and *S. americanum* with valid expression profiles. We generated a series of QT-related gene mutants using the CRISPR-Cas9 system with various molecular dosages of the target genes. Through phenotyping of the mutants, we isolated yield-maximized black nightshade plants in the hexaploid *S. nigrum* due to the broader spectrum of phenotypic variations in the polyploid crop. Furthermore, we applied our strategy to a black nightshade variety, Boranong, and successfully produced genetically improved plants with a dramatically increased fruit yield. Although the *S. nigrum* genome is still at the scaffold level and the origins of each chromosome have not been identified, we were able to define and edit homoeologous genes to regulate agricultural traits. Thus, our approach could be useful for engineering other polyploid crop species for which genomic information has not yet been established.

By taking advantage of the functional redundancy among homoeologous genes in *S. nigrum*, we discovered 17 929 sets of genes that could partially compensate for each other within a set. We isolated a variety of mutant combinations that showed a broad range of intermediate traits by editing homoeologous genes. Specifically, we targeted *S*, *SP* *CLV3*, and *CLE9* genes, which resulted in improved plant structures. These observations provide three considerations for engineering polyploid genomes. First, we suggest that improving plant architecture by manipulating functional redundancy among homoeologous genes could be applied to other major crops. In hexaploid wheat, *TaGW2* homoeologous genes affect grain weight and are generally increased in mutants, especially in the lines lacking two homoeologous out of three (Zhang *et al.*, 2018). Secondly, homoeologous gene editing can improve and optimize disease resistance in polyploid crops. This has been demonstrated to be effective by knocking out three *MLO* homoeologous in wheat (Wang *et al.*, 2014), and by combining mutations in *SHATTERPROOF1/2* homoeologous in *Brassica napus* (Zaman *et al.*, 2021). Thirdly, our approach can generate phenotypic diversity and improve the metabolite content of useful fruits in polyploid crops. For instance, in the hexaploid oilseed crop *Camelina sativa*, combining different alleles of *FAD2* generated a diverse range of lipid profiles in the oil, allowing for an unbiased gene dosage analysis and providing a source of plant breeding genetic variability (Morineau *et al.*, 2017).

Our study also provides an important strategy for manipulating functional redundancy among paralogs to achieve desired crop traits. Paralogs are genes that arise through duplication events and have similar or overlapping functions, such as homoeologous genes of polyploids. For example, it has been reported that expression regulation or copy number variation of *CLV3* and *CLE9* quantitatively regulates the meristem size in the Solanaceae family (Kwon *et al.*, 2022). Here, we generated various mutant combinations of paralogs and homoeologs according to floral organ number development and optimized fruit size by modulating the functional *CLV3* and *CLE9* genes. In flowering crops such as tomato and rice, combining genetic variations in paralogs of *FT* can induce flowering at an optimal time during a given season and increase grain yield (Andrés and Coupland, 2012). In rice, *Sc-j* allele contains a pollen-essential gene, whereas the *Sc-i* allele contains two or three tandem-duplicated ~28-kb segments. Knocking out one or two of the three *Sc-i* copies rescued *Sc-j*

expression and male fertility (Shen *et al.*, 2017). Moreover, soybean, maize, and tobacco, which underwent diploidization after genome duplication, possess redundant paralogs and are crops that can be genetically applied to our technology. For example, *GmFT2a* and *GmFT2b* redundantly act as florigens (Nan *et al.*, 2014), and mutant combinations or heterozygosity of the *Dt1* and *Dt2* genes that regulate stem termination display intermediate shoot growth patterns in soybean (Bernard, 1972). Taken together, engineering of paralogs or homoeologs can be extremely useful in discovering novel quantitative phenotypes that have not been seen before by increasing phenotypic diversity.

In this study, we verified the dosage effects of functional homoeologous expression by editing coding sequences and found that phenotypic variations of the mutants ranged from 0 (KO) to 100% (WT). To create dosage variations in gene expression, researchers usually try to edit *cis*-elements of the genes, especially promoters. In tomato, several 'promoter-bashing' mutant lines have been characterized that exhibit phenotypic variations according to the expression levels of the target gene (Rodríguez-Leal *et al.*, 2017; Wang *et al.*, 2021). However, altered expression levels do not always correlate with phenotypic consequences, possibly because unwanted spatio-temporal expression changes make them unpredictable (Rodríguez-Leal *et al.*, 2017; Wang *et al.*, 2021). Our approach of editing homoeologous genes could exclude unwanted side effects from the ectopic expression of the target gene generated by editing the regulatory regions.

In addition to the knockout mutants, we obtained several weak allele mutants by CRISPR-Cas9 mutagenesis. Since hexaploids have more copies of homologous genes, the chances of finding different mutation types were higher in *S. nigrum* than in *S. americanum* (Figure S15). An interesting example is the *SP3<sup>a1</sup>* allele of *S. nigrum* cv. Boranong (Figure 6). This gene encodes a 1-a.a.-deleted *SP3* protein, which possibly possesses altered protein activity according to the phenotypic effects in the *sp1sp2sp3<sup>a1</sup>* mutant; however, detailed biochemical analyses should be performed. Weak allele mutants were rarely found in diploid *S. americanum*, as phenotypic variations were much less than in the hexaploids. Thus, using polyploid plants, we may have more opportunities to identify additional weak mutations that could be easily dismissed in diploids. This could expand our knowledge of detailed protein functions.

The *s1s2s3* triple mutants of *S. nigrum* and *S. nigrum* cv. Boranong developed flowers and fruits with viable seeds, unlike the *s* mutants in diploids. This finding suggests that more complex compensation mechanisms to restore *S* gene function deficiency might have been adopted in black nightshade species during polyploid evolution and genome triplication. One possible explanation is that the neo-functionalization of other *WOX* family genes by coding sequence mutations might functionally (perhaps partially) complement the *S* gene. This type of genetic change can occur more frequently in increased polyploid genomes to overcome deleterious mutations (Comai, 2005).

Although our attempts were individually restricted to specific genes that are functionally known and related to certain phenotypes, expression analyses and validation results from transcriptome data (Figure 2 and Figure S11) further suggest that our approach could be generalized. Lower accumulation of sequence divergence and highly conserved expression profiles in homoeologous gene sets might give rise to functional redundancy, which generates a genetic dosage affecting QT in the

hexaploid *S. nigrum*. By editing these genes, QT can be precisely controlled to improve the polyploid crops quality.

Furthermore, our strategy can be applied to synthetic polyploid plants. Using techniques such as colchicine treatment, we artificially generated auto- or allo-polyploid crops via self-fertilization or crossing with closely related species, respectively. By combining new polyploid speciation and engineering of duplicated homoeologous genes, we can rapidly improve diploid crops, such as tomato and rice, with optimized quantitative traits.

## Methods

### Permission

No specific permits were required to grow *S. nigrum* plants in the greenhouse at Wonkwang University, Iksan, Republic of Korea. Transgenic and engineered mutants were grown in an LMO growth room (LML16-1201) permitted by the National Research Safety Headquarters in the Republic of Korea. For field experiments, we used the LMO quarantine space in Ochang (LML16-750) permitted by the National Research Safety Headquarters of the Republic of Korea. All the methods complied with relevant institutional, national, and international guidelines and legislation for scientific research.

### Plant materials and growth conditions

*Solanum americanum* (accession SP2273 and SP2275), *S. nigrum* (accession NIBRGR0000189638 and cultivar Boranong) and *S. lycopersicum* (cultivar M82) were used in this study. *S. nigrum* cv. Boranong was generated by crossing a regional *S. nigrum* cultivar Naju and a Chinese cultivar MPR-0156 (registered in the National Institute of Horticultural and Herbal Science) by Korean breeders in the Republic of Korea and was obtained from the Korea Seed and Variety Service (KSVS, registration number 5804). Other plant species are in our stock. Plants were grown in a greenhouse under natural long-day conditions (14–16 h light, 26–28 °C/8–10 h dark, 18–20 °C; 40%–60% relative humidity) supplemented with artificial light from 200 W halogen lamps. Seeds were directly sown in the soil in 50-cell plastic flats and transferred to pots after 4 weeks. All the plants were grown under drip irrigation and standard fertilizer regimes. To test yield performance, plants were grown under natural field conditions during summer in the Republic of Korea. We used two field conditions: one in Iksan and the other in Ochang. Field management, including irrigation, fertilizer application and pest control, followed standard agricultural practices.

### Genome information of *S. nigrum*

The whole process and results of genome sequencing, assembly, functional annotation of hexaploid *S. nigrum* and comparative analyses with other species, including fluorescence *in situ* hybridization (FISH), are described in the Supplementary Information.

### CRISPR-mutagenesis and plant transformation

CRISPR-Cas9 mutagenesis in *S. americanum* and *S. nigrum* was performed as described previously (Park et al., 2020). Briefly, four gRNAs for each gene were designed using the CRISPRdirect web tool (Naito et al., 2015) and cloned into the pAGM4723 binary vector using Golden Gate cloning as previously described (Werner et al., 2012). The CRISPR-Cas9 constructs on pAGM4723 were inserted into *S. americanum* or *S. nigrum* seedlings by *Agrobacterium tumefaciens* (strain EHA105)-mediated transformation, as

previously described (Park et al., 2020). Briefly, excised cotyledon explants from 10 days after germination (DAG) seedlings were co-cultivated with *Agrobacterium* carrying the transgene for 2 days and incubated in shoot initiation media (2 mg/L zeatin, 0.2 mg/L IAA, 100 mg/L kanamycin, and 200 mg/L cefotaxime) for 10 days, followed by incubation in shoot differentiation media (1 mg/L zeatin and 0.1 mg/L IAA) for 2 weeks; then roots were initiated in rooting media with 100 mg/L kanamycin. All gRNA and primer sequences used for cloning are listed in Table S10.

### Plant genotyping

The standard cetyl-trimethyl-ammonium bromide (CTAB) protocol was used to extract high-molecular-weight DNA (Doyle and Doyle, 1987).  $T_0$  transgenic plants were genotyped to validate the CRISPR-generated mutations using targeted deep sequencing as described previously (Oh et al., 2020). Briefly, high-throughput sequencing was performed using the MiniSeq platform (Illumina). Mutations were analysed using the Cas-Analyser web tool (Park et al., 2017) available from CRISPR RGEN Tools (<http://www.rogenome.net>). Indel frequency (%) was estimated by dividing the number of sequencing reads containing mutations by the total number of sequencing reads. For hexaploid *S. nigrum*, homoeologs in the sequencing reads were differentiated using polymorphic markers, and the indel frequency was manually calculated.  $T_1$  transgenic plants were sprayed with 400 mg/L kanamycin to confirm the absence of the transgene as previously reported (Kim et al., 2022). Nontransgenic plants were confirmed and genotyped by PCR and Sanger sequencing. To genotype the F2 segregants, we designed PCR primers for the derived cleaved amplified polymorphic sequence (dCAPS) or CAPS markers (Neff et al., 1998) and performed targeted Sanger sequencing. All primer sequences used for genotyping are listed in Table S10.

### Plant phenotyping and yield trials

To phenotype the *s* mutants, at least 12 plants per genotype were grown in a greenhouse. The branch and flower numbers of five to six inflorescences on each plant were counted. For fruit yield for the *s* mutants, at least seven plants were grown in the greenhouse with pruning for each plant to have only the main stem and the first sympodial stem or grown in the Ochang field without pruning in the summer season in 2021 and 2022. To phenotype the *sp* mutants, at least six plants per genotype were grown in the Wonkwang field without pruning during the summer season in 2021. Average internode lengths were calculated by dividing stem length by inflorescence number, and the lengths of the first internodes of the eighth and ninth axillary shoots from the top and the number of emerging inflorescences in the eighth and ninth axillary shoot branching regions (within 1 cm) from the top were measured. The sympodial index represents the number of leaf nodes per sympodial unit, and the flowering time was calculated as the number of true leaves per primary shoot. For SID plants of Boranong, we defined an axillary shoot with  $\geq 15$  cm in length as a long branch, counted, and then pruned them out. For the fruit yield of the *sp* mutants, at least six plants per genotype were grown in the Ochang field without pruning during the summer season in 2021. To phenotype the *cle9* mutants in the *clv3Q* background, F2 segregants were grown in a greenhouse. Although we obtained only two plants for *clv3Q cle9h2 cle9h3* from the segregants, at least five plants were examined for other genotypes. Sepal and locule numbers of the first floral organ of each plant were counted, and the average fruit weight was

calculated from five inflorescences by dividing the inflorescence weight by the number of fruits.

Fruit yield potential was calculated by multiplying the total inflorescence number by the average inflorescence weight of each plant. Other yield-related traits, such as plant height, plant weight (fresh weight, including fruits), 10-fruit weight, and Brix of ripe fruits, were measured for each genotype. Damaged or diseased plants were excluded from the analysis.

## Acknowledgements

We thank members of the Park's laboratory for valuable comments and discussions. We thank A. Cho, Y.S. La and Y.S. Chae for technical support. We thank Prof. S.K. Oh, Dr. C.G. Kim and staff from the Korean Research Institute of Bioscience and Biotechnology for assistance with plant care. We thank M. E. Bartlett for assistance with the phylogenetic tree. This work was supported by a grant from the New Breeding Technologies Development Program (project no. PJ01653802) to S.J.P., also supported by the National Research Foundation of Korea grant numbers 2020R1A6A1A03044344, 2021R1C1C2006483 and 2020R1A2C1101915 to J.-C.H., K.K. and S.J.P., respectively, and supported by the USDA ARS 80622100 044 000D to K.M.C. and D.W.

## Author contributions

E.S.L., J.H., W.Y.B., K.M.C., D.W., K.K. and S.J.P. conceived the original research plan; E.S.L., K.K. and S.J.P. generated CRISPR mutants and plant phenotyping; J.H., W.Y.B., K.M.C., S.B., J.D.G.J., K.H.S. and D.W. performed genome sequencing and bioinformatic analyses; N.E.W., N.T.H. and H.H.K. performed FisH experiment; E.S.L., J.H., M.J.K., H.K.B., Y.J.K., R.A.P., J.I.J., K.I.C., S.H.S. and S.R. performed yield trials; Y.-K.C., J.H.B., C.M.K., Y.K.L., J.L., J.C.H., K.K. and S.J.P. analysed the data; D.W., K.K. and S.J.P. supervised the work; E.S.L., J.H., K.K. and S.J.P. wrote the manuscript with contributions from all authors.

## Competing interests

The authors declare no competing interests.

## Data availability statement

All the raw sequencing data generated in this study have been deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) with BioProject accession number PRJNA946382. The assembled genomes and annotations are available at Figshare ([https://figshare.com/articles/dataset/Genome\\_assembly\\_and\\_annotation\\_for\\_Solanum\\_amERICANum\\_and\\_Solanum\\_nigrum/22316296](https://figshare.com/articles/dataset/Genome_assembly_and_annotation_for_Solanum_amERICANum_and_Solanum_nigrum/22316296)).

## References

Andrés, F. and Coupland, G. (2012) The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* **13**, 627–639.

Bernard, R.L. (1972) Two genes affecting stem termination in soybeans 1. *Crop. Sci.* **12**, 235–239.

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illia, E., Díez, M.J., Francis, D. et al. (2015) Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics* **16**, 257.

Comai, L. (2005) The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836–846.

Doyle, J.J. and Doyle, J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15.

Edmonds, J.M. and Chwya, J.A. (1997) *Black Nightshades: Solanum Nigrum L. and Related Species*. BPromoting the conservation and use of underutilized and neglected crops 15, pp. 1–113. Rome, Italy: Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute.

Hendelman, A., Zebell, S., Rodriguez-Leal, D., Dukler, N., Robitaille, G., Wu, X., Kostyuk, J. et al. (2021) Conserved pleiotropy of an ancient plant homeobox gene uncovered by cis-regulatory dissection. *Cell* **184**, 1724–1739.e16.

Heo, J., Bang, W.Y., Jeong, J.C., Park, S.C., Lee, J.M., Choi, S. et al. (2022) The comparisons of expression pattern reveal molecular regulation of fruit metabolites in *S. nigrum* and *S. lycopersicum*. *Sci. Rep.* **12**, 5001.

Jabamalairaj, A., Priatama, R.A., Heo, J. and Park, S.J. (2019) Medicinal metabolites with common biosynthetic pathways in *Solanum nigrum*. *Plant Biotechnol. Rep.* **13**, 315–327.

Kim, M.J., Beak, H.K., Choi, J.E., Lee, E.S., Kim, K., Kim, C.M. and Park, S.J. (2022) Simple methods for selection of T-DNA-free segregants from offspring of gene-edited *Solanum nigrum*. *Plant Biotechnol. Rep.* **16**, 257–264.

Krieger, U., Lippman, Z.B. and Zamir, D. (2010) The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat. Genet.* **42**, 459–463.

Kwon, C.T., Tang, L., Wang, X., Gentile, I., Hendelman, A., Robitaille, G. et al. (2022) Dynamic evolution of small signalling peptide compensation in plant stem cell control. *Nat. Plants* **8**, 346–355.

Leitch, A.R. and Leitch, I.J. (2008) Genomic plasticity and the diversity of polyploid plants. *Science* **320**, 481–483.

Lemmon, Z.H., Park, S.J., Jiang, K., Van Eck, J., Schatz, M.C. and Lippman, Z.B. (2016) The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation. *Genome Res.* **26**, 1676–1686.

Lin, X., Jia, Y., Heal, R., Prokchorchik, M. and Sindalovskaya, M. (2022) The *Solanum* pan-genome and effectoromics reveal new resistance genes against potato late blight. *bioRxiv*, 2022.08.11.503608.

Lippman, Z.B., Cohen, O., Alvarez, J.P., Abu-Abied, M., Pekker, I., Paran, I. et al. (2008) The making of a compound inflorescence in tomato and related nightshades. *PLoS Biol.* **6**, 2424–2435.

Morineau, C., Belloc, Y., Tellier, F., Gissot, L., Kelemen, Z., Nogué, F. and Faure, J.D. (2017) Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa*. *Plant Biotechnol. J.* **15**, 729–739.

Naito, Y., Hino, K., Bono, H. and Ui-Tei, K. (2015) CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* **31**, 1120–1123.

Nan, H., Cao, D., Zhang, D., Li, Y., Lu, S., Tang, L., Yuan, X. et al. (2014) GmFT2a and GmFT5a redundantly and differentially regulate flowering through interaction with and upregulation of the bZIP transcription factor GmFDL19 in soybean. *PLoS One* **9**, e97669.

Neff, M.M., Neff, J.D., Chory, J. and Pepper, A.E. (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant J.* **14**, 387–392.

Oh, Y., Lee, B., Kim, H. and Kim, S.G. (2020) A multiplex guide RNA expression system and its efficacy for plant genome engineering. *Plant Methods* **16**, 37.

Park, S.J., Jiang, K., Schatz, M.C. and Lippman, Z.B. (2012) Rate of meristem maturation determines inflorescence architecture in tomato. *Proc. Natl. Acad. Sci. USA* **109**, 639–644.

Park, S.J., Jiang, K., Tal, L., Yichie, Y., Gar, O., Zamir, D. et al. (2014) Optimization of crop productivity in tomato using induced mutations in the florigen pathway. *Nat. Genet.* **46**, 1337–1342.

Park, J., Lim, K., Kim, J.S. and Bae, S. (2017) Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* **33**, 286–288.

Park, S., Lee, E., Heo, J., Kim, D.H., Chun, H.J., Kim, M.C. et al. (2020) Rapid generation of transgenic and gene-edited *Solanum nigrum* plants using *Agrobacterium*-mediated transformation. *Plant Biotechnol. Rep.* **14**, 497–504.

Pnueli, L., Carmel-Goren, L., Hareven, D., Gutfinger, T., Alvarez, J., Ganal, M., Zamir, D. et al. (1998) The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. *Development* **125**, 1979–1989.

Rebocho, A.B., Bliek, M., Kusters, E., Castel, R., Procissi, A., Roobek, I., Souer, E. et al. (2008) Role of EVERGREEN in the development of the cymose petunia inflorescence. *Dev. Cell* **15**, 437–447.

Rodríguez-Leal, D., Lemmon, Z.H., Man, J., Bartlett, M.E. and Lippman, Z.B. (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell* **171**, 470–480.e8.

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S. et al. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641.

Shen, R., Wang, L., Liu, X., Wu, J., Jin, W., Zhao, X. et al. (2017) Genomic structural variation-mediated allelic suppression causes hybrid male sterility in rice. *Nat. Commun.* **8**, 1310.

Somssich, M., Je, B.I., Simon, R. and Jackson, D. (2016) CLAVATA-WUSCHEL signaling in the shoot meristem. *Development* **143**, 3238–3248.

Soyk, S., Lemmon, Z.H., Oved, M., Fisher, J., Liberatore, K.L., Park, S.J., Goren, A. et al. (2017) Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* **169**, 1142–1155.e12.

Soyk, S., Benoit, M. and Lippman, Z.B. (2020) New horizons for dissecting epistasis in crop quantitative trait variation. *Annu. Rev. Genet.* **54**, 287–307.

Thouet, J., Quinet, M., Ormenese, S., Kinet, J.M. and Périlleux, C. (2008) Revisiting the involvement of Self-pruning in the sympodial growth of tomato. *Plant Physiol.* **148**, 61–64.

Train, C.M., Glover, N.M., Gonnet, G.H., Altenhoff, A.M. and Dessimoz, C. (2017) Orthologous Matrix (OMA) algorithm 2.0: more robust to asymmetric evolutionary rates and more scalable hierarchical orthologous group inference. *Bioinformatics* **33**, i75–i82.

Waminal, N.E., Pellerin, R.J., Kim, N.S., Jayakodi, M., Park, J.Y., Yang, T.J. and Kim, H.H. (2018) Rapid and efficient fish using pre-labeled oligomer probes. *Sci. Rep.* **8**, 8224.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. and Qiu, J.L. (2014) Simultaneous editing of three homeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **32**, 947–951.

Wang, X., Aguirre, L., Rodríguez-Leal, D., Hendelman, A., Benoit, M. and Lippman, Z.B. (2021) Dissecting cis-regulatory control of quantitative trait variation in a plant stem cell circuit. *Nat. Plants* **7**, 419–427.

Weeks, D.P. (2017) Gene editing in polyploid crops: wheat, camelina, canola, potato, cotton, peanut, sugar cane, and citrus. *Prog. Mol. Biol. Transl. Sci.* **149**, 65–80.

Wendel, J.F. (2000) Genome evolution in polyploids. *Plant Mol. Biol.* **42**, 225–249.

Werner, S., Engler, C., Weber, E., Gruetzner, R. and Marillonnet, S. (2012) Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. *Bioeng. Bugs* **3**, 38–43.

Xu, C., Liberatore, K.L., Macalister, C.A., Huang, Z., Chu, Y.H., Jiang, K. et al. (2015) A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nat. Genet.* **47**, 784–792.

Zaman, Q.U., Wen, C., Yuqin, S., Mengyu, H., Desheng, M., Jacqueline, B., Baohong, Z. et al. (2021) Characterization of SHATTERPROOF homeologs and CRISPR-Cas9-mediated genome editing enhances pod-shattering resistance in *Brassica napus* L. *CRISPR J.* **4**, 360–370.

Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., Liu, J. et al. (2018) Analysis of the functions of TaGW2 homeologs in wheat grain weight and protein content traits. *Plant J.* **94**, 857–866.

Zsögön, A., Čermák, T., Naves, E.R., Notini, M.M., Edel, K.H., Weinl, S., Freschi, L. et al. (2018) De novo domestication of wild tomato using genome editing. *Nat. Biotechnol.* **36**, 1211–1216.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Table S1** Chloroplast genomes of various species used for the phylogenetic tree.

**Table S2** Chromosome compositions and FISH karyotypes of *Solanum americanum* and *Solanum nigrum*.

**Table S3** Read quality of whole-genome sequencing in *Solanum nigrum*.

**Table S4** Contigs and scaffolds in genome assembly of *Solanum nigrum*.

**Table S5** Comprehensive statistics of *Solanum americanum* and *Solanum nigrum* gene annotations.

**Table S6** Quality control of *Solanum americanum* and *Solanum nigrum* proteins using BUSCO.

**Table S7** Repeat annotation of *Solanum americanum* and *Solanum nigrum* genomes.

**Table S8** Phylogenetic analysis of largest synteny blocks in pepper, tomato, *Solanum americanum* and *Solanum nigrum*.

**Table S9** Read quality of mRNA sequencing in *Solanum americanum* and *Solanum nigrum*.

**Table S10** Primers used in this study.

**Table S11** Probe sequences used for FISH experiments.

**Data S1** Supplementary Dataset.

**Appendix S1** Supplementary Information and Supplementary Figures.