



## Article

# Characterization of Genomic Variation from Lotus (*Nelumbo Adans.*) Mutants with Wide and Narrow Tepals

Fengluan Liu , Mi Qin, Shuo Li, Dasheng Zhang, Qingqing Liu, Mengxiao Yan and Daike Tian \*

Shanghai Key Laboratory of Plant Functional Genomics and Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China; 19liu19@163.com (F.L.); qm9314z@163.com (M.Q.); lishuo456lp@163.com (S.L.); dszhang@cemps.ac.cn (D.Z.); nicecocoliu@sina.com (Q.L.); mxyan@cemps.ac.cn (M.Y.)

\* Correspondence: dktian@cemps.ac.cn

**Abstract:** Compared with rose, chrysanthemum, and water lily, the absence of short-wide and long-narrow tepals of ornamental lotus (*Nelumbo Adans.*) limits the commercial value of flowers. In this study, the genomes of two groups of lotus mutants with wide-short and narrow-long tepals were resequenced to uncover the genomic variation and candidate genes associated with tepal shape. In group NL (short for *N. lutea*, containing two mutants and one control of *N. lutea*), 716,656 single nucleotide polymorphisms (SNPs) and 221,688 insertion-deletion mutations (Indels) were obtained, while 639,953 SNPs and 134,6118 Indels were obtained in group WSH (short for 'Weishan Hong', containing one mutant and two controls of *N. nucifera* 'Weishan Hong'). Only a small proportion of these SNPs and Indels was mapped to exonic regions of genome: 1.92% and 0.47%, respectively, in the NL group, and 1.66% and 0.48%, respectively, in the WSH group. Gene Ontology (GO) analysis showed that out of 4890 (NL group) and 1272 (WSH group) annotated variant genes, 125 and 62 genes were enriched ( $Q < 0.05$ ), respectively. Additionally, in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, 104 genes (NL group) and 35 genes (WSH group) were selected ( $p < 0.05$ ). Finally, there were 306 candidate genes that were sieved to determine the development of tepal shape in lotus plants. It will be an essential reference for future identification of tepal-shaped control genes in lotus plants. This is the first comprehensive report of genomic variation controlling tepal shape in lotus, and the mutants in this study are promising materials for breeding novel lotus cultivars with special tepals.

**Keywords:** Asian lotus; genomic variation; *Nelumbo*; SNP; petal shape; whole-genome resequencing



**Citation:** Liu, F.; Qin, M.; Li, S.; Zhang, D.; Liu, Q.; Yan, M.; Tian, D. Characterization of Genomic Variation from Lotus (*Nelumbo Adans.*) Mutants with Wide and Narrow Tepals. *Horticulturae* **2021**, *7*, 593. <https://doi.org/10.3390/horticulturae7120593>

Academic Editors: Hamidou F. Sakhanokho, Nurul Islam-Faridi and Jiafu Jiang

Received: 10 November 2021

Accepted: 10 December 2021

Published: 20 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Lotus (*Nelumbo Adans.*) is one of the most economically aquatic plants in Southeast Asia, especially in China [1]. The genus *Nelumbo* has two species: *N. nucifera* Gaertn., commonly known as Asian lotus, and *N. lutea* Willd., commonly known as American lotus [2,3]. Asian lotus is one of the top 10 traditional flowers in China and the national flower of India [4]. Lotus, particularly Asian lotus, is usually divided into three types, depending on its agricultural application: ornamental lotus, rhizome lotus, and seed lotus [1,5]. Approximately 2000 commercial cultivars of ornamental lotus have been developed to date, accounting for 96% of the total number of the three types [6]. However, as an ornamental character of importance [7], the shape of petals (referred to as tepals in the lotus) exhibits little phenotypic variation in these cultivars. Few lotus germplasms exhibit short-wide tepals (similar to rose) and long-narrow tepals (as in chrysanthemum), which limits the ornamental value and market requirement of lotus [8]. Therefore, to develop lotus cultivars with unique tepal shapes, understanding the genetic basis of its determination is critical.

Genetic mapping of quantitative trait loci (QTL), genome-wide association studies (GWAS), and transcriptome analysis using modern high-throughput sequencing methods are powerful approaches that facilitate the identification of novel genes and gene regulatory

networks [9,10]. The completion of *de novo* sequencing of the lotus genome [11,12] and the rapid development of high-throughput sequencing technology have made it possible to analyze nucleotide sequence variation among various lotus cultivars. Whole-genome resequencing is used to detect single nucleotide polymorphisms (SNPs) and insertion-deletion mutations (Indels) in the whole genome [13,14] and to mine structural variation and copy number variation by high-depth sequencing [15,16]. Using this technique in lotus, Hu et al. [17] identified a large number of nucleotide sequence variations and 103,656 simple sequence repeats (SSRs) between the genomes of ‘Chiang Mai Wild’ and ‘Middle Lake Wild’, which could be used for QTL mapping and molecular-assisted breeding. In another study, whole-genome resequencing of 19 accessions belonging to three subgroups of cultivated temperate lotus revealed that rhizome lotus showed the lowest genomic diversity, and the genes that found in the genomic region and related to temperate lotus and tropical lotus always exhibited divergent expression patterns [18].

Petal growth depends on the precise orchestration of the frequency of cell division, the orientation of cell division planes, and the extent of cell elongation [19–21]. For example, non-uniform cell expansion within a petal resulted in distorted petals in *Eustoma grandiflorum* [22]. Although the molecular basis of petal shape is largely unclear, it has been shown that several genes, such as *LEUNIG* [23,24], *SEUSS* and *APETALA2* [25,26], *JAGGED* [27–29], *CYCLOIDEA* [30], *INDOLE-3-BUTYRIC ACID-RESPONSE5* [31], and *SPIKE1* [32], regulate petal shape by affecting cell proliferation and or cell expansion. However, the roles of these genes in petal shape regulation were mainly verified in model plants including *Arabidopsis* and *Antirrhinum*.

In this study, three lotus mutants with wide-short and narrow-long tepals were divided into two groups and analyzed by whole-genome resequencing. We explored the differences in the number and distribution of SNPs and Indels in each group and compared them between the two lotus groups. In the meantime, functional genes and metabolic pathways related to the development of broad and narrow tepals were screened. The results of this study provide important clues for the exploration of genes affecting tepal shape in lotus plants.

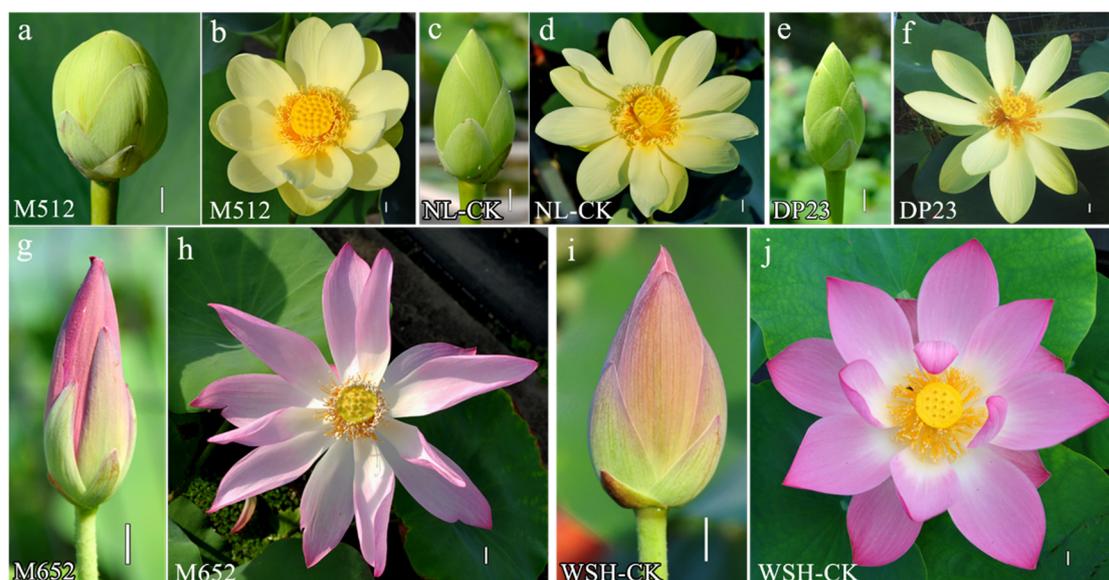
## 2. Materials and Methods

### 2.1. Plant Material

In 2015, two accessions of American lotus (*N. lutea*) with significant differences in tepal shape were obtained from the ‘International Nelumbo Collection’ (certified by International Waterlily and Water Gardening Society in 2016) of Shanghai Chenshan Botanical Garden, Shanghai, China. Both these accessions were germinated from wild lotus seeds collected from Florida, USA. Compared with the wild American lotus, one genotype (M512) showed wide and short tepals, while the other genotype (DP23) exhibited narrow and long tepals (Figure 1), and the tepal shapes of both accessions were stable after years of vegetative propagation via rhizomes. In 2016, 1024 seeds of ‘Weishan Hong’, a wild Asian lotus (*N. nucifera*), were irradiated with cobalt-60, which emits gamma rays, and a mutant (M652) with narrow and long tepals was obtained (Figure 1), and the tepal shape was stable after three years of clonal propagation.

### 2.2. Grouping of Lotus Accessions

The lotus genotypes investigated in this study were divided into two groups: NL (short for *N. lutea*), which comprised American lotus genotypes, M512, DP23, and their wild-type control (NL-CK), and WSH (short for ‘Weishan Hong’), which comprised M652 and its two wild-type controls, WSH-CK1 and WSH-CK2. Each control contained three wild seedlings.



**Figure 1.** Buds and second-day flowers of NL group (a–f) and WSH group (g–j). The NL group accessions included M512 ((a,b), wide-short tepals), DP23 ((e,f), narrow-long tepals), and NL-CK ((c,d), wild American lotus; the control). The WSH group accessions included M652 ((g,h), derived from ‘Weishan Hong’ by gamma irradiation; narrow-long tepals) and WSH-CK ((i,j), ‘Weishan Hong’ belonging to wild Asian lotus; the control). Scale bars = 1 cm.

### 2.3. DNA Extraction

Three young leaves of each mutant (M512, DP23, and M652) were harvested separately from three pots of their clones, and one young leaf of three seedlings of each control (NL-CK, WSH-CK1, and WSH-CK2) were collected. DNA was extracted separately from the three leaves of the same genotype using the modified cetyl triethylammonium bromide (CTAB) method [33], in which the sugar removal was performed three times, then they were pooled together at equimolar concentrations. The integrity of the pooled DNA samples was detected by electrophoresis on 1% agarose gels, and the DNA quality and concentration were measured by NanoDrop 2000c and Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). DNA samples with OD260/OD280 (optical density, OD) values of 1.8–2.0 and with contents greater than 1.5  $\mu\text{g}$  were used for library construction.

### 2.4. Library Construction and Whole Genome Resequencing

The purified genomic DNA was randomly sheared into 350 bp fragments using a COVARIS M220 Focused-ultrasonicator<sup>TM</sup> (Covaris, Woburn, MA, USA). The genomic DNA library was constructed using the TruSeq Nano DNA HT Sample Preparation Kit (Illumina Inc., San Diego, CA, USA), and its quality was assessed on the Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY, USA). The size of inserts was determined using Agilent Technologies 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Finally, whole-genome resequencing was conducted by Beijing Novogene Bioinformatics Technology Co. Ltd. (Beijing, China) at a depth of 10 $\times$  using the Illumina HiSeq 2500 platform.

### 2.5. Data Filtering

The original image data generated by the sequencing machine were converted into sequence data via base calling (Illumina pipeline CASAVA v1.8.2). The sequence data were then subjected to the quality control (QC) procedure to remove unusable reads: (1) the reads with a Phred quality (Q) score < 20; (2) the reads contain the Illumina library construction adapters; (3) the reads contain more than 10% unknown bases (N bases); (4) one end of the read contains more than 50% of low-quality bases (sequencing quality value  $\leq$  5).

## 2.6. Read Mapping, Variant Detection, and Annotation

The clean reads were mapped to the reference genomes of ‘China Antique’ (accession number: GCA\_003033685.1) using the Burrows-Wheeler Aligner (BWA) program [34] using the following parameters: ‘mem-t4-k32-M’. Subsequent processing, including duplicate removal, was performed using SAM tools [35] with the parameter ‘rmdup’ and PICARD (<http://broadinstitute.github.io/picard/>, 11 December 2021). Raw SNP and Indel sets were called with SAM tools using the following parameters: ‘mpileup-m2-F0.002-d1000’. Then, these sets were filtered using the following criteria: mapping quality >20; depth of the variate position >4. Additionally, the missing data of SNPs and Indels of six samples from two groups were filtered. ANNOVAR [36] was used for the functional annotation of variants.

## 2.7. Screening and Enrichment Analysis of Variant Genes

Genes showing SNPs and Indels between any two samples (except between WSH-CK1 and WSH-CK2) within each group were selected, that is, M512 vs. NL-CK, DP23 vs. NL-CK, M512 vs. DP23, M652 vs. WSH-CK1, and M652 vs. WSH-CK2. Then, these genes from pairwise comparisons were combined and removed duplicates in each group. Among these genes, those showing SNPs and Indels in the exonic regions were manually extracted and searched in the NCBI non-redundant protein (Nr) database for annotation and in the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases for enrichment analyses.

## 2.8. The Assessment of the Reliability of Candidate Genes

Based on the blasting in the GO and KEGG databases, genes from NL and WSH groups showing significant enrichment ( $Q < 0.05$ ) were identified and considered as the candidate genes that may relate to the development of tepal shape in lotus. Among those genes, the genes common to both groups and the top 10% genes with the highest mutation rate (SNP and Indel density per kb in the coding region) were extracted and searched in NCBI and UniProtKB databases to clarify their biological information. Both types of genes were served as a rough ruler to measure the reliability of the candidate genes.

# 3. Results

## 3.1. Whole-Genome Resequencing Data of the Six Genotypes

A total of six samples belonging to groups NL and WSH were subjected to the whole-genome resequencing, and 9.23–12.85 Gb clean data, with ‘Q20 rate’ > 97% and ‘Q30 rate’ > 92%, were generated (Table S1). Among the six samples, M652 produced the lowest amount of clean data (9.23 Gb), whereas NL-CK produced the highest amount (12.85 Gb). The GC content of all samples varied from 39.28% to 40.14%, and no significant difference was detected in GC content between and within the two groups (Table S1). The mapping rate of NL group genotypes (93.63–94.43%) was lower than that of WSH group accessions (98.22–98.43%). Additionally, the 1× and 4× coverage of NL group samples was also lower than those of WSH group accessions; the 4× coverage of the NL group ranged from 80.41% to 83.54%, whereas that of the WSH group varied from 94.65% to 98.67% (Table S1).

## 3.2. SNPs and Indels in the Six Genotypes

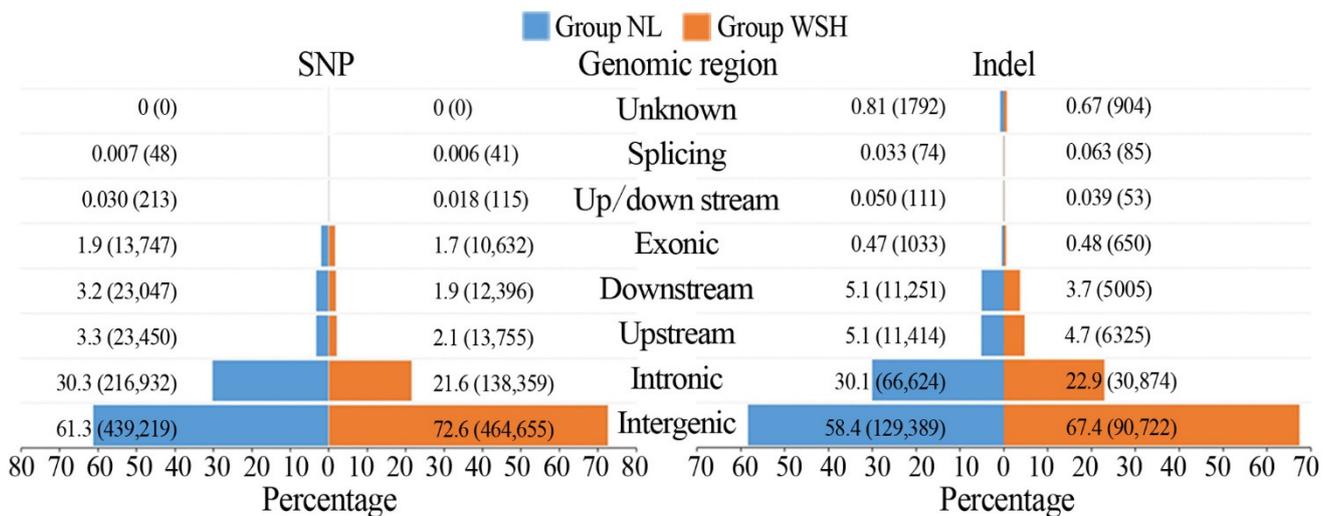
The sequence reads of all six accessions were mapped onto the reference genome sequence of lotus, and SNPs and Indels were called for each accession. The total number of SNPs identified in each sample of NL group (3,928,934; 4,197,575; 4,383,471) was about 7–10-fold higher than that of WSH group (444,148; 542,312; 569,431), and it was about 8–10-fold higher in Indel, with the number of 399,031, 430,348, and 450,433 in NL group and 43,181, 51,780, and 50,568 in the WSH group (Figure S1). Of all the SNPs identified in the NL group, 6.15% (M512), 5.90% (DP23), and 5.73% (NL-CK) were mapped to exonic regions, which were approximately 3.5 times greater than the number of exonic SNPs in the WSH group. Of all the Indels identified in the NL group, 0.84% (M512), 0.81% (DP23),

and 0.79% (NL-CK) were mapped to exonic regions, which was approximately 1.5 times those of the WSH group (Figure S1).

Within each group, the controls contained more SNPs and Indels than their mutants (Figure S1), which was probably associated with the increase in variant sites caused by the pooling of DNA extracted from three seedlings per control. Additionally, the ratio of nonsynonymous to synonymous SNPs (Nonsyn/Syn), transitions to transversions (Ts/Tv), and exonic sequence variation to total variation (exonic/total) varied slightly among the three accessions within each group (Figure S1), to some extent, suggesting that the three genotypes within each group exhibit very low genetic diversity.

### 3.3. Analysis of SNPs and Indels between Mutant Accessions and Their Controls

Within each group, the SNPs and Indels were filtered between any two samples (except between WSH-CK1 and WSH-CK2). A total of 716,656 SNPs and 221,688 Indels were obtained from the NL group, while 639,953 SNPs and 134,6118 Indels were obtained from the WSH group (Figure 2). Compared with the controls, nucleotide polymorphisms in the mutant samples of the two groups occurred mainly in the intergenic regions: intergenic SNPs accounted for 61.3% and 72.6% of all SNPs identified in NL and WSH groups, respectively, and intergenic Indels accounted for 58.4% and 67.4% of all Indels identified in NL and WSH groups, respectively (Figure 2). By contrast, the proportion of SNPs and Indels that mapped to exonic regions was very small in both groups (1.9% SNPs and 0.47% Indels in the NL group; 1.7% SNPs and 0.48% Indels in the WSH group) (Figure 2). This showed that the ratio of exonic SNPs to total SNPs was slightly higher in the NL group than in the WSH group.



**Figure 2.** Distribution, proportion, and number of SNPs and Indels identified within each group. Each pair of data represented the percentage (numbers out the brackets) and the amount (numbers in the brackets), respectively.

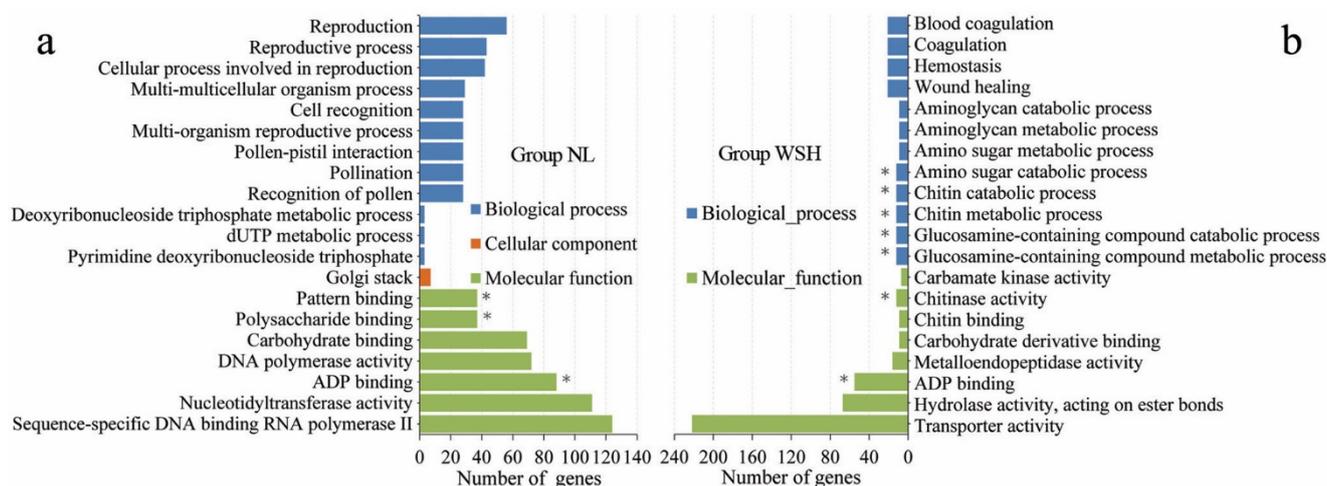
### 3.4. Annotation and Analysis of Genes Harboring SNPs and Indels in the Exonic Regions

The predicted amino acid sequences of the variant genes with exonic SNPs and Indels were blasted in the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases to annotate their functions. In the GO database, 4890 (72.8%) of the 6715 variant genes in the NL group were annotated and assigned to three functional categories for a total of 19,785 times (Table 1). The most distributed times of these variant genes were assigned to the biological process category (99,311 times (50.2%)), followed by the cellular component and molecular function categories (Table 1). In the entire genetic background and variant genes, GO terms including ‘ADP binding’, ‘pattern binding’, and ‘polysaccharide binding’ were significantly ( $Q < 0.05$ ) enriched in the molecular function category and contained 88, 37, and 37 candidate genes, respectively (Figure 3a). After

excluding duplicates, a total of 125 candidate genes were selected from the above three GO items (Table S2).

**Table 1.** Annotation and distribution of variant genes in GO and KEGG databases.

Database	Sample Group	No. of Variant Genes	No. of Annotated Genes	Percentage of Annotated Genes	Distributed Times of Annotated Genes			
					Total Times	Biological Process	Cellular Component	Molecular Function
GO	NL	6715	4890	72.8	197,885	99,311	72,072	26,502
	WSH	5353	1272	23.8	52,425	25,824	20,751	5850
Database	Sample group	No. of variant genes	No. of genes annotated	Percentage of annotated genes	No. of pathways	No. of annotated genes in the top three pathways		
						Metabolic pathways	Biosynthesis of secondary metabolites	Starch and sucrose metabolism
KEGG	NL	6715	2286	34.0	122	430	256	67
	WSH	5353	645	12.0	68	128	84	39



**Figure 3.** Top 20 of most enriched Gene Ontology (GO) terms among variant genes in NL group (a) and WSH group (b). GO terms showing significant differences ( $Q < 0.05$ ) within each group were indicated with an asterisk (\*).

In the WSH group, 1272 genes were annotated in the GO database, accounting for 23.8% of the 5353 variant genes, which was significantly lower than the proportion of annotated genes in the NL group (Table 1). These annotated genes of the WSH group received a total of 52,425 assignments in the three functional categories (in the same order as that observed in the NL group): 25,824 times (49.3%) in the biological process category, followed by 20,751 times in the cellular component category, and 5850 times in the molecular function category (Table 1). Seven GO terms, such as ‘ADP binding’, ‘chitin metabolic process’, ‘chitin catabolic process’, showed significant differences ( $Q < 0.05$ ). Among them, ‘ADP binding’ contained 55 genes, while each of the remaining six GO terms contained 12 identical genes (Figure 3b). Finally, A total of 62 candidate genes were identified in the 7 GO terms, after excluding gene duplicates (Table S2).

In the KEGG database, 2286 variant genes of the NL group were assigned into 122 pathways, of which the top three pathways were ‘metabolic pathways’ (430 genes), ‘biosynthesis of secondary metabolites’ (256 genes), and ‘starch and sucrose metabolism’ (67 genes) (Table 1). By contrast, only 645 genes of the WSH group were annotated in 68 pathways, and the top three pathways with the most annotated genes and their order were the same as those in the NL group, each with 128, 84, and 39 genes, respectively (Table 1). However, no pathway that showed a significant overrepresentation, based on the Q-value ( $Q < 0.05$ ), was found in the two groups. While four pathways (comprising 104 genes, Table S2) with

significant  $p$ -value ( $p < 0.05$ ) were detected in the NL group, including ‘ribosome biogenesis’, ‘ABC transporters’, ‘propanoate metabolism’, and ‘phenylpropanoid biosynthesis’. Additionally, 5 pathways (comprising 35 genes, Table S2) were detected in the WSH group ( $p < 0.05$ ), including ‘ABC transporters’, ‘starch and sucrose metabolism’, ‘plant–pathogen interaction’, ‘ether lipid metabolism’, and ‘amino sugar and nucleotide sugar metabolism’.

After excluding gene duplicates, a total of 306 enriched genes were collected from the NL (125 GO + 104 KEGG) and WSH (62 GO + 35 KEGG) groups, and which comparison group each candidate gene came from was also displayed (Table S2). These are essential references for locating and screening tepal shape-regulating genes of lotus in the future.

### 3.5. Sketchy Sampling of Biological Information of Candidate Genes

According to the 306 genes enriched in the GO and KEGG databases, those shared by both groups and with the highest mutation rate (top 10%) in the exonic regions were selected to measure the 306 candidate genes. Thus, a total of 41 genes were sampled, in which 8 genes were shared by the 2 groups, and 24 and 9 genes were exclusive to the NL and WSH groups, respectively. Among the eight shared genes, three (LOC104590721, LOC104605275, and LOC104607443) were enriched in the GO but not the KEGG database, and the remaining five were enriched in both GO and KEGG databases (Table S3).

Of the 41 selected genes, 35 were annotated with a clear function, including 17 disease resistance-related genes, 6 receptor (like-) protein kinases, and 12 genes related to chitinase, transferase, synthase, and other functions. The lengths of the 41 genes varied from 891 to 17,163 bp, in which 2 candidate genes (LOC104607832 and LOC104609114) produced 4 transcript variants, and 10 genes produced 2–3 transcript variants (Table S3).

In terms of ‘family and domain’, the genes harboring nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4 (NB-ARC) and the genes containing leucine-rich repeat (LRR) domains were found the most (18 genes, 44%) and mainly related to the function of ‘ADP binding’ (Table S3). While the domains of the six receptor (like-) protein kinases were epidermal growth factor (EGF)-like, Pkinase, and wall-associated receptor-kinase galacturonan-binding (GUB\_WAK), which are associated with the morphological development of plants due to their functions such as ‘ATP binding’, ‘calcium ion binding’, ‘polysaccharide binding’, and ‘protein serine/threonine kinase activity’. In the KEGG database, only 18 of the 41 candidate genes were assigned to specific pathways, while the remaining 23 genes were unknown (Table S3).

## 4. Discussion

### 4.1. Genetic Relationship of the Individuals in Each Group

The mutants in both groups germinated from the seeds of wild populations. American lotus (*N. lutea*) grows in a wild state in North America, their populations showed low genetic diversity but a high level of differentiation with very low genetic flow [37]. The three accessions of the NL group were the seedlings from the Okeechobee Lake population of Florida, USA, which largely differentiated and separated from the population outside Florida according to our previous work [38]. For the accessions of the WSH group, they were obtained from the seeds of the Weishan Lake population, which has always been regarded as one of the lotus populations native to China [39,40]. Additionally, high uniformity and stability were observed in the phenotypic (especially floral) characteristics of the population and the 1024 seedlings, which were irradiated with the gamma rays of cobalt-60. Thus, the origins of the mutants and controls in each group implied a close genetic relationship among themselves, which was beneficial to rule out the noise, to some extent, in the genetic background that was irrelevant to the tepal-shape variation. In other words, the three mutants in this study are good materials for producing novel lotus cultivars with extreme-shaped tepals and for exploring the genetic regulation of tepal shape in plants.

#### 4.2. Differences in Genome Mapping between the NL and WSH Groups

The two groups of samples investigated in this study were representative wild-type germplasms of the two species of *Nelumbo*. Our results by whole-genome resequencing revealed many differences at the DNA level. For example, the mapping rate, 1× coverage, and 4× coverage of the three lotus accessions in the NL group were lower than those in the WSH group (Table S1), more SNPs and Indels were detected in the NL group than in the WSH group (Table 1). It was probably due to ‘China Antique’ (source of the reference genome) exhibiting a closer genetic relationship with each accession of the WSH group (native to China) than with the geographically isolated American lotus accessions in the NL group. This fact was supported by the fact that ‘China Antique’ and the WSH accessions are the wild lotus genotypes belonging to *N. nucifera*, while the NL accessions are the wild lotus belonging to *N. lutea*, which are two allopatric species of *Nelumbo* that diverged for approximate 12 million years [41]. Therefore, there could be a substantial number of sequences and genes showing presence and absence between *N. nucifera* and *N. lutea*, for instance, between ‘China Antique’ and the accessions of the NL group in this study. In this sense, the *N. nucifera* reference genome (‘China Antique’) has its limitation to studying the *N. lutea* genetic variations.

Previously, we showed that many SNPs and Indels among M512, DP23, and ‘China Antique’ could be detected using expressed sequence tag-derived simple sequence repeat (EST-SSR) makers, and polymorphic DNA bands were produced by 26/217 EST-SSR makers between M512 and DP23, from which the nucleotide sequences or genes obtained were associated with organ development, hormone metabolism, signal transduction, etc. [42]. This indirectly supported the above inference that the reference genome exhibited lower sequence similarity with the American lotus genome than with the ‘Weishan Hong’ genome. In addition, the three accessions in the WSH group exhibited higher Nonsyn/Syn ratio and Ts/Tv ratio than those in the NL group (Figure S1), which was consistent with the results of our previous study [42]. In other words, fewer non-synonymous mutations and transitions occurred in the genomes of NL group accessions, probably because of the lack of gene transfer between Asian lotus and American lotus during the long period of geographical isolation [18,43]. Additionally, genes with non-synonymous mutant SNPs are the focus of future research, and they will be screened and annotated in combination with a comparative analysis of transcriptomes.

#### 4.3. The Reliability of the Candidate Genes That May Associate with Tepal Shape in Lotus

We picked 41 genes to roughly estimate the 306 candidate genes (NL 277 + WSH 89), which were assumed to associate with the development of tepal shape in lotus. Tepal shape and size are remarkably consistent within a given ecotype [44,45] but are influenced by the environment [46,47]. Among the 41 genes in this study, 17 genes encoded disease resistance proteins belonging to the NB-LRR family, of which 12 were exclusive to the NL group, 4 were common to both groups, and only 1 gene was unique to the WSH group (Table S3). The contrasting number of disease resistance genes between the two groups was consistent with, at least to some degree, the fact that M512 and DP23 were derived from natural mutation or selection of lotus seeds, whereas M652 was induced by artificial radiation.

The development of tepal shape is mainly regulated by cellular proliferation, expansion, and the direction of cell division [20,21]. In this study, 6 genes from the 41 sampled genes predicted to encode EGF-like or WAK domain-containing proteins (Table S3), required for cell expansion [48–50], may play essential roles in the morphogenesis of lotus tepals. *ERECTA*, which encodes an LRR receptor-like serine-threonine kinase (STK), was verified to have a major effect locus for shaping petals in *Arabidopsis thaliana* [51]. A BLAST search of this *Arabidopsis* gene sequence in the lotus genome identified its homolog, LOC104600563, which was also annotated as *ERECTA*. However, this gene was not one of the 306 genes screened in this study. Nonetheless, the LOC104604923 gene identified in the WSH group, as well as the previously reported WAK domain-containing genes

LOC104590721, LOC104596880, LOC104608917, and LOC109115528, were all STK-type genes [48], and similar to *ERECTA*, they will be the focus of subsequent research.

To summarize this study, we believe that the genomic data and the 306 candidate genes in this study are valuable for future studies on the tepal shape of lotus plants, combined with other analyses such as transcriptomic data or epigenomic study.

## 5. Conclusions

The shape and size of petals are essential for the survival and reproduction of plants. Both these morphological features are important targets in ornamental crop breeding. Lotus germplasms with wide and narrow tepals generated by natural mutation and artificial irradiation, respectively, represent ideal materials for exploring the mechanism of tepal shape development. Compared with the mutant samples to their respective controls within each group, a large number of SNPs, Indels, and candidate genes can be found, which also exhibited differences between group NL and WSH. The 306 candidate genes enriched from the GO and KEGG databases, especially the EGF and WAK domain-containing genes related to cell expansion and division, must be the research focus in future studies for understanding tepal development in lotus plants. Overall, the results of this study provide a fruitful resource for future identification of petal-shaped control genes study and a valuable database that can be used combined with transcriptome technology and QTL mapping for understanding the mechanism of flower morphogenesis in *Nelumbo*.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/horticulturae7120593/s1>, Figure S1: Distributions of SNPs and Indels in the genomes of six accessions, Table S1: The statistics of the data generated by whole-genome resequencing and genome mapping in the six samples, Table S2: The genes enriched in the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database, Table S3: The statistics of candidate genes that may associate with tepal shape in lotus.

**Author Contributions:** Conceptualization, F.L. and D.T.; methodology, F.L.; validation, M.Y.; formal analysis, M.Q.; investigation, F.L. and Q.L.; resources, S.L.; data curation, M.Q. and D.Z.; writing—original draft preparation, F.L. and M.Q.; writing—review and editing, D.T. and D.Z.; visualization, S.L. and Q.L.; supervision, M.Y.; project administration, D.T.; funding acquisition, F.L. and D.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant Number 32071821) and by the Shanghai Landscaping Administration Bureau (Grant Number G162402).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** We thank Ken Tilt (Auburn University, Alabama, USA) and Lyn Gettys (University of Florida, Florida, USA) for facilitating the collection of American lotus seeds. We also thank Xiao-Li Chen (Beijing Genomics Institution) for providing suggestions on data processing.

**Conflicts of Interest:** The authors declare no conflict of interest. The authors declare that they have no known competing financial interests.

## References

1. Guo, H.B. Cultivation of lotus (*Nelumbo nucifera* Gaertn. ssp. *nucifera*) and its utilization in China. *Genet. Resour. Crop Evol.* **2009**, *56*, 323–330. [[CrossRef](#)]
2. Li, H.L. Classification and phylogeny of Nymphaeaceae and allied families. *Am. Midl. Nat.* **1955**, *54*, 33–41. [[CrossRef](#)]
3. Li, Y.; Svetlana, P.; Yao, J.X.; Li, C.S. A review on the taxonomic, evolutionary and phytogeographic studies of the lotus plant (*Nelumbonaceae: Nelumbo*). *Acta Geol. Sin.* **2014**, *88*, 1252–1261. [[CrossRef](#)]
4. Gowthami, R.; Sharma, N.; Pandey, R.; Agrawal, A. A model for integrated approach to germplasm conservation of Asian lotus (*Nelumbo nucifera* Gaertn.). *Genet. Resour. Crop Evol.* **2021**, *68*, 1269–1282. [[CrossRef](#)]

5. Wang, Q.C.; Zhang, X.Y. *Lotus Flower Cultivars in China*; Bao, M.Z., Translator; Forestry Publishing House Press: Beijing, China, 2004; p. 59.
6. Liu, L.; Li, Y.C.; Min, J.; Tian, D.K. Analysis of the cultivar names and characteristics of global lotus (*Nelumbo*). *Agric. Sci.* **2019**, *9*, 163–181. (In Chinese with English Abstract) [[CrossRef](#)]
7. Lin, Z.Y.; Zhang, C.H.; Cao, D.D.; Damaris, R.N.; Yang, P.F. The latest studies on lotus (*Nelumbo nucifera*)-an emerging horticultural model plant. *Int. J. Mol. Sci.* **2019**, *20*, 3680. [[CrossRef](#)]
8. Min, J.; Xiang, Y.C.; Tian, D.K. Consumption demand and market potential of cut lotus flower in China. *J. Anhui Agric. Sci.* **2018**, *46*, 212–217. (In Chinese with English Abstract)
9. Huang, L.Y.; Li, M.; Cao, D.D.; Yang, P.F. Genetic dissection of rhizome yield-related traits in *Nelumbo nucifera* through genetic linkage map construction and QTL mapping. *Plant Physiol. Bioch.* **2021**, *160*, 155–165. [[CrossRef](#)] [[PubMed](#)]
10. Zou, Y.; Chen, G.L.; Jin, J.; Wang, Y.; Xu, M.L.; Peng, J.; Ding, Y. Small RNA and transcriptome sequencing reveals miRNA regulation of floral thermogenesis in *Nelumbo nucifera*. *Int. J. Mol. Sci.* **2020**, *21*, 3324. [[CrossRef](#)]
11. Gui, S.T.; Peng, J.; Wang, X.L.; Wu, Z.; Cao, R.; Salse, J.; Zhang, H.Y.; Zhu, Z.X.; Xia, Q.J.; Quan, Z.W.; et al. Improving *Nelumbo nucifera* genome assemblies using high-resolution genetic maps and BioNano genome mapping reveals ancient chromosome rearrangements. *Plant J.* **2018**, *94*, 721–734. [[CrossRef](#)] [[PubMed](#)]
12. Ming, R.; VanBuren, R.; Liu, Y.; Yang, M.; Han, Y.; Li, L.T.; Zhang, Q.; Kim, M.J.; Schatz, M.C.; Campbell, M.; et al. Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol.* **2013**, *14*, R41. [[CrossRef](#)]
13. Nielsen, R.; Paul, J.S.; Albrechtsen, A.; Song, Y.S. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* **2011**, *12*, 443–451. [[CrossRef](#)]
14. Say, Y.H. The association of insertions/deletions (INDELs) and variable number tandem repeats (VNTRs) with obesity and its related traits and complications. *J. Physiol. Anthropol.* **2017**, *36*, 25. [[CrossRef](#)] [[PubMed](#)]
15. Magi, A.; Pippucci, T.; Sidore, C. XCAVATOR: Accurate detection and genotyping of copy number variants from second and third generation whole-genome sequencing experiments. *BMC Genom.* **2017**, *18*, 747. [[CrossRef](#)]
16. Neerman, N.; Faust, G.; Meeks, N.; Modai, S.; Kalfon, L.; Falik-Zaccai, T.; Kaplun, A. A clinically validated whole genome pipeline for structural variant detection and analysis. *BMC Genom.* **2019**, *20*, 545. [[CrossRef](#)] [[PubMed](#)]
17. Hu, J.H.; Gui, S.T.; Zhu, Z.X.; Wang, X.L.; Ke, W.D.; Ding, Y. Genome-wide identification of SSR and SNP markers based on whole-genome re-sequencing of a Thailand wild sacred lotus (*Nelumbo nucifera*). *PLoS ONE* **2015**, *11*, e0143765. [[CrossRef](#)]
18. Huang, L.Y.; Yang, M.; Li, L.; Li, H.; Yang, D.; Shi, T.; Yang, P.F. Whole genome re-sequencing reveals evolutionary patterns of sacred lotus (*Nelumbo nucifera*). *J. Integr. Plant Biol.* **2018**, *60*, 2–15. [[CrossRef](#)]
19. Hase, Y.; Fujioka, S.; Yoshida, S.; Sun, G.Q.; Umeda, M.; Tanaka, A. Ectopic endoreduplication caused by sterol alteration results in serrated petals in *Arabidopsis*. *J. Exp. Bot.* **2005**, *56*, 1263–1268. [[CrossRef](#)]
20. Hill, J.P.; Lord, E.M. Floral development in *Arabidopsis thaliana*: A comparison of the wild type and the homeotic pistillata mutant. *Can. J. Bot.* **1989**, *67*, 2922–2936. [[CrossRef](#)]
21. Peaucelle, A.; Wightman, R.; Höfte, H. The control of growth symmetry breaking in the *Arabidopsis* hypocotyl. *Curr. Biol.* **2015**, *25*, 1746–1752. [[CrossRef](#)]
22. Kawabata, S.; Nii, K.; Yokoo, M. Three-dimensional formation of corolla shapes in relation to the developmental distortion of petals in *Eustoma grandiflorum*. *Sci. Hort.* **2011**, *132*, 66–70. [[CrossRef](#)]
23. Conner, J.; Liu, Z.C. *LEUNIG*, a putative transcriptional corepressor that regulates *AGAMOUS* expression during flower development. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12902–12907. [[CrossRef](#)] [[PubMed](#)]
24. Liu, Z.C.; Meyerowitz, E.M. *LEUNIG* regulates *AGAMOUS* expression in *Arabidopsis* flowers. *Development* **1995**, *121*, 975–991. [[CrossRef](#)] [[PubMed](#)]
25. Franks, R.G.; Wang, C.X.; Levin, J.Z.; Liu, Z.C. *SEUSS*, a member of a novel family of plant regulatory proteins, represses floral homeotic gene expression with *LEUNIG*. *Development* **2002**, *129*, 253–263. [[CrossRef](#)] [[PubMed](#)]
26. Franks, R.G.; Liu, Z.C.; Fischer, R.L. *SEUSS* and *LEUNIG* regulate cell proliferation, vascular development and organ polarity in *Arabidopsis* petals. *Planta* **2006**, *224*, 801–811. [[CrossRef](#)]
27. Dinneny, J.R.; Yadegari, R.; Fischer, R.L.; Yanofsky, M.F.; Weigel, D. The role of *JAGGED* in shaping lateral organs. *Development* **2004**, *131*, 1101–1110. [[CrossRef](#)] [[PubMed](#)]
28. Sauret-Güeto, S.; Schiessl, K.; Bangham, A.; Sablowski, R.; Coen, E. *JAGGED* controls *Arabidopsis* petal growth and shape by interacting with a divergent polarity field. *PLoS Biol.* **2013**, *11*, e1001550. [[CrossRef](#)] [[PubMed](#)]
29. Schiessl, K.; Muiño, J.M.; Sablowski, R. *Arabidopsis* *JAGGED* links floral organ patterning to tissue growth by repressing Kip-related cell cycle inhibitors. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2830–2835. [[CrossRef](#)] [[PubMed](#)]
30. Juntheikki-Palovaara, I.; Tähtiharju, S.; Lan, T.Y.; Broholm, S.K.; Rijpkema, A.S.; Ruonala, R.; Kale, L.; Albert, V.A.; Teeri, T.H.; Elomaa, P. Functional diversification of duplicated *CYC2* clade genes in regulation of inflorescence development in *Gerbera hybrida* (Asteraceae). *Plant J.* **2014**, *79*, 783–796. [[CrossRef](#)]
31. Johnson, K.L.; Ramm, S.; Kappel, C.; Ward, S.; Leyser, O.; Sakamoto, T.; Kurata, T.; Bevan, M.W.; Lenhard, M. The *Tinkerbelle* (*Tink*) Mutation Identifies the dual-specificity MAPK phosphatase *INDOLE-3-BUTYRIC ACID-RESPONSE5* (*IBR5*) as a novel regulator of organ size in *Arabidopsis*. *PLoS ONE* **2015**, *10*, e0131103.
32. Ren, H.B.; Dang, X.; Yang, Y.Q.; Huang, D.Q.; Liu, M.T.; Gao, X.W.; Lin, D.S. *SPIKE1* activates ROP GTPase to modulate petal growth and shape. *Plant Physiol.* **2016**, *172*, 358–371. [[CrossRef](#)] [[PubMed](#)]

33. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
34. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
35. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079.
36. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [[CrossRef](#)] [[PubMed](#)]
37. Islam, M.R.; Zhang, Y.; Li, Z.Z.; Liu, H.; Chen, J.M.; Yang, X.Y. Genetic diversity, population structure, and historical gene flow of *Nelumbo lutea* in USA using microsatellite markers. *Aquat. Bot.* **2020**, *160*, 103162. [[CrossRef](#)]
38. Li, C.; Mo, H.B.; Tian, D.K.; Xu, Y.X.; Meng, J.; Tilt, K. Genetic diversity and structure of American lotus (*Nelumbo lutea* Willd.) in North America revealed from microsatellite markers. *Sci. Hortic.* **2015**, *189*, 17–21. [[CrossRef](#)]
39. Guo, H.B.; Li, S.M.; Peng, J.; Ke, W.M. Genetic diversity of *Nelumbo* accessions revealed by RAPD. *Genet. Resour. Crop Evol.* **2007**, *54*, 741–748. [[CrossRef](#)]
40. Liu, Z.W.; Zhu, H.L.; Zhou, J.H.; Jiang, S.J.; Wang, Y.; Kuang, J.; Ji, Q.; Peng, J.; Wang, J.; Gao, L.; et al. Resequencing of 296 cultivated and wild lotus accessions unravels its evolution and breeding history. *Plant J.* **2020**, *104*, 1673–1684. [[CrossRef](#)]
41. Cao, Y.N.; Comes, H.P.; Sakaguchi, S.; Chen, L.Y.; Qiu, Y.X. Evolution of East Asia's Arcto-Tertiary relict *Euptelea* (Eupteleaceae) shaped by Late Neogene vicariance and Quaternary climate change. *BMC Evol. Biol.* **2016**, *16*, 66. [[CrossRef](#)] [[PubMed](#)]
42. Liu, F.L.; Qin, M.; Liu, Q.Q.; Zhang, D.S.; Tian, D.K. Detection of genetic variation between broad and narrow-tepalled American lotus (*Nelumbo lutea* Willd.) by EST-SSR markers. *Acta Agric. Boreali-Occident. Sin.* **2020**, *29*, 306–314. (In Chinese with English Abstract)
43. Li, Z.; Liu, X.Q.; Gituru, R.W.; Juntawong, N.; Zhou, M.Q.; Chen, L.Q. Genetic diversity and classification of *Nelumbo* germplasm of different origins by RAPD and ISSR analysis. *Sci. Hortic.* **2010**, *125*, 724–732. [[CrossRef](#)]
44. Brock, M.T.; Weinig, C. Plasticity and environment-specific covariances: An investigation of floral-vegetative and within flower correlations. *Evolution* **2007**, *61*, 2913–2924. [[CrossRef](#)] [[PubMed](#)]
45. Pyke, K.A.; Page, A.M. Plastid ontogeny during petal development in *Arabidopsis*. *Plant Physiol.* **1998**, *116*, 797–803. [[CrossRef](#)] [[PubMed](#)]
46. Chacón, B.; Ballester, R.; Birlanga, V.; Rolland-Lagan, A.G.; Pérez-Pérez, J.M. A quantitative framework for flower phenotyping in cultivated carnation (*Dianthus caryophyllus* L.). *PLoS ONE* **2013**, *8*, e82165. [[CrossRef](#)] [[PubMed](#)]
47. Marsden-Jones, E.M.; Turrill, W.B. A quantitative study of petal size and shape in *Saxifraga granulata* L. *J. Genet.* **1947**, *48*, 206–218. [[CrossRef](#)] [[PubMed](#)]
48. He, Z.H.; Cheeseman, I.; He, D.Z.; Kohorn, B.D. A cluster of five cell wall-associated receptor kinase genes, *Wak1–5*, are expressed in specific organs of *Arabidopsis*. *Plant Mol. Biol.* **1999**, *39*, 1189–1196. [[CrossRef](#)] [[PubMed](#)]
49. Kohorn, B.D.; Kohorn, S.L. The cell wall associated kinases, WAKs, as pectin receptors. *Front. Plant Sci.* **2012**, *3*, 88. [[CrossRef](#)]
50. Wagner, T.A.; Kohorn, B.D. Wall-associated kinases are expressed throughout plant development and are required for cell expansion. *Plant Cell* **2001**, *13*, 303–318. [[CrossRef](#)] [[PubMed](#)]
51. Abraham, M.C.; Methetrairut, C.; Irish, V.F. Natural variation identifies multiple loci controlling petal shape and size in *Arabidopsis thaliana*. *PLoS ONE* **2013**, *8*, e56743.