



Article

# Genome-Wide Association Analysis of Fruit Shape-Related Traits in *Areca catechu*

Hao Ding <sup>†</sup>, Guangzhen Zhou <sup>†</sup>, Long Zhao , Xinyu Li, Yicheng Wang, Chengcai Xia, Zhiqiang Xia and Yinglang Wan <sup>\*</sup>

Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, College of Tropical Crops, Hainan University, Haikou 570228, China

<sup>\*</sup> Correspondence: ylwan@hainanu.edu.cn

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** The areca palm (*Areca catechu* L.) is one of the most economically important palm trees in tropical areas. To inform areca breeding programs, it is critical to characterize the genetic bases of the mechanisms that regulate areca fruit shape and to identify candidate genes related to fruit-shape traits. However, few previous studies have mined candidate genes associated with areca fruit shape. Here, the fruits produced by 137 areca germplasms were divided into three categories (spherical, oval, and columnar) based on the fruit shape index. A total of 45,094 high-quality single-nucleotide polymorphisms (SNPs) were identified across the 137 areca cultivars. Phylogenetic analysis clustered the areca cultivars into four subgroups. A genome-wide association study that used a mixed linear model identified the 200 loci that were the most significantly associated with fruit-shape traits in the germplasms. In addition, 86 candidate genes associated with areca fruit-shape traits were further mined. Among the proteins encoded by these candidate genes were UDP-glucosyltransferase 85A2, the ABA-responsive element binding factor GBF4, E3 ubiquitin-protein ligase SIAH1, and LRR receptor-like serine/threonine-protein kinase ERECTA. Quantitative real-time polymerase chain reaction (qRT-PCR) analysis showed that the gene that encoded UDP-glycosyltransferase, *UGT85A2*, was significantly upregulated in columnar fruits as compared to spherical and oval fruits. The identification of molecular markers that are closely related to fruit-shape traits not only provides genetic data for areca breeding, but it also provides new insights into the shape formation mechanisms of drupes.



**Citation:** Ding, H.; Zhou, G.; Zhao, L.; Li, X.; Wang, Y.; Xia, C.; Xia, Z.; Wan, Y. Genome-Wide Association Analysis of Fruit Shape-Related Traits in *Areca catechu*. *Int. J. Mol. Sci.* **2023**, *24*, 4686. <https://doi.org/10.3390/ijms24054686>

Academic Editor: Daniela Trono

Received: 2 February 2023

Revised: 20 February 2023

Accepted: 24 February 2023

Published: 28 February 2023



**Copyright:** © 2023 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/>).

**Keywords:** *Areca catechu* L.; fruit shape; single-nucleotide polymorphism (SNP); genome-wide association analysis (GWAS); candidate gene

## 1. Introduction

Fruits, generally developed from the ovary of a plant, protect the seed for propagation and reproduction during plant growth and development [1,2]. Fruit is also an economically important organ for numerous crops. As one of the characteristics of fruit appearance, shape directly affects the commercial value and economic benefits of fruit [3,4]. Therefore, fruit shape is an influential selection feature in a variety of breeding programs [5]. Hormonal regulation, varietal characteristics, plant nutrient levels, and the features of the cultivation environment (such as temperature, light, and moisture) are the main factors that control fruit shape [6,7].

Fruit shape is a complex trait that is genetically controlled by multiple genes in different ways [8]. Recent extensive genetic studies of fruit shapes have resulted in a series of advances [3,9,10]. Tomato fruit shape has received particularly thorough research attention, and several genes have been shown to participate in regulating fruit shape [11–14]. For example, *SUN*, a member of the IQ67-domain (IQD) protein family that encodes calmodulin binding proteins, is highly expressed in tomato ovaries and young fruits; *SUN* expression not only affects the morphology of the vegetative organs, but it also leads to the

elongation of the flower organs and fruits during peak cell division after pollination [11]. Transposon-mediated doubling events led to a significant increase in the *SUN* expression level and promoted fruit elongation [11,15]. *OVATE*, located on tomato chromosome 2, is one of the main members of the *OVATE* family, which inhibits fruit elongation [16,17]. The early appearance of a stop codon in the *OVATE* protein sequence, caused by a single base mutation, can alter the shape of the tomato fruit from round to pear-like. Previous studies have shown that both the *SOV1* locus on chromosome 10 and the *SOV2* locus on chromosome 11 can promote tomato elongation by inhibiting *OVATE* expression [16,17]. Furthermore, *FASCIATED* (*FAS*) and *LOCULE NUMBER* (*LC*) affect fruit shape by changing the locule number in tomatoes [18–20]. The previous study showed that MuMADS1 could interact with the *OVATE* family protein MaOFP1 and regulate fruit quality in a tomato ovate mutant [21]. In apples, the ectopic expression of *Malus domestica* *PISTILLATA* (*MdPI*), the ortholog of the floral organ identity gene *PISTILLATA* (*PI*), regulates apple fruit tissue growth and shape [22].

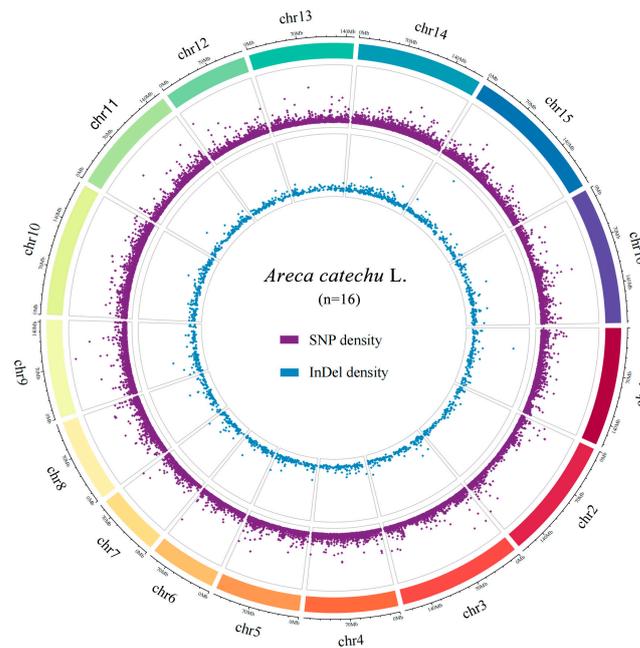
With the rapid development of high-throughput genotyping, genome-wide association analysis (GWAS) has recently become a powerful approach for explorations of the genetic mechanisms that underlie fruit-shape traits [23–25]. A previous GWAS identified a 1.7 Mb inversion on peach chromosome 6 that was associated with peach fruit shape [25]. Also near this region is *PpOFP1*, which inhibits vertical peach growth by interacting with an activated factor of fruit elongation, *PpTRM17*, resulting in flat-shaped fruits [25]. In addition, grape fruit-shape-related genes were obtained by GWAS analysis based on the whole-genome resequencing of 472 grape germplasm resources [26]. This study identified a SNP site (Chr7.20085154; mutation of “C” to “T”) on grape chromosome 7 associated with fruit shape regulation; this site was associated with a gene that encodes the serine/threonine protein kinase *SRK2A* [26,27]. Fruit-shape regulation mechanisms have been analyzed using GWAS in a variety of other fruits, including melon and watermelon [10,28]. These previous studies provide a framework that we can use to explore the mechanisms that regulate areca fruit shape.

*A. catechu*, a perennial casuarina in the palm family (*Areaceae*), is endemic to Malaysia. The areca palm is one of the most economically important trees in the tropics due to its high medicinal and edible value [29,30]. Areca nuts, which are dried mature areca fruit, are also a popular food, and more than 400 million people habitually eat areca nuts in China, South Asia, East Africa, and other countries and regions [29,31]. As Chinese consumers have a strong preference for oval areca nuts, it is important to clarify the molecular mechanisms that regulate areca fruit shape for use by the areca industry. Recently, substantial progress has been made in the understanding of fruit-shape regulation mechanisms in tomato, a berry, but little is known about the shape-forming mechanisms of areca fruit, a drupe. In this study, we performed the first GWAS of areca nuts to identify the loci associated with fruit-shaped regulation and to highlight useful molecular markers. We also explored the genetic structure and genetic diversity of areca palm populations.

## 2. Results

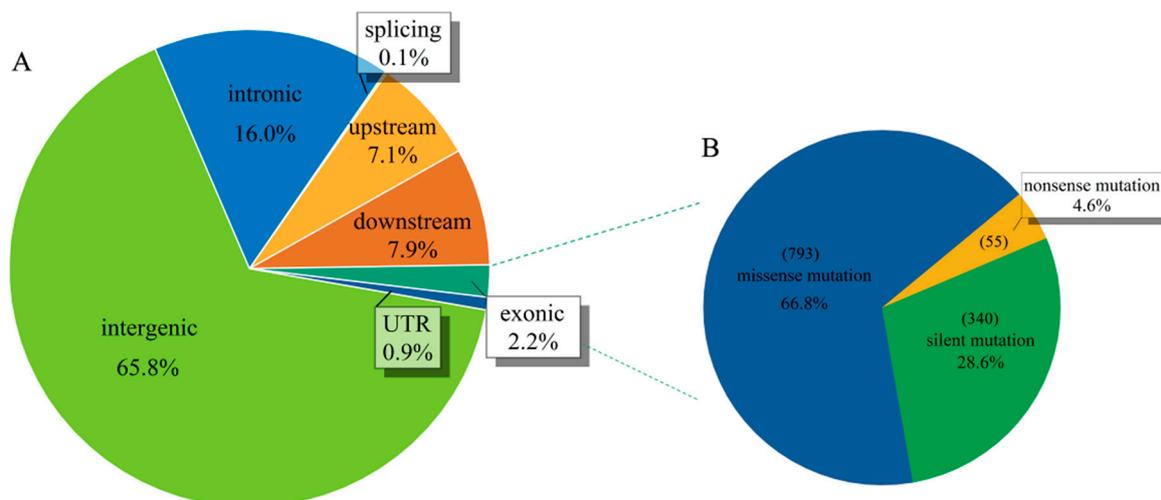
### 2.1. Genotype Analysis of *A. catechu* Populations

After the filtering of data, a total of 714.30 Gb of data were obtained through the sequencing of 137 areca genomic DNA sequences. Then, after filtering all the SNPs and indels (max missing > 0.5, minor allele frequency, MAF > 0.05 and Hardy–Weinberg equilibrium *p*-value > 0.001), we obtained 45,094 high-quality SNPs and 2347 indels across all 16 areca chromosomes (Figure 1). The numbers of variants per chromosome ranged from 1591 (chromosome 7) to 3992 (chromosome 10). These variants covered a total physical distance of ~2.45 Gb, with the shortest distance on chromosome 7 (92.0 Mb) and the longest distance on chromosome 15 (204.2 Mb). The genome-wide average variant density was 51.9 kb per variant and varied from 46.2 kb (chromosome 14) to 60.9 kb (chromosome 2) (Figure 1; Table S2).



**Figure 1.** Distribution of variation detection types in areca catechu genome. SNP: single nucleotide polymorphism; InDel: insertion and deletion. The outermost circle shows the 16 chromosomes of *A. catechu*; purple dots in the middle circle show the distribution of SNPs across the genome; blue dots in the innermost circle show the distribution of InDels across the genome.

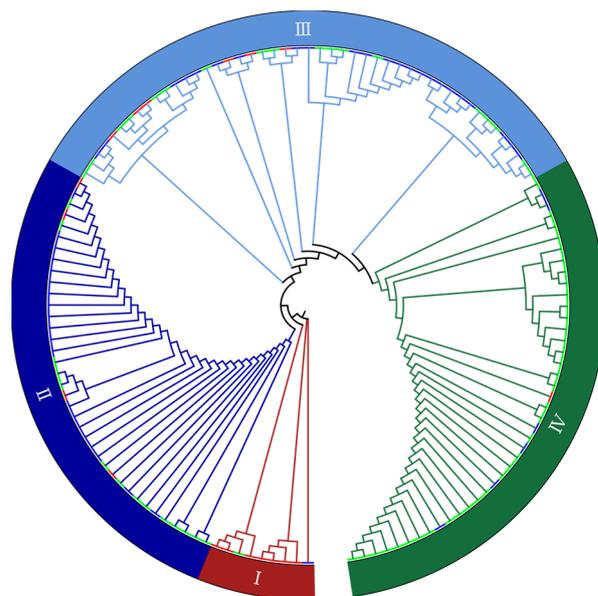
We detected the following two types of SNP markers in the areca genome: transitions (ts) and transversions (tv). ‘GA’-type mutations were the most common transitions (5927) and ‘TA’-type mutations were the most common transversions (3697). Across the 137 individuals tested, the average value of ts/tv was 2.56, the maximum ts/tv for a single sample was 2.82, and the minimum ts/tv for a single sample was 2.00 (Figure S1). Distribution analysis of the high-quality SNPs showed that 37,003 (65.8%) were located in the intergenic regions, 483 (0.9%) were located in the untranscribed regions, 8984 (15.98%) were located in introns, and only 1219 SNPs (2.17%) were located in coding regions (Figure 2; Table S3). Notably, 340 SNPs (28.6%) in the coding regions produced silent mutations, and 793 SNPs (66.8%) produced missense mutations (Figure 2).



**Figure 2.** SNP location and functional annotation. (A) The distribution of SNP variation sites in different locations. (B) The relative abundance of types of SNP mutations (missense, nonsense, and silent).

### 2.2. Genetic Diversity Revealed by SNP Markers

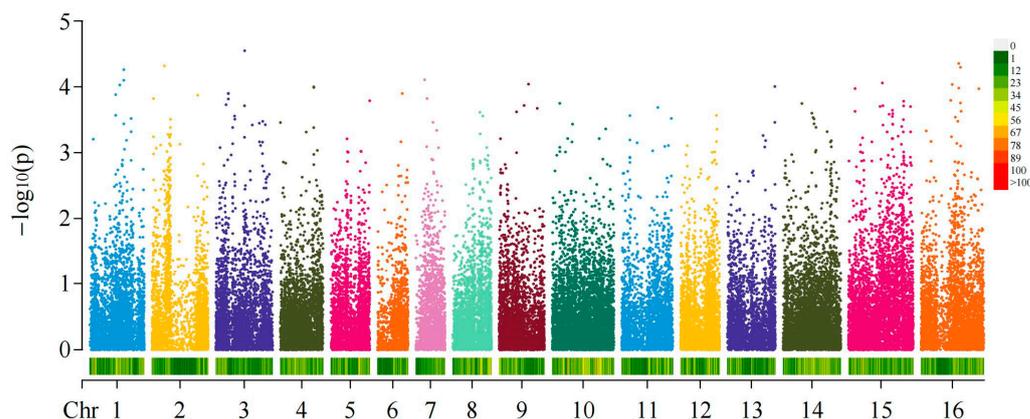
Across the 137 materials tested, fruit shape was roughly divided into spherical, oval, and columnar based on the fruit shape index. PCA and kindship analysis, performed using all high-quality SNPs and indel data, identified an obvious population structure that separated the Hainanese and Vietnamese populations (Figure S2). Our neighbor-joining phylogenetic tree clustered the areca germplasms with different fruit-shaped traits into four population subgroups. Areca trees with oval fruit were mainly clustered into subgroup 1, those with spherical fruit were mainly clustered into subgroup 2, and those with columnar fruit were mainly clustered into subgroup 4; subgroup 3 contained a mixture of fruit shapes (Figure 3).



**Figure 3.** Genetic relationship analysis. Neighbor-joining phylogeny constructed based on 137 areca germplasms. Roman numbers in the outermost circle indicate subgroups. Branch color corresponds to fruit shape, with red (I), blue (II), light blue (III) and green (IV) corresponding to areca germplasms with oval, spherical, mixed shape and columnar fruits, respectively.

### 2.3. GWAS of Fruit Shapes and Areca Germplasms

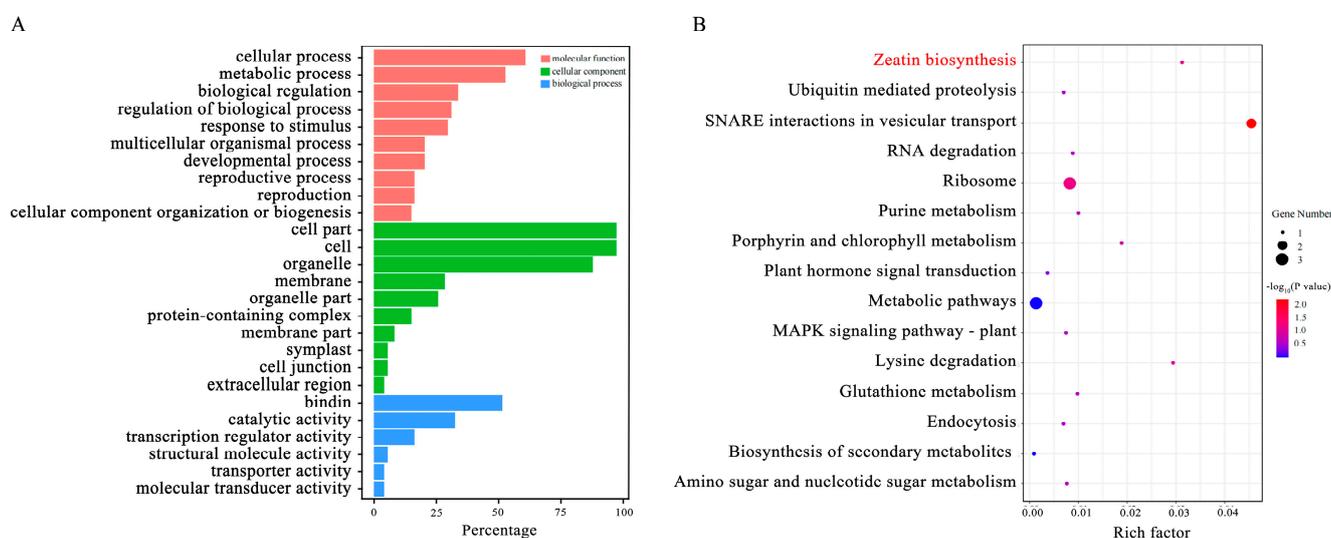
Among the materials tested of the 137 areca germplasms, the fruit shapes were roughly divided into spherical (fruit shape index, FSI < 1.38), oval (1.38 < FSI < 1.73) and columnar (FSI > 1.73) based on the areca FSI (Figure S3; Table S1). We performed genome-wide association analysis (GWAS) by using a compressed MLM model on the data of fruit-shape traits (Table S1). Then, we set the  $p$ -value of the threshold as  $1.25 \times 10^{-3}$  to detect significant loci. In the MLM association analysis of the 45,094 SNPs and areca fruit shapes, we finally obtained 200 loci that were significantly associated with fruit-shape traits. After annotating genes located at or within 50 kb of the top 200 most significant loci, 86 candidate genes were identified (Figure 4 and Figure S3; Table S4).



**Figure 4.** Genome-wide association analysis of the fruit shapes of areca germplasms. Manhattan plot shows the results of the whole-genome association analysis of fruit shape.

#### 2.4. Candidate Genes

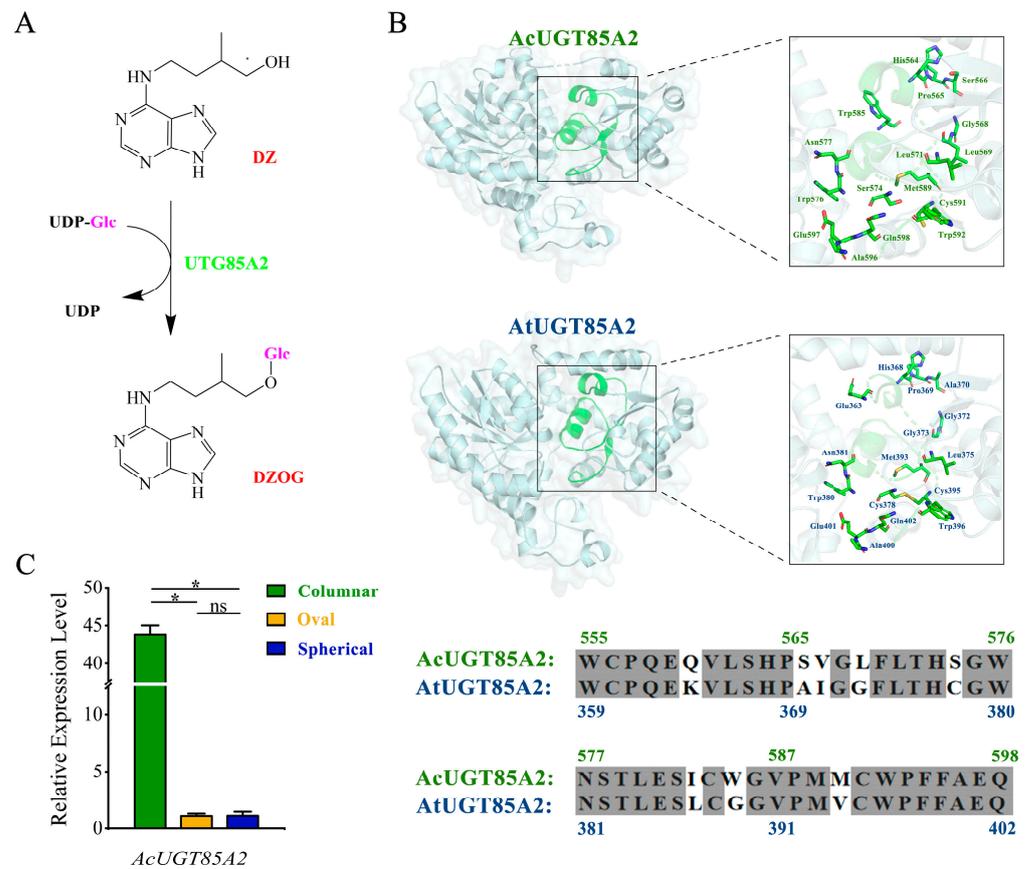
We obtained 86 candidate genes associated with fruit-shape traits, each of which was located within 50 kb of a significant mutation site. We next analyzed the Gene Ontology (GO) enrichment of the 86 candidate genes. Most genes were enriched in cellular component terms. The three most enriched molecular terms were the cellular process part, metabolic process, and regulation of the biological process (Figure 5A). In the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, the three most enriched pathways were SNARE interactions in vesicular transport and ribosome, and zeatin biosynthesis (Figure 5B; Table S5).



**Figure 5.** Functional annotation of candidate genes. (A) Gene Ontology (GO) annotation of the candidate genes. (B) Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation of the candidate genes.

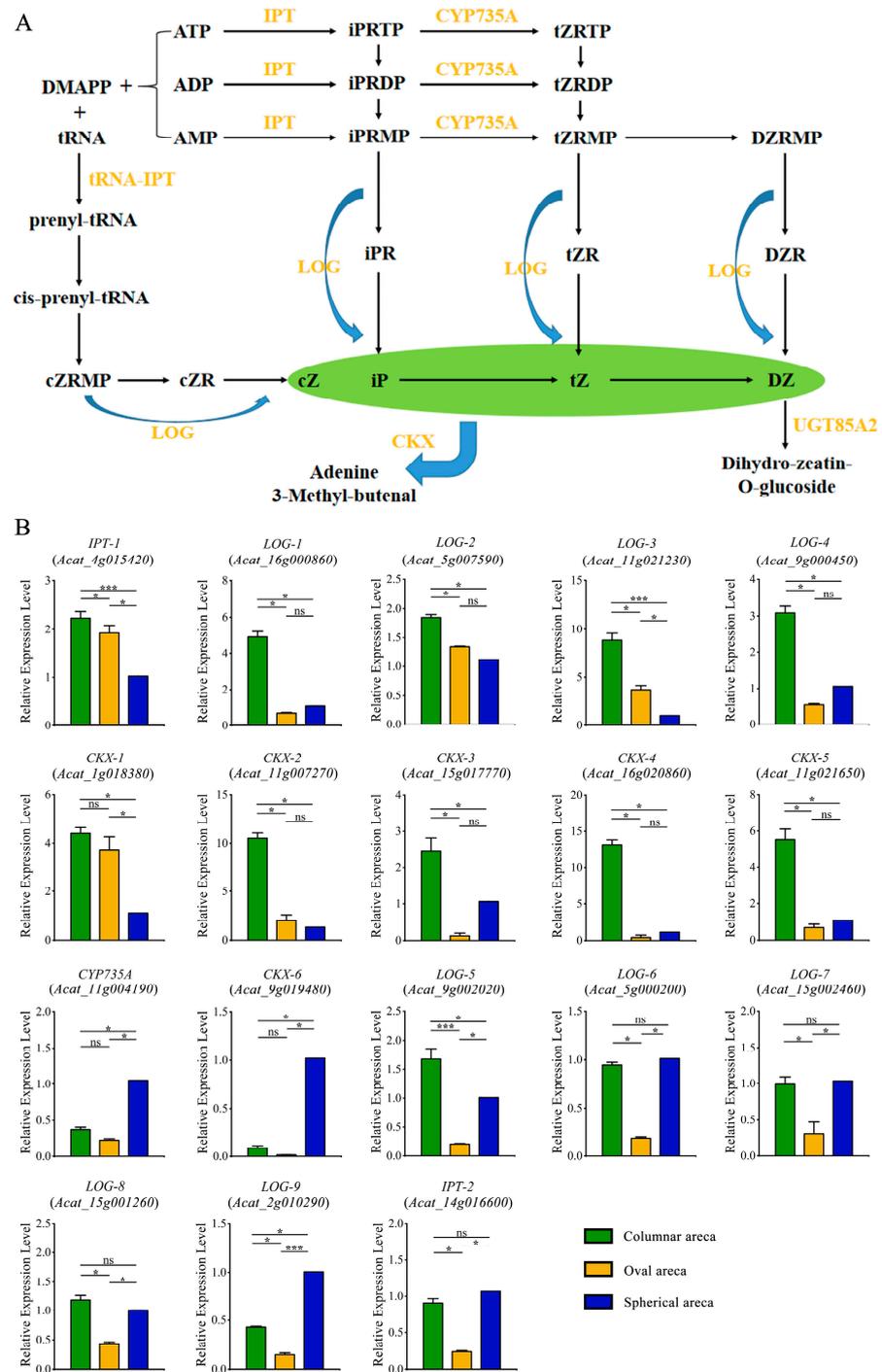
Notably, one of the candidate genes on chromosome 2 encoded UDP-glucosyltransferase in the zeatin synthesis pathway. Three ribosomal protein coding genes were found near significant sites on chromosomes 2 (one gene) and 15 (two genes). Furthermore, we identified genes that encoded the biodegradation-related proteins E3 ubiquitin protein ligase, *n*-trimethyltransferase, and *CCR4-NOT* gene on chromosomes 6, 8 and 13, respectively. Three genes that encoded proteins involved in the transport of plant organic matter (syntaxin, protein transport protein SFT1, and charged multivesicular body protein 1) were identified on chromosomes 2, 10, and 13, respectively. Because areca fruit-shape traits may be regulated by many factors, the identification of these candidate genes provides important genetic resources for the analysis of fruit shape regulation mechanisms (Table S5).





**Figure 7.** Verification of candidate genes. (A) Chemical structures of compounds and biosynthetic pathways related to UGT85A2. (B) The complete predicted structure of AcUGT85A2 (**upper left**), the functional domain of UDP-glycosyltransferase (**upper right**), the complete predicted structure of AtUGT85A2 (**lower left**), and the functional domain of UDP-glycosyltransferase (**lower right**). Differences in the amino acid sequences of the conserved functional domains of AcUGT85A2 and AtUGT85A2; fully conserved residues are shaded in gray. (C) Experimental verification of *AcUGT85A2* expression levels using qRT-PCR. Statistical significance was assessed by an unpaired two-tailed Student's *t*-test. Symbols for statistical significance levels: \*: significant differences ( $p < 0.05$ ), ns: no significant differences,  $n = 3$ .

To further investigate whether cytokinins participate in the development of fruit shape, we identified 18 genes related to the cytokinin metabolism in the areca genome. Then, the expression profiles of these genes in different fruit shapes were obtained using qRT-PCR (Figure 8A,B). We found that most genes were differentially expressed in differently shaped areca nuts. Notably, 11 of the 18 genes (e.g., *IPT-1*, *LOG-1*, and *CKX-3*) were significantly upregulated in columnar areca fruits, as compared to spherical or oval fruits. This was consistent with the expression pattern of *AcUGT85A2* (Figure 7B). In addition, three genes (*CYP735A*, *CKX-6* and *LOG-9*) were significantly upregulated in spherical areca fruits, as compared to columnar or oval fruits. In general, genes tended to be upregulated in columnar and spherical areca, as compared to oval areca. The shape-related differences in the expression patterns of genes related to the zeatin metabolic pathway in areca nuts may indicate that this pathway is involved in the regulation of areca fruit shape (Figure 8B).



**Figure 8.** Expression patterns of genes involved in the cytokinin (CK) synthesis pathway. **(A)** The CK synthesis pathway. **(B)** Experimental verification of the expression patterns of genes associated with the CK synthesis pathway using real-time fluorescent quantitative polymerase chain reaction (qRT-PCR). Statistical significance was assessed by an unpaired two-tailed Student’s t-test Symbols for statistical significance levels: \*: significant differences ( $p < 0.05$ ), \*\*\*: extremely significant differences ( $p < 0.001$ ), ns: no significant differences,  $n = 3$ .

### 3. Discussion

In this study, we found a high diversity of areca fruit shapes, including oval, spherical, and columnar shapes. Fruit morphology is determined by the coordination of cell division and expansion, which are basic cellular processes required for the development of all plant organs [32,33]. Phytohormones are integral to the regulation of fruit development and

maturation; auxin, gibberellic acid (GA), cytokinins (CKs), and others are thought to coordinate the development of fruit shape during cell division and expansion [2]. The molecular mechanisms that underlie the auxin- and GA-mediated regulation of cell division and cell expansion have made great progress. The auxin signaling and auxin–GA interaction pathways rely on Aux/IAA, ARF, and GH3 to affect fruit cell division [34]. Previous studies have shown that four auxin-/GA-responsive ARFs (SlARF6, 8, 10, and 16) redundantly restrict cell division in the tomato pericarp during early growth [33]. However, during auxin accumulation, the transcription of these four *ARF* genes was inhibited by *miR160* and *miR167*, thus promoting cell division in the early tomato pericarp [35,36]. Furthermore, at the beginning of tomato fruit cell expansion, the increased auxin and GA levels in seeds promote the expression of *GA3ox* and *GA20ox* in fruit tissues, thereby increasing the cellular concentration of bioactive GA [37]. The ability of auxin to stimulate GA biosynthesis has been demonstrated in pears (*Pyrus* spp.), where the exogenous application of 2,4-D (a synthetic auxin mimic) stimulated the expression of *GA20ox* and *GA3ox*, increased the accumulation of bioactive GA in the fruit tissues, and upregulated both cell division and expansion [38].

In addition to auxin and GA, several studies have shown that CK input can also influence cell division and expansion. For example, Zhang et al. reported the specific downregulation of CK dehydrogenase (CKX) in short bottle gourds, indicating that CKs can act in concert with or replace auxin and GA to establish fruit morphology [39]. Our results are consistent with the previous study, as most of the *CKX* genes tended to be upregulated in columnar and spherical areca as compared to oval areca, indicating that *CKX* genes may be involved in the regulation of areca fruit shape. In *Arabidopsis*, CKs directly regulate plant size. Specifically, mutations that reduce AHP activity lead to CK insensitivity and smaller siliques, while decreased *CKX* expression results in increased CK levels and larger siliques [40]. The shape-related differences in the expression patterns of genes related to the CKs synthesis pathway in areca nuts indicate that this pathway is involved in the regulation of areca fruit shape. Furthermore, *UGT85A2* encodes UDP-glycosyltransferase, which is an important regulator of CK levels in the cytokinin synthesis pathway. A previous phenotypic analysis of two *ugt85a3* mutant lines found that *UGT85A3*, a member of the *UGT85A* family, negatively regulated the number of ovules and seeds in *Arabidopsis thaliana* [41]. It has also been demonstrated that the knockdown of *CsUGT85A2* in cucumbers induced the formation of a super ovary phenotype, leading to fuller fruits [42]. Similarly, *AcUGT85A2* was significantly upregulated in columnar areca nuts as compared to fruits of other shapes. This suggested that *AcUGT85A2* may play a role in the formation of areca fruit morphology by regulating CKs, especially DZ.

Fruit shape is a polygenic quantitative genetic trait that has a complex genetic mechanism. Previous studies have performed relatively thorough genetic mappings of fruit shape in fruit vegetable crops, such as tomato, and four important genes that regulate tomato fruit shape have been successfully located and cloned, indicating that fruit shape in tomato is co-regulated by multiple genes [11–14]. In general, gene regulation patterns are polygenic and have complex quantitative genetic characteristics. In grapes, studies based on GWAS have identified some candidate genes related to fruit shape and these genes are mainly related to LRR receptor-like serine/threonine-protein kinase, transcription factors, ubiquitin ligases and plant hormones [27]. Here, we also identified several genes related to fruit-shape traits in areca. *Acat\_6g004190*, which encodes RING-type E3 ubiquitin-protein ligase SiAH1, is involved in ubiquitin-mediated proteolysis. RING-type E3 is a ubiquitin ligase that determines seed size by regulating gametogenesis and cell cycle processes. Additionally, the regulation of plant seed organ morphology by RING-type E3 has been studied extensively [43,44]. *Acat\_3g012500* encodes the LRR receptor-like serine/threonine-protein kinase ERECTA, which is involved in the MAPK signaling pathway in plants. Mapping-based cloning methods have identified LRR-RLK protein kinase as a possible candidate gene for the shape of peach pits [45]. Further studies of the candidate genes identified

herein using GWAS are required to further clarify the molecular mechanisms that regulate nut shape in areca palms.

## 4. Materials and Methods

### 4.1. Material Collection and Phenotype Evaluation

To dissect the genetic bases of areca nut shape, we performed a GWAS including 137 areca accessions collected from areca plantations. Of these, 125 accessions originated from the Hainan Province, China, and 12 accessions originated from Vietnam. Vernier calipers were used to measure fruit length and diameter. The fruit shape index (FSI), which is defined as the ratio of fruit length to fruit diameter, and other statistics were calculated using Microsoft Excel 2021. Areca fruit shapes were distinguished, identified, and recorded based on the FSI (Table S1).

### 4.2. DNA Preparation and Sequencing

A modified cetyltrimethylammonium bromide (CTAB) method was used to extract genomic DNA from areca leaves [46]. After quantification using 1% agarose gel electrophoresis, the working DNA solutions were diluted to 100 ng/ $\mu$ L and stored at  $-20^{\circ}\text{C}$ . The hyper-seq method was used to construct sequencing libraries for 137 areca DNA samples [47]. After monoclonal detection met the requirements, the HiSeq 2500 platform was used to perform paired-end 150-bp sequencing for each constructed sequencing library.

### 4.3. SNP Calling and Annotation

We tested the raw data using the FastQC software (<https://github.com/s-andrews/FastQC>, accessed on 20 June 2022) [48]. The clean reads were obtained using the fastp software (<https://github.com/OpenGene/fastp>, accessed on 20 June 2022) by filtering the raw data that passed the test. The BWA software was not only used to establish the index file of the areca reference genome (<https://ftp.cngb.org/pub/CNSA/data3/CNP0000517>, accessed on 20 June 2022), but was also used for the sequence comparison of the clean reads of each sample to obtain the SAM file [49,50]. Samtools was used to reorder and convert the SAM file into a BAM file, followed by another sorting process, and the final generation of the index file [51]. Picard (<https://github.com/broadinstitute/picard/releases/latest>, accessed on 20 June 2022) was used to mark the java environment duplicates, which were then used to create citations for the new BAM files.

The variation detection followed the best practice workflow recommended by GATK [52]. The HaplotypeCaller application of the GATK software was used to generate each sample GVCF file (GVCF format includes all variant types, including SNPs and Indel, which need to be further filtered) [52]. Finally, the GVCF was merged using CombineGVCFs to obtain the total merged GVCF file, followed by the use of GenotypeGVCFs to extract genotypes. The filtered VCF files were then obtained using VCFtools (<https://vcftools.github.io/index.html>, accessed on 20 June 2022) with an MAF  $> 0.05$  (minimum allele frequency of 0.05, post removal of rare alleles) and HWE  $> 0.001$  (removal of loci that do not satisfy the Hardy–Weinberg Equilibrium ( $p < 0.001$ )) [53]. The VCFtools were further used to remove the variant loci with a  $>50\%$  deletion rate of all materials, to obtain the final filtered high-quality variant loci [53]. Using SnpEff software (<http://pcingola.github.io/SnpEff/>, accessed on 20 June 2022), we compared the sequencing reads to the reference genome to identify mutation locations (i.e., intergenic regions, untranslated regions, upstream gene regions, and downstream gene regions) and mutation types (i.e., missense, nonsense, and silent), while simultaneously annotating the mutations [54].

### 4.4. Genetic Diversity Analysis

First, we used the dnadist program by PHYLIP (<http://evolution.genetics.washington.edu/phylip.html>, accessed on 22 June 2022) to calculate the genetic distance matrix for all the samples and notepad++ software (<https://notepad-plus.en.softonic.com/>, accessed on 22 June 2022) to save the genetic distance matrix file in an appropriate format [55]. A

phylogenetic tree was constructed based on this distance matrix using the neighbor-joining method and drawn, based on the spanning tree file, using the online tool iTOL (<https://itol.embl.de/itol.cgi>, accessed on 22 June 2022). A principal component analysis (PCA) of the SNPs detected in the *Areca catechu* populations was performed using GCTA software [56]. Finally, we used R software (<https://www.r-project.org/>, accessed on 22 June 2022) to draw the PCA scatter plot and to construct a population–genetic structure matrix based on the genetic composition coefficient (Q) of each sample within each subpopulation.

#### 4.5. Association Analysis

We used TASSEL 5.0 (<https://tassel.bitbucket.io/>, accessed on 23 June 2022) to analyze the associations between areca fruit shape and these SNPs using a mixed linear model (MLM) [57]. In this analysis, the first three PCA values (eigenvectors) were used as fixed effects in the mixed model to correct for stratification. The formula used was  $Y = \alpha X + \beta Q + \mu K + e$ , where X represents the genotype, Y represents the phenotype, Q represents the results of the PCA, K represents the indicator matrix of the individual genetic relationships,  $\alpha$  represents the estimated fixed effect parameter,  $\beta$  represents the SNP effect,  $\mu$  represents the predicted random individual, and e represents the random residual, such that  $e \sim (0, \delta_e^2)$ . We calculated the association values and identified the SNPs that were significantly associated with areca fruit-shape traits. The contribution rates of the different SNP loci to the phenotype were calculated. The significance limit of the tolerance level was set at 0.05. Finally, we used the CMplot R package (<https://github.com/YinLiLin/CMplot>, accessed on 23 June 2022) to draw Manhattan plots.

#### 4.6. Candidate Gene Screening

We set 100 kb as the size of the flanking region (50 kb upstream to 50 kb downstream of the significantly associated SNP markers) that corresponded to the relevant SNP loci. Based on areca SNP annotations and the functional annotations of the loci, the genes on which the loci were located were considered as candidate genes. If a locus was located simultaneously upstream and downstream of other genes, the upstream and downstream genes were also considered as candidate genes. If a locus was located in the intergenic region, the upstream and downstream genes closest to the locus were considered as candidate genes.

#### 4.7. Bioinformatics Analysis

To perform bioinformatic analysis of the target protein, the online tool Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=help>, accessed on 15 August 2022) was used to predict the tertiary structure, function, and conserved domains of UTG85A2 in *A. thaliana* and *A. catechu*, using the homologous modeling method. The protein model with the highest predictive confidence was selected, and the 3D structure and conserved domains of the enzyme protein were mapped using PyMOL (<https://www.pymol.org/>, accessed on 16 August 2022).

#### 4.8. Real-Time Fluorescence Quantitative PCR (qRT-PCR) Verification

RNA was extracted from areca nuts with different shapes using extraction reagents (Tiangen, Beijing, China). In addition, reverse transcription was performed with the One-Step gDNA Removal and cDNA Synthesis SuperMix (Tiangen, Beijing, China). The candidate genes were analyzed using qRT-PCR. The qRT-PCR primers of the candidate genes listed in Table S6 were designed by Primer Premier 5.0 (<http://www.premierbiosoft.com/primerdesign/>, accessed on 24 September 2022) and  $\beta$ -actin was used as an internal control. The total reaction volume of qRT-PCR was 20  $\mu$ L, containing 10  $\mu$ L of 2  $\times$  SYBR Green Master Mix (Tiangen, China), 1  $\mu$ L of the forward and reverse primers, 1  $\mu$ L of the cDNA template, and 7  $\mu$ L of ddH<sub>2</sub>O. The qRT-PCR program was as follows: 94 °C for 30 s, 40 cycles at 94 °C for 30 s, and 60 °C for 30 s. The relative gene expressions were calculated using the  $2^{-\Delta\Delta CT}$  method. In this experiment, each sample was represented by three biological and technical replicates.

## 5. Conclusions

To identify candidate genes associated with areca-nut shape formation, we indexed the fruit shapes of 137 areca accessions. The phylogenetic analysis of high-quality SNP markers grouped these areca accessions into four subpopulations, with close genetic relationships among the populations. A GWAS was performed to identify the significant loci related to fruit-shape traits in the germplasm resources. The proteins encoded by the screened candidate genes, which were mainly associated with biosynthesis, degradation, signal transduction, transport, and metabolic pathways, included the plant hormone glycosyl-transferase, ubiquitin-protein ligase, and LRR receptor-like serine/threonine-protein kinase. Further analysis showed that *UGT85A2*, which regulates the CK level, was differentially expressed across different areca accessions. Our results help to clarify the genetic control of areca fruit-shape traits. The identification of molecular markers closely related to fruit-shape traits is of great significance for the breeding of specific shape varieties, especially those of drupes.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24054686/s1>.

**Author Contributions:** Y.W. (Yinglang Wan) and Z.X. conceived and designed the study. H.D., L.Z., Y.W. (Yicheng Wang) and C.X. performed the experiments. H.D. carried out the statistical analysis. H.D. wrote the manuscript. Y.W. (Yinglang Wan), G.Z. and X.L. revised the manuscript. All authors contributed to manuscript revision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Hainan Province Science and Technology Special Fund (ZDYF2021XDNY121), the National Natural Science Foundation of China (31960064), and the Developing Bioinformatics Platform in the Hainan Yazhou Bay Seed Lab (B21HJ0001).

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

**Informed Consent Statement:** Informed consent was obtained from all the subjects involved in the study.

**Data Availability Statement:** The original contributions presented in the study are publicly available. These data can be found at the following link: <https://bigd.big.ac.cn/gvm/getProjectDetail?Project=GVM000483> (accessed on 1 February 2023).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Gillaspay, G.; Ben-David, H.; Gruissem, W. Fruits: A Developmental Perspective. *Plant Cell* **1993**, *5*, 1439–1451. [[CrossRef](#)]
2. Fenn, M.A.; Giovannoni, J.J. Phytohormones in fruit development and maturation. *Plant J.* **2021**, *105*, 446–458. [[CrossRef](#)]
3. Visa, S.; Cao, C.; Gardener, B.M.; van der Knaap, E. Modeling of tomato fruits into nine shape categories using elliptic fourier shape modeling and Bayesian classification of contour morphometric data. *Euphytica* **2014**, *200*, 429–439. [[CrossRef](#)]
4. Ziaf, K.; Amjad, M.; Shakeel, A.; Saeed, A. Assessment of genetic diversity in tomato for fruit morphology, composition and yield. *Pak. J. Bot.* **2016**, *48*, 2477–2483.
5. Paran, I.; van der Knaap, E. Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J. Exp. Bot.* **2007**, *58*, 3841–3852. [[CrossRef](#)]
6. Christodoulou, A.; Pool, R.; Weaver, R. Prebloom thinning of Thompson Seedless grapes is feasible when followed by bloom spraying with gibberellin. *Calif. Agric.* **1966**, *20*, 8–10.
7. Beppu, K.; Yanaihara, M.; Ikeda, T.; Kataoka, I. Effects of phytohormones on the occurrence of double pistils in sweet cherry (*Prunus savium*). *Hortic. Res.* **2007**, *6*, 123–125.
8. Chunthawodtiporn, J.; Hill, T.; Stoffel, K.; Van, D.A. Quantitative trait loci controlling fruit size and other horticultural traits in bell pepper (*Capsicum annuum*). *Plant Genome* **2018**, *11*, 160125. [[CrossRef](#)]
9. Wei, Q.Z.; Fu, W.Y.; Wang, Y.Z.; Qin, X.D.; Wang, J.; Li, J.; Lou, Q.F.; Chen, J.F. Rapid identification of fruit length loci in cucumber (*Cucumis sativus* L.) using next-generation sequencing (NGS)-based QTL analysis. *Sci. Rep.* **2016**, *6*, 27496. [[CrossRef](#)]
10. Dou, J.; Zhao, S.; Lu, X.; He, N.; Zhang, L.; Ali, A.; Kuang, H.; Liu, W. Genetic mapping reveals a candidate gene (*CIF51*) for fruit shape in watermelon (*Citrullus lanatus* L.). *Theor. Appl. Genet.* **2018**, *131*, 947–958. [[CrossRef](#)]
11. Xiao, H.; Jiang, N.; Schaffner, E.; Stockinger, E.J.; van der Knaap, E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **2008**, *319*, 1527–1530. [[CrossRef](#)]
12. Liu, J.; Van, E.J.; Cong, B.; Tanksley, S.D. A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13302–13306. [[CrossRef](#)]

13. Li, H.; Qi, M.; Sun, M.; Liu, Y.; Liu, Y.; Xu, T.; Li, Y.; Li, T. Tomato transcription factor *SIWUS* plays an important role in tomato flower and locule development. *Front. Plant Sci.* **2017**, *8*, 457. [[CrossRef](#)] [[PubMed](#)]
14. Xu, C.; Liberatore, K.L.; MacAlister, C.A.; Huang, Z.; Chu, Y.H.; Jiang, K.; Brooks, C.; Ogawa-Ohnishi, M.; Xiong, G.; Pauly, M.; et al. A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nat. Genet.* **2015**, *47*, 784–792. [[CrossRef](#)]
15. Jiang, N.; Gao, D.; Xiao, H.; van der Knaap, E. Genome organization of the tomato sun locus and characterization of the unusual retrotransposon rider. *Plant J.* **2009**, *60*, 181–193. [[CrossRef](#)]
16. Rodríguez, G.R.; Kim, H.J.; van der Knaap, E. Mapping of two suppressors of *OVATE* (*sov*) loci in tomato. *Heredity* **2013**, *111*, 256–264. [[CrossRef](#)]
17. Wang, Y.; Clevenger, J.P.; Illa-Berenguer, E.; Meulia, T.; van der Knaap, E.; Sun, L. A Comparison of *sun*, *ovate*, *fs8.1* and auxin application on tomato fruit shape and gene expression. *Plant Cell Physiol.* **2019**, *60*, 1067–1081. [[CrossRef](#)] [[PubMed](#)]
18. Rodríguez, G.R.; Muños, S.; Anderson, C.; Sim, S.C.; Michel, A.; Causse, M.; Gardener, B.B.; Francis, D.; van der Knaap, E. Distribution of *SUN*, *OVATE*, *LC*, and *FAS* in the tomato germplasm and the relationship to fruit shape diversity. *Plant Physiol.* **2011**, *156*, 275–285. [[CrossRef](#)]
19. Cong, B.; Barrero, L.S.; Tanksley, S.D. Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat. Genet.* **2008**, *40*, 800–804. [[CrossRef](#)] [[PubMed](#)]
20. Muños, S.; Ranc, N.; Botton, E.; Bérard, A.; Rolland, S.; Duffé, P.; Carretero, Y.; Le, P.M.C.; Delalande, C.; Bouzayen, M.; et al. Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near *WUSCHEL*. *Plant Physiol.* **2011**, *156*, 2244–2254. [[CrossRef](#)] [[PubMed](#)]
21. Liu, J.; Zhang, J.; Wang, J.; Zhang, J.; Miao, H.; Jia, C.; Wang, Z.; Xu, B.; Jin, Z. *MuMADS1* and *MaOFF1* regulate fruit quality in a tomato ovate mutant. *Plant Biotechnol. J.* **2016**, *16*, 989–1001. [[CrossRef](#)]
22. Yao, J.L.; Xu, J.; Tomes, S.; Cui, W.; Luo, Z.; Deng, C.; Ireland, H.S.; Schaffer, R.J.; Gleave, A.P. Ectopic expression of the *PISTILLATA* homologous *MdPI* inhibits fruit tissue growth and changes fruit shape in apple. *Plant Direct* **2018**, *2*, e00051. [[CrossRef](#)]
23. Nordborg, M.; Weigel, D. Next-generation genetics in plants. *Nature* **2008**, *456*, 720–723. [[CrossRef](#)]
24. Korte, A.; Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **2013**, *9*, 29. [[CrossRef](#)] [[PubMed](#)]
25. Guo, J.; Cao, K.; Deng, C.; Li, Y.; Zhu, G.; Fang, W.; Chen, C.; Wang, X.; Wu, J.; Guan, L.; et al. An integrated peach genome structural variation map uncovers genes associated with fruit traits. *Genome Biol.* **2020**, *21*, 258. [[CrossRef](#)] [[PubMed](#)]
26. Liang, Z.; Duan, S.; Sheng, J.; Zhu, S.; Ni, X.; Shao, J.; Liu, C.; Nick, P.; Du, F.; Fan, P.; et al. Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nat. Commun.* **2019**, *10*, 1190. [[CrossRef](#)]
27. Zhang, C.; Cui, L.; Fang, J. Genome-wide association study of the candidate genes for grape berry shape-related traits. *BMC Plant Biol.* **2022**, *22*, 42. [[CrossRef](#)] [[PubMed](#)]
28. Liu, S.; Gao, P.; Zhu, Q.; Zhu, Z.; Liu, H.; Wang, X.; Weng, Y.; Gao, M.; Luan, F. Resequencing of 297 melon accessions reveals the genomic history of improvement and loci related to fruit traits in melon. *Plant Biotechnol. J.* **2020**, *18*, 2545–2558. [[CrossRef](#)]
29. Amudhan, M.S.; Begum, V.H.; Hebbar, K.B. A review on phytochemical and pharmacological potential of *Areca catechu* L. *Seed. Int. J. Pharm. Sci. Res.* **2012**, *3*, 4151–4157.
30. Meng, K.; Mei, F.; Zhu, L.; Chen, H. Arecanut (*Areca catechu* L.) seed polyphenol improves osteoporosis via gut-serotonin mediated Wnt/ $\beta$ -catenin pathway in ovariectomized rats. *J. Funct. Foods* **2021**, *84*, 104598. [[CrossRef](#)]
31. Li, J.; Jia, X.; Liu, L.; Cao, X.; Xiong, Y.; Yang, Y.; Zhou, H.; Yi, M.; Li, M. Comparative biochemical and transcriptome analysis provides insights into the regulatory mechanism of striped leaf albinism in arecanut (*Areca catechu* L.). *Ind. Crop. Prod.* **2020**, *154*, 112734. [[CrossRef](#)]
32. Pattison, R.J.; Catalá, C. Evaluating auxin distribution in tomato (*Solanum lycopersicum*) through an analysis of the *PIN* and *AUX/LAX* gene families. *Plant J.* **2012**, *70*, 585–598. [[CrossRef](#)]
33. Ripoll, J.J.; Zhu, M.; Brocke, S.; Hon, C.T.; Yanofsky, M.F.; Boudaoud, A.; Roeder, A.H.K. Growth dynamics of the *Arabidopsis* fruit is mediated by cell expansion. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 25333–25342. [[CrossRef](#)]
34. Devoghalaere, F.; Doucen, T.; Guitton, B.; Keeling, J.; Payne, W.; Ling, T.J.; Ross, J.J.; Hallett, I.C.; Gunaseelan, K.; Dayatilake, G.A.; et al. A genomics approach to understanding the role of auxin in apple (*Malus domestica*) fruit size control. *BMC Plant Biol.* **2012**, *12*, 7. [[CrossRef](#)] [[PubMed](#)]
35. Liu, X.; Xu, T.; Dong, X.; Liu, Y.; Liu, Z.; Shi, Z.; Wang, Y.; Qi, M.; Li, T. The role of gibberellins and auxin on the tomato cell layers in pericarp via the expression of ARFs regulated by miRNAs in fruit set. *Acta Physiol. Plant* **2016**, *38*, 77. [[CrossRef](#)]
36. He, H.; Yamamuro, C. Interplays between auxin and GA signaling coordinate early fruit development. *Hortic. Res.* **2022**, *9*, uhab078. [[CrossRef](#)]
37. McAtee, P.; Karim, S.; Schaffer, R.; David, K. A dynamic interplay between phytohormones is required for fruit development, maturation, and ripening. *Front. Plant Sci.* **2013**, *4*, 79. [[CrossRef](#)]
38. Cong, L.; Yue, R.; Wang, H.; Liu, J.; Zhai, R.; Yang, J.; Wu, M.; Si, M.; Zhang, H.; Yang, C.; et al. 2,4-D-induced parthenocarpy in pear is mediated by enhancement of GA4 biosynthesis. *Physiol. Plant* **2019**, *166*, 812–820. [[CrossRef](#)]
39. Zhang, H.; Tan, J.; Zhang, M.; Huang, S.; Chen, X. Comparative transcriptomic analysis of two bottle gourd accessions differing in fruit size. *Genes* **2020**, *11*, 359. [[CrossRef](#)]
40. Hussain, Q.; Shi, J.; Scheben, A.; Zhan, J.; Wang, X.; Liu, G.; Yan, G.; King, G.J.; Edwards, D.; Wang, H. Genetic and signalling pathways of dry fruit size: Targets for genome editing-based crop improvement. *Plant Biotechnol. J.* **2020**, *18*, 1124–1140. [[CrossRef](#)]

41. Cucinotta, M.; Manrique, S.; Cuesta, C.; Benkova, E.; Novak, O.; Colombo, L. CUP-SHAPED COTYLEDON1 (*CUC1*) and *CUC2* regulate cytokinin homeostasis to determine ovule number in *Arabidopsis*. *J. Exp. Bot.* **2018**, *69*, 5169–5176. [[CrossRef](#)]
42. Li, X.; Cui, H.; Song, X.; Sun, C.; Ding, Z.; Zhu, X.; Liu, X.; Yan, L. *CsIPT1b* and *CsUGT85A2* delay female corolla opening in cucumber by regulating the content of cytokinins. *Hortic. Plant J.* 2022; *pre-proof*. [[CrossRef](#)]
43. Shu, K.; Yang, W. E3 Ubiquitin Ligases: Ubiquitous actors in plant development and abiotic stress responses. *Plant Cell Physiol.* **2017**, *58*, 1461–1476. [[CrossRef](#)]
44. Trujillo, M. News from the PUB: Plant U-box type E3 ubiquitin ligases. *J. Exp. Bot.* **2018**, *69*, 371–384. [[CrossRef](#)] [[PubMed](#)]
45. López-Girona, E.; Zhang, Y.; Eduardo, I.; Mora, J.R.H.; Alexiou, K.G.; Arús, P.; Aranzana, M.J. A deletion affecting an LRR-RLK gene co-segregates with the fruit flat shape trait in peach. *Sci. Rep.* **2017**, *7*, 6714. [[CrossRef](#)]
46. Stewart, C.N., Jr.; Via, L.E. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques* **1993**, *14*, 748–750.
47. Zou, M.; Xia, Z. Hyper-seq: A novel, effective, and flexible marker-assisted selection and genotyping approach. *Innovation* **2022**, *3*, 100254. [[CrossRef](#)]
48. Vashishtha, K.; Gaud, C.; Andrews, S.; Krueger, C. Librarian: A quality control tool to analyse sequencing library compositions. *F1000Research* **2022**, *11*, 1122. [[CrossRef](#)]
49. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
50. Yang, Y.; Huang, L.; Xu, C.; Qi, L.; Wu, Z.; Li, J.; Chen, H.; Wu, Y.; Fu, T.; Zhu, H.; et al. Chromosome-scale genome assembly of areca palm (*Areca catechu*). *Mol. Ecol. Resour.* **2021**, *21*, 2504–2519. [[CrossRef](#)]
51. 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. [[CrossRef](#)] [[PubMed](#)]
52. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)]
53. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)] [[PubMed](#)]
54. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)]
55. Retief, J.D. Phylogenetic analysis using PHYLIP. *Methods Mol. Biol.* **2000**, *132*, 243–258.
56. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [[CrossRef](#)]
57. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.