horticulturae

MDPI

Article

Identification, Characterization and Comparison of the Genome-Scale UTR Introns from Six Citrus Species

Chunzhen Cheng [1,2,*,†], Xiaobao Shi [2,†], Yongyan Zhang [1], Bin Wang [2], Junwei Wu [2], Shizao Yang [3] and Shaohua Wang [3,*]

[1] College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China; zhyy0425@126.com
[2] College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou 350002, China; galelilr@126.com (X.S.); wb971220@163.com (B.W.); a839941883@126.com (J.W.)
[3] Institute of Tropical and Subtropical Cash Crops, Yunnan Academy of Agricultural Sciences, Baoshan 678000, China; yangshizao2022@163.com
* Correspondence: ld0532cheng@sxau.edu.cn (C.C.); rjswsh@yaas.org.cn (S.W.)
† These authors contributed equally to this work.

Abstract: Ever since their discovery, introns within the coding sequence (CDS) of transcripts have been paid great attention. However, the introns located in the untranslated regions (UTRs) are often ignored. Here, we identified, characterized and compared the UTR introns (UIs) from six citrus species. Results showed that the average intron number of UTRs is greatly lower than that of CDSs. Among all six citrus species, the number and density of 5′UTR introns (5UIs) are higher than those of 3′UTR introns (3UIs). The UI densities varied greatly among different citrus species. There are 11 and 9 types of splice site (SS) pairs for the UIs of *C. sinensis* and *C. medica*, respectively. However, the UIs of the other four citrus species all own only three kinds of SS pairs. The 'GT-AG', accounting for more than 95% of both 5UIs and 3UIs SS pairs for all the six species, is the most popular type. Moreover, 81 5UIs and 26 3UIs were identified as common UIs among the six citrus species, and the transcripts containing these common UIs were mostly involved in gene expression or gene expression regulation. Our study revealed that the UIs' length, abundance, density and SS pair types varied among different citrus species and that many UI-containing genes play important roles in gene expression regulation. Our findings have great implications for future citrus UI function research.

Keywords: citrus; UTR intron; genome-wide identification; gene expression regulation; splice site

Citation: Cheng, C.; Shi, X.; Zhang, Y.; Wang, B.; Wu, J.; Yang, S.; Wang, S. Identification, Characterization and Comparison of the Genome-Scale UTR Introns from Six Citrus Species. *Horticulturae* **2022**, *8*, 434. https://doi.org/10.3390/horticulturae8050434

Academic Editor: Stefano Giovanni La Malfa

Received: 17 April 2022
Accepted: 10 May 2022
Published: 13 May 2022

1. Introduction

Introns, the noncoding regions of RNA transcripts, were independently discovered in 1977 by Philip A. Sharp and Richard J. Roberts using the adenovirus model system [1,2]. Since then, progressive research was performed for the explorations of their origin, emergence, evolution, functions and so on [3]. Noteworthily, introns were found to be harbored in the genomic structure of all eukaryotes but not prokaryotes [4]. The total size of introns is larger than that of exons and the density of introns is greater than that of exons in genomes, which imposed a huge energetic burden on the cell [5,6]. Introns were once thought to be 'selfish DNAs' or even 'junk DNAs' that consumed large amounts of energy but contributed nothing to the protein synthesis [7]. However, in the past half a century, the biological functions of introns were increasingly verified, including regulating alternative splicing (AS) [8], enhancing gene expression [9–11], controlling mRNA transport and chromatin assembly [12,13], affecting nonsense-mediated decay (NMD) [14,15], introducing new genes [16], generating functional noncoding RNA sequences and so on [4,17,18].

The introns that are removed from protein-coding genes by spliceosomes are the most common and widely studied ones [19]. Introns in the coding region (CDS) of protein-coding genes have been proved to play roles affecting almost every step of gene expression [20,21]. Moreover, the intron-mediated enhancement (IME) effect on gene expression has been

identified in many plants and fungi [22–26]. However, compared to the CDS introns (CIs), introns located in the untranslated regions (UTRs) of protein-coding genes were much less investigated or even ignored during analysis [27,28]. Like the CIs, introns in the 5′UTRs (5UIs) and 3′UTRs (3UIs) were also recognized to function in gene expression regulation [29–32]. The 5UIs are very close to the start codons of protein-coding genes, so they were thought to function during transcription initiation. Moreover, many 5UIs serve as repositories of *cis*-elements [28], indicating that they participate in the gene transcription regulation [6,33,34]. David-Assael et al. [35] found that the 416 bp 5UI of the arabidopsis *AtMHX* gene considerably enhanced gene expression by about 86 times. Akua and Shaul [36] further discovered that the extra sequence of the *AtMHX* 5UI was necessary for the separation of different functional intronic elements. Cenik et al. [37] reported that the presence/absence of 5UIs strongly correlated with sequence features and that the 5UIs could dictate the mRNA export pathway used for the gene, i.e., mRNAs with 5UIs are generally exported through the canonical transcription export (TREX) pathway, whereas those without 5UIs are exported through an alternative mRNA export (ALREX) pathway. It was also shown that 5UIs and 3UIs play significant roles in regulating the NMD sensitivity of transcripts [15]. Unlike 5UIs, 3UIs were usually thought to function in downregulating gene expression levels [38–40], and transcripts containing 3UIs were once generally considered nonfunctional [38,41,42]. However, nowadays, the roles of 3UIs in modulating gene expression have also been verified, [38] and many transcripts containing 3UIs have been proved to be functional [43,44]. Moreover, many noncoding RNAs were identified to be preferentially located within intron regions [17,45,46], indicating that UIs might harbor many ncRNAs [4]. Taken together, it can be concluded that UIs are obviously not junk but functional DNAs.

Given the importance of UIs in gene expression regulation, genome-wide identification and characterization of UIs have been performed in some plants, such as arabidopsis [47,48], sweet orange [28] and *Atalantia buxifolia* [49]. Citrus is one of the most important and most widely cultivated fruit crops in the world due to its global availability and popularity. The *Citrus* genus is comprises some of the most widely cultivated fruit crops, and the diversity and evolution of citrus have been well addressed at the species level through data generated by whole-genome sequencing [50,51]. These well-annotated genome data provide an opportunity to identify, characterize and compare UIs among citrus species. In our previous study, we identified and characterized the UIs in sweet orange genes at a genome-wide level [28]. To explore the UIs and to show their differences among different citrus species, we computationally identified the 5UIs and 3UIs from *C. hindsii*, *C. maxima*, *C. reticulata*, *C. cavaleriei* and *C. medica* based on the citrus genome data and compared their size and nucleotide distribution characteristics with *C. sinensis* UIs [28]. Moreover, UIs commonly identified in these six citrus species were further focused on, and the transcripts containing them were subjected to enrichment analysis to show their possible roles.

2. Materials and Methods

2.1. Genome Data Preparation

The genome data files of six citrus species, including *C. hindsii*, *C. maxima*, *C. reticulata*, *C. cavaleriei*, *C. medica* and *C. sinensis*, were downloaded from the orange annotation project (http://citrus.hzau.edu.cn/orange/download/index.php/, accessed on 28 February 2020).

2.2. Intron Extraction and UI Identification

Based on the genome annotation data, the CDS introns (CIs), 5′UTR introns (5UIs) and 3′UTR introns (3UIs) were separately extracted from the genomes of the six citrus species. For the identification of UTR introns (UIs), only intron sequences located between two UTR exons that never showed retention in any alternative transcripts were retained for further analysis [28]. The obtained UIs were named according to their corresponding transcript IDs and their chromosomal location information. Information regarding the identified

5UIs and 3UIs from *C. hindsii*, *C. maxima*, *C. reticulata*, *C. cavaleriei* and *C. medica* is listed in Supplemental Tables S1–S10, respectively.

2.3. UI Characterization and Identification of Common UIs among the Six Citrus Species

By using the method described by Chung et al. [47] and Shi et al. [28], the identified 5UIs and 3UIs of *C. hindsii, C. maxima, C. reticulata, C. cavaleriei* and *C. medica*, together with *C. sinensis* 5UIs and 3UIs [28], were subjected to density, length distribution and position distribution relative to the start and stop ends of their corresponding UTRs, and splice site pair analysis.

For the identification of the common UIs that existed in all the six citrus species, all the UI sequences were firstly subjected to cluster analysis using CD-HIT. Common UIs were identified under the following criteria: sequence similarity $\geq$ 90%, length difference $\leq$ 10% and coverage ratio $\geq$ 90% [52].

Sequences of the identified common UIs were used for phylogenetic analysis to show the UI conservation among these citrus species. Briefly, tandemly linked common 5UI and/or 3UI sequences were firstly aligned using MAFFT [53]. Then, a phylogenetic tree was constructed by the neighbor-joining method using Mega 7 with 1000 bootstrap replicates [54]. Information pertaining to common 3UIs and 5UIs is provided in Supplemental Tables S11 and S12, respectively.

2.4. Annotation and Pathway Enrichment Analysis of Common UI-Ts

The *C. sinensis* protein sequences downloaded from the orange annotation project were submitted to Mercator v.3.6 (https://plabipd.de/portal/mercator-sequence-annotation, accessed on 3 March 2020) to obtain the mapping files used for the MapMan annotation analysis of the transcripts containing common 5UIs or/and 3UIs [49,55]. After removing the repeated sequences, transcripts containing common UIs (UI-Ts) were subjected to MapMan annotation and PageMan enrichment analysis under default parameters. The obtained MapMan annotation results of the transcripts containing common UIs are shown in Supplemental Table S13.

3. Results

3.1. Identification of Introns in CDSs, 5′UTRs and 3′UTRs in Six Citrus Species

Totally, we extracted 23,394, 32,257, 30,123, 28,833, 32,067 and 32,579 complete CDSs; 16,916, 17,275, 13,330, 13,784, 14,336 and 15,240 5′UTRs and 17,408, 17,160, 14,144, 14,127, 14,116 and 15,502 3′UTRs from the genome data of *C. sinensis* [28], *C. hindsii*, *C. maxima*, *C. reticulata*, *C. cavaleriei* and *C. medica*, respectively (Table 1). Among these sequences, there were, respectively, 617 5′UTRs (corresponding to 965 5UIs), 17,897 CDSs and 469 3′UTRs (corresponding to 745 3UIs) in *C. sinensis*; 567 5′UTRs (corresponding to 935 5UIs), 24,326 CDSs and 482 3′UTRs (corresponding to 854 3UIs) in *C. hindsii;* 408 5′UTRs (corresponding to 604 5UIs), 22,186 CDSs and 244 3′UTRs (corresponding to 385 3UIs) in *C. maxima;* 430 5′UTRs (corresponding to 675 5UIs), 21,372 CDSs and 274 3′UTRs (corresponding to 446 3UIs) in *C. reticulata*; 450 5′UTRs (corresponding to 683 5UIs), 24,499 CDSs and 280 3′UTRs (corresponding to 450 3UIs) in *C. cavaleriei*; and 444 5′UTRs (corresponding to 682 5UIs), 24,411 CDSs and 349 3′UTRs (corresponding to 572 3UIs) in *C. medica* containing introns (Table 1).

Table 1. Statistics information of 5′ UTR, CDS and 3′ UTR in six citrus species. CDS: coding sequence; UI: UTR intron; *: cited from reference [28].

Species	Position	No. of Sequences	Sequences with Introns	Total Bases (Genomic)	Intron/Sequence	No. of UIs	No. of Introns/Nucleotides (mRNA)
C. sinensis *	5′UTR	16,916	617	6.80×10^6	0.06	965	1.42×10^{-4}
	CDS	23,394	17,897	3.77×10^7	4.81	-	2.98×10^{-3}
	3′UTR	17,408	469	1.17×10^7	0.04	745	6.36×10^{-5}
C. hindsii	5′UTR	17,275	567	1.32×10^7	0.05	935	7.07×10^{-5}
	CDS	32,257	24,326	4.18×10^7	4.16	-	3.21×10^{-3}
	3′UTR	17,160	482	2.61×10^7	0.05	854	3.27×10^{-5}
C. maxima	5′UTR	13,330	408	4.12×10^6	0.05	604	1.47×10^{-4}
	CDS	30,123	22,186	3.42×10^7	3.68	-	3.25×10^{-3}
	3′UTR	14,144	244	8.94×10^6	0.03	385	4.31×10^{-5}
C. reticulata	5′UTR	13,784	430	5.32×10^6	0.05	675	1.27×10^{-4}
	CDS	28,833	21,372	3.48×10^7	3.87	-	3.20×10^{-3}
	3′UTR	14,127	274	1.08×10^7	0.03	446	4.14×10^{-5}
C. cavaleriei	5′UTR	14,336	450	5.45×10^6	0.05	683	1.25×10^{-4}
	CDS	32,067	24,499	3.45×10^7	3.74	-	3.47×10^{-3}
	3′UTR	14,116	280	9.74×10^6	0.03	450	4.62×10^{-5}
C. medica	5′UTR	15,240	444	6.48×10^6	0.04	682	1.05×10^{-4}
	CDS	32,579	24,411	3.56×10^7	3.72	-	3.40×10^{-3}
	3′UTR	15,502	349	1.40×10^7	0.04	572	4.07×10^{-5}

On average, there were 3.68~4.81 introns for each CDS sequence in the six citrus species (the average intron number for CDSs in *C. sinensis* was the highest, followed by *C. hindsii*, *C. reticulata*, *C. cavaleriei*, *C. media* and *C. maxima*), which is considerably higher than those in both 5′UTRs (0.04~0.06 introns for each 5′UTR) and 3′UTRs (0.03~0.05 introns for each 3′UTR). Results showed that the UI numbers varied in different citrus species (Table 1). The total UI number and 3UI number annotated in *C. hindsii* genome were both the largest, which were 1789 and 854, respectively. The total UI number and 3UI number of *C. sinensis* both ranked the second, which were 1710 and 745, respectively. Its 5UI number was also the largest. *C. maxima* had the least 5UI and 3UI numbers (604 5UIs and 385 3UIs, 989 in total). The 3UI number of *C. hindsii* was about 2.22-fold of that of *C. maxima*. However, the 5UI density of *C. maxima* was the highest, which was 2.08-fold of that of *C. hindsii*. Moreover, the 3UI density of *C. sinensis* was the highest, which was about 1.94-fold that of *C. maxima*. These results show that both the abundance and the density of UIs varied a lot among different citrus species.

3.2. Size and Position Distributions of UIs in Six Citrus Species

Intron size distribution analysis revealed that, compared with the CIs, the UIs of all the six citrus species were less conserved in size. CIs with lengths ranging from 100 to 300 bp peaked obviously; the UIs with lengths in this range also peaked but not as obviously as the CIs (Figure 1). Moreover, the relative frequency of 3UIs within this range was higher than that of 5UIs in *C. sinensis*, *C. medica* and *C. cavaleriei* (Figure 1).

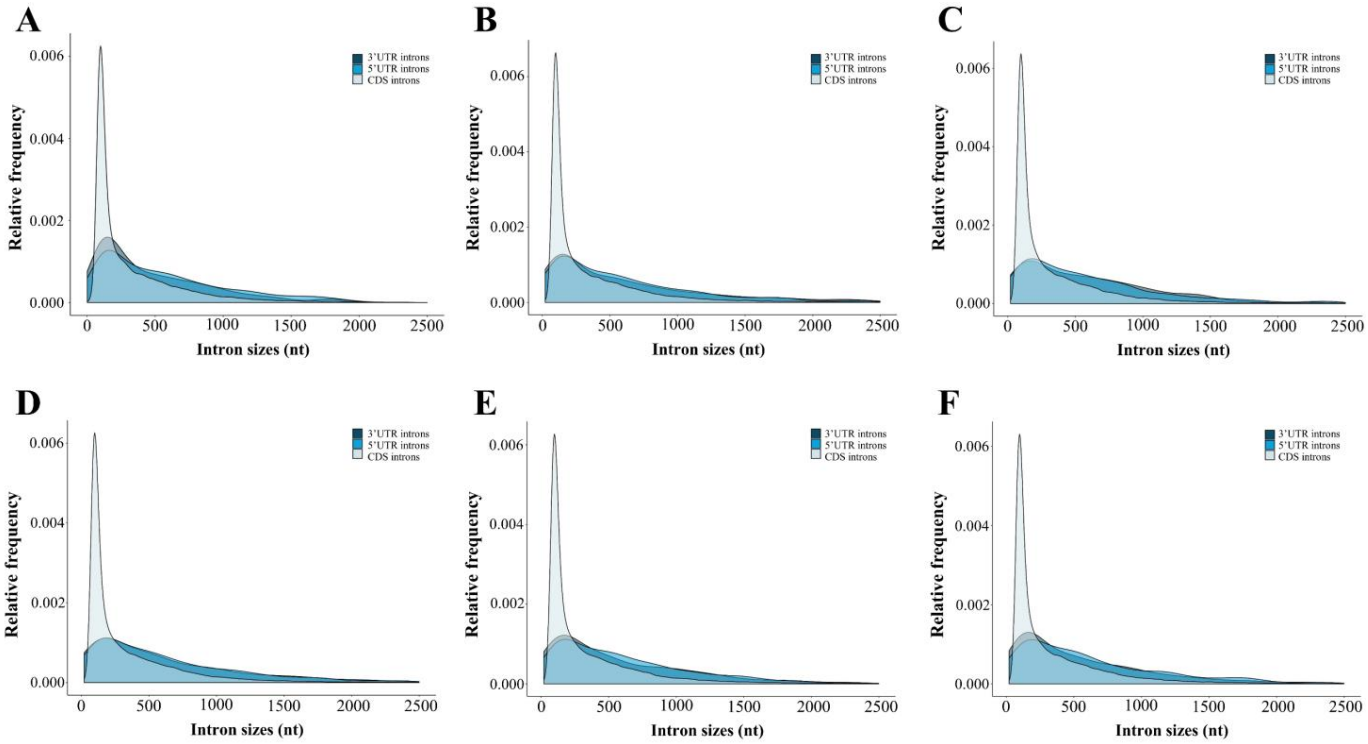

Figure 1. Length distributions of 5UI, 3UI and CIs in six citrus species. The CDS introns of all the six citrus species were found to be much more conserved in size than both the 5′UTR introns and 3′UTR introns, peaking in the range of 100~300 bp. (**A**–**F**) represents the intron length distribution result for *C. sinensis* [28], *C. hindsii*, *C. maxima*, *C. reticulata*, *C. cavaleriei* and *C. medica*, respectively. The X-axis and Y-axis represent the intron sizes and the relative frequency of introns.

The result of the intron position distribution analysis showed that, consistent with *C. sinensis* [28], both 5UIs and 3UIs in the other five citrus species were also preferentially located at the ends of the UTRs, and the proximity of 5UIs to the stop end of 5′UTR and the proximity of 3UIs to the start end of 3′UTR were both obvious.

3.3. Intron Size Comparisons among Six Citrus Species

The sizes of the 5′UTR, CDS and 3′UTR introns were calculated and compared among the six citrus species (Table 2). Results showed that the mean 5UI, CI and 3UI lengths of *C. sinensis* were all the lowest. The mean 5UI average length of *C. reticulata* was the highest (1594.74 nucleotides), which was about 2.71-fold that of *C. sinensis*. The mean CI length of *C. maxima* was the highest (468.96 nucleotides), which was about 1.37-fold that of *C. sinensis*. The mean 3UI length of *C. hindsii* was the highest (930.25 nucleotides), which was about 1.65-fold that of *C. sinensis*. Moreover, we also found that the mean lengths of 5UIs and 3UIs were both higher than that of CIs, indicating that UTRs are inclined to possess longer introns. Consistently, the mean, median, lower quartile and upper quartile lengths of 5UIs and 3UIs were all higher than those of CIs. For instance, the mean length of 5UIs was about 3.40-fold that of its CIs in *C. maxima*.

Table 2. Summary statistics for intron lengths in 5′UTR, 3′UTR and CDS in six citrus species. CI: CDS intron; LQ: lower quartile; UQ: upper quartile. *: cited from reference [28].

Species	Intron Type	Mean	Median	LQ	UQ
C. sinensis *	5UI	587.50	450	168	836
	CI	343.20	171	102	441
	3UI	563.5	335	139	730
C. hindsii	5UI	818.16	472	165	925
	CI	463.07	172	102	457
	3UI	930.25	459	141	981.25
C. maxima	5UI	665.42	475.5	180.75	851.5
	CI	468.96	174	102	449
	3UI	751.44	508	159	960
C. reticulata	5UI	1594.74	494	186.5	977
	CI	461.40	176	102	459
	3UI	858.35	473	154.5	955.5
C. cavaleriei	5UI	654.86	506	197.5	881.5
	CI	433.85	177	102	463
	3UI	691.95	435	153.25	928.25
C. medica	5UI	682.38	485	198.25	914.75
	CI	403.59	177	102	462
	3UI	742.33	428.5	150	861

3.4. Splice Site Conservation Analysis of UIs in Six Citrus Species

Splice sites, including the 5′ donor sites and 3′ receptor sites, are required for intron removal [56]. By analyzing the nucleotide preferences surrounding the UI splice junctions, we found that 'GT-AG' type splice site (SS) pairs constituted the largest part, accounting for 95.62~97.18% of 3UIs and 97.21~97.92% of 5UIs in the six citrus species (Table 3). Furthermore, the 'GC-AG' type SS pairs ranked the second for both 5UIs and 3UIs of all the six citrus species except the *C. medica* 3UIs (ranked the third). The 5UIs and 3UIs in *C. hindsii*, *C. maxima*, *C. reticulata* and *C. cavaleriei* had three types of SS pairs, i.e., 'GT-AG', 'GC-AG' and 'AT-AC' types. Moreover, the 'AT-AC' SS pairs ranked the third for *C. sinensis* 5UIs and 3UIs and *C. medica* 5UIs but ranked the second in *C. medica* 3UIs.

Table 3. Information for the splice site (SS) pair types of UIs in six citrus species.

Citrus Species	UI Type	SS Pair Type	No.	Percentage
C. sinensis	3UIs	GT-AG	724	97.18%
		GC-AG	16	2.14%
		AT-AC	2	0.26%
		GT-TG	1	0.13%
		TA-AG	1	0.13%
		TT-AG	1	0.13%
	5UIs	GT-AG	945	97.92%
		GC-AG	14	1.45%
		AT-AC	2	0.20%
		CT-AC	2	0.20%
		GT-GG	1	0.10%
		TG-AG	1	0.10%
C. hindsii	3UIs	GT-AG	823	96.37%
		GC-AG	16	1.87%
		AT-AC	15	1.75%
	5UIs	GT-AG	909	97.21%
		GC-AG	20	2.13%
		AT-AC	6	0.64%

Table 3. *Cont.*

Citrus Species	UI Type	SS Pair Type	No.	Percentage
C. maxima	3UIs	GT-AG	371	96.36%
		GC-AG	10	2.59%
		AT-AC	4	1.03%
	5UIs	GT-AG	589	97.51%
		GC-AG	9	1.49%
		AT-AC	6	0.99%
C. reticulata	3UIs	GT-AG	429	96.18%
		GC-AG	15	3.36%
		AT-AC	2	0.44%
	5UIs	GT-AG	657	97.33%
		GC-AG	13	1.92%
		AT-AC	5	0.74%
C. cavaleriei	3UIs	GT-AG	434	96.44%
		GC-AG	13	2.88%
		AT-AC	3	0.66%
	5UIs	GT-AG	666	97.51%
		GC-AG	16	2.34%
		AT-AC	1	0.14%
C. medica	3UIs	GT-AG	547	95.62%
		AT-AC	12	2.09%
		GC-AG	9	1.57%
		CA-AG	1	0.17%
		TA-CA	1	0.17%
		TT-AT	1	0.17%
		TT-CT	1	0.17%
	5UIs	GT-AG	665	97.50%
		GC-AG	9	1.31%
		AT-AC	3	0.43%
		AT-GA	1	0.14%
		GG-GC	1	0.14%
		TA-CT	1	0.14%
		TT-AT	1	0.14%
		TT-TA	1	0.14%

For *C. sinensis*, there were also 'GT-TG', 'TA-AG' and 'TT-AG' SS pair types for 3UIs and 'CT-AC', 'GT-GG' and 'TG-AG' SS pair types for 5UIs. In total, nine types of SS pairs were identified in UIs of *C. sinensis* [28]. For *C. medica*, there were also 'CA-AG', 'TA-CA', 'TT-AT' and 'TT-CT' SS pair types for 3UIs and 'AT-GA', 'GG-GC', 'TA-CT', 'TT-AT' and 'TT-TA' SS pair types for 5UIs. In all, 11 types of SS pairs were identified in the UIs of *C. medica*.

3.5. Coexistence Analysis of UIs in All the Six Citrus Species

Through coexistence analysis, we found that most 5UIs and 3UIs were species specific (Figure 2A,B). There were 429 (44.46%), 377 (40.32%), 232 (34.02%), 208 (30.45%), 186 (27.56%) and 128 (21.19%) species-specific 5UIs for *C. sinensis*, *C. hindsii*, *C. medica*, *C. cavaleriei*, *C. reticulata* and *C. maxima*, respectively. In addition, 457 (53.51%), 414 (55.57%), 266 (46.50%), 172 (38.22%), 148 (33.18%) and 148 (28.05%) species-specific 3UIs were identified in *C. hindsii*, *C. sinensis*, *C. medica*, *C. cavaleriei*, *C. reticulata*, and *C. maxima*, respectively. These results indicated that the UI sequences varied a lot among different citrus species. Additionally, *C. sinensis* shared, respectively, 246, 276, 301, 250 and 237 5UIs and 146, 142, 159, 122 and 108 3UIs with *C. hindsii, C. maxima, C. reticulata, C. cavaleriei* and *C. medica.*

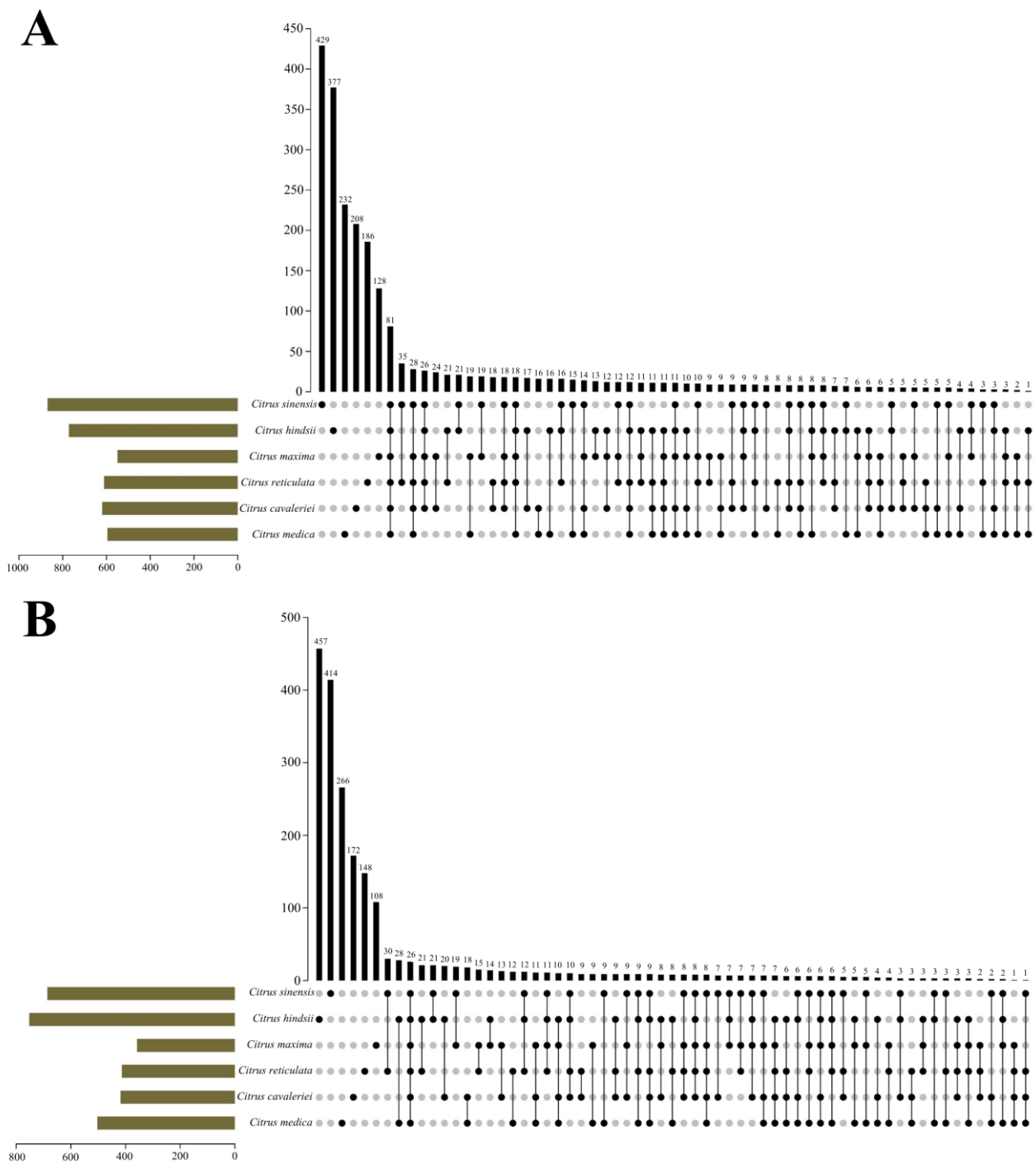

Figure 2. Coexistence analysis result of 5UIs and 3UIs identified in six citrus species. (**A**) 5UIs identified in six citrus species. (**B**) 3UIs identified in six citrus species. The vertical columns represent number of common UIs among different citrus species, and the transverse columns represent the total number of identified UIs in each citrus species.

We also identified 81 5UIs and 26 3UIs coexisting in all the six citrus species (Figure 2). Phylogenic analysis results based on the tandem common 5UIs or/and 3UIs showed that the UIs' evolutionary relationship between *C. sinensis* and *C. hindsii* was the closest, followed by that between *C. maxima* and *C. cavaleriei* (Figure 3). The common 5UIs and 3UIs between *C. sinensis* and *C. maxima* and *C. reticulata* ranked in the top two. As sweet orange originates from *C. maxima* and *C. reticulata* [50,51], the common 5UI and 3UI sequences

shared by them, especially the 5UIs, might have potential to be used in the study of the origin and evolution of citrus.

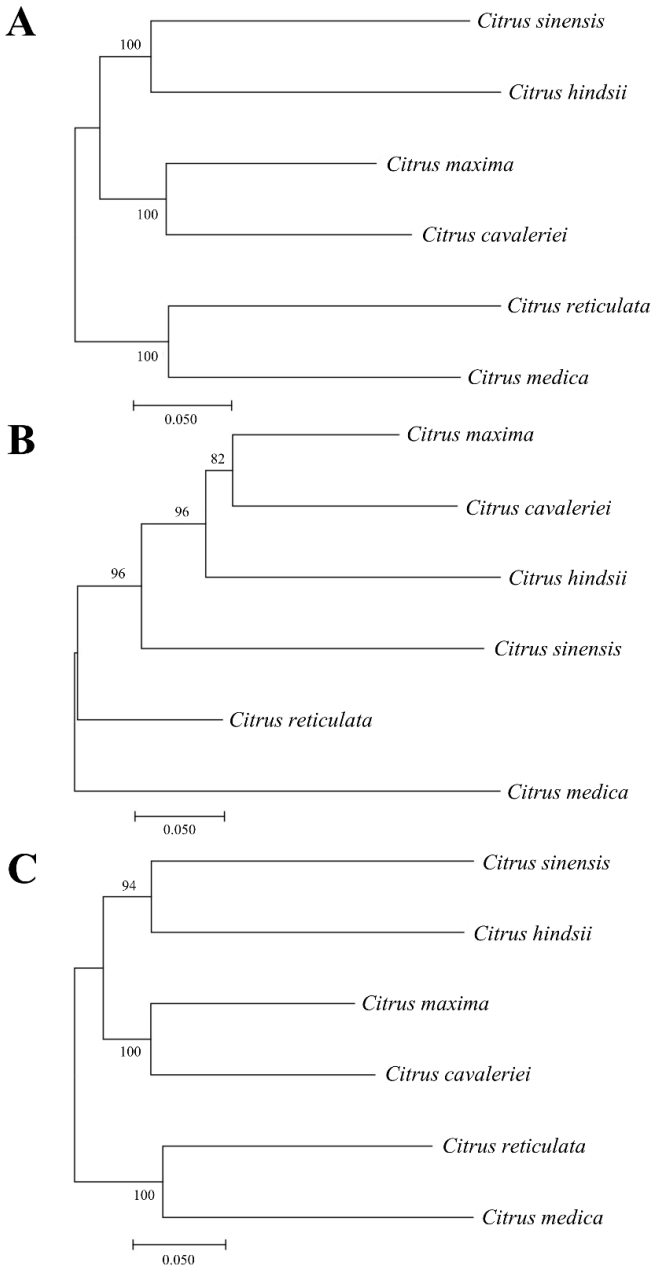

Figure 3. Phylogenic tree constructed based on the common 5UI and 3UI sequences. (**A**–**C**) Phylogenic tree constructed based on the tandom 81 common 5UI sequences, 26 common 3UI sequences and all the common UIs, respectively.

3.6. Annotation and Enrichment Analysis of the Common UI-Containing Transcripts (UI-Ts)

After removing the repeated transcripts, 99 transcripts containing common UIs (UI-Ts) were identified and subjected to MapMan annotation based on the *C. sinensis* mapping file generated using Mercator v.3.6. These UI-Ts could be mapped to 107 data points of the *C. sinensis* mapping file. Annotation results showed that these UI-Ts were involved in 'RNA biosynthesis' (14), 'RNA processing' (6), 'protein modification' (5), 'protein homeostasis' (5), 'protein biosynthesis' (4), 'vesicle trafficking' (4), 'phytohormone action' (3), 'cell wall organization' (3), 'chromatin organisation' (2), 'coenzyme metabolism' (2), 'cell division' (2), 'protein translocation' (2), 'carbohydrate metabolism' (1), 'solute transport' (1), 'external

stimuli response' (1), 'multi-process regulation' (1) and 'redox homeostasis' (1) (Table 4). Moreover, 49 UI-Ts (including 26 annotated and 23 unannotated ones) were categorized as 'not assigned'.

Table 4. Annotation results for the common UI-containing transcripts identified in this study. 5UI-T: transcripts containing common 5UI; 3UI-T: transcripts containing common 3UI; 5&3UI: transcripts containing both common 5UI and 3UI. MapMan annotation was performed using the *C. sinensis* gene mapping files.

BINcode	BinName	Gene ID	Description	Type
3.13.3.1	Carbohydrate metabolism. nucleotide sugar biosynthesis. UDP-D-glucuronic acid biosynthesis. UDP-D-glucose 6-dehydrogenase	cs6g22050.1	UDP-D-glucose 6-dehydrogenase	5UI-T
7.3.1	Coenzyme metabolism. S-adenosyl methionine (SAM) cycle. S-adenosyl methionine synthetase (MAT)	cs6g01310.2	S-adenosyl methionine synthetase	5UI-T
		cs9g01410.1	S-adenosyl methionine synthetase	5UI-T
10.4.2.4	Redox homeostasis. thiol-based redox regulation. peroxiredoxin activities. type-2 peroxiredoxin (PrxII)	cs6g15550.2	type-2 peroxiredoxin (PrxII)	3UI-T
11.1.2.1.3	Phytohormone action. abscisic acid. perception and signalling. receptor activities. regulatory protein (EAR1)	cs8g14510.2	regulatory protein (EAR1) of abscisic acid signaling	5UI-T
11.2.4.2	Phytohormone action. auxin. transport. auxin efflux transporter (PILS)	cs2g13710.1	auxin efflux transporter (PILS)	5UI-T
11.10.2.4.2	Phytohormone action. signalling peptides. CRP (cysteine-rich-peptide) category. RALF/RALFL-peptide activity. RALF-peptide receptor (CrRLK1L)	cs6g10250.2	RALF-peptide receptor (CrRLK1L)	5UI-T
12.2.2.1	Chromatin organisation. histone chaperone activities. FACT histone chaperone complex. component SPT16	cs6g01200.1	component SPT16 of FACT histone chaperone complex	5UI-T
12.3.1.1.7	Chromatin organisation. post-translational histone modification. histone methylation. lysine methylation. class-VI histone methyltransferase (SMYD)	cs8g12080.2	class-VI histone methyltransferase (SMYD)	3UI-T
13.2.1.2.5	Cell division. cell cycle organisation. cell cycle control. CYCLIN-dependent protein kinase complex. catalytic component CDKE	cs5g01900.2	protein kinase (CDKE/CDK8)	5UI-T
13.4.2.3.1	Cell division. cytokinesis. cell-plate formation. SNARE cell-plate vesicle fusion complex. Qa-SNARE component KNOLLE	cs5g26090.1	SYP1-group Qa-type SNARE component	5UI-T
15.1.4.6	RNA biosynthesis. DNA-dependent RNA polymerase complexes. RNA polymerase IV complex. subunit NRPD7	cs9g11150.1	subunit NRPD7 of RNA polymerase IV complex	5UI-T
15.1.5.6	RNA biosynthesis. DNA-dependent RNA polymerase complexes. RNA polymerase V complex. subunit NRPE7	cs9g11150.1	subunit NRPE7 of RNA polymerase V complex	5UI-T
15.3.4.2.6	RNA biosynthesis. RNA polymerase II-dependent transcription. transcription co-activation. TFIId complex. component TAF8	cs6g22280.1	component TAF8 of TFIId basal transcription regulation complex	5UI-T
15.3.4.3.3.2	RNA biosynthesis. RNA polymerase II-dependent transcription. transcription co-activation. SAGA complex. SPT recruitment module. component ADA1	cs1g21850.2	component ADA1 of SAGA transcription co-activator complex	5UI-T
		cs7g07280.1	component ADA1 of SAGA transcription co-activator complex	5UI-T
15.3.4.4.4.3	RNA biosynthesis. RNA polymerase II-dependent transcription. transcription co-activation. MEDIATOR complex. regulatory kinase module. component CDK8	cs5g01900.2	protein kinase (CDKE/CDK8)	5UI-T
15.5.2.2	RNA biosynthesis. transcriptional regulation. MYB transcription factor superfamily. transcription factor (MYB-related)	cs1g24225.1	transcription factor (MYB-related)	3UI-T

Table 4. *Cont.*

BINcode	BinName	Gene ID	Description	Type
15.5.12	RNA biosynthesis. transcriptional regulation. transcription factor (GRAS)	cs4g12130.1	transcription factor (GRAS)	5UI-T
		cs7g02550.1	transcription factor (GRAS)	5UI-T
15.5.20	RNA biosynthesis. transcriptional regulation. transcription factor (Trihelix)	cs4g16730.1	transcription factor (Trihelix)	3UI-T
15.5.30	RNA biosynthesis. transcriptional regulation. transcription factor (bHLH)	cs4g02590.1	transcription factor (bHLH)	5UI-T
		cs5g30170.2	transcription factor (bHLH)	5UI-T
15.5.32	RNA biosynthesis. transcriptional regulation. transcription factor (BBR/BPC)	orange1.1t01638.1	transcription factor (BBR/BPC)	5UI-T
15.6.2.2	RNA biosynthesis. organelle machinery. transcriptional regulation. transcription factor (mTERF)	cs5g31960.1	transcription factor (mTERF)	5UI-T
		cs8g01080.1	transcription factor (mTERF)	5UI-T
16.1.1.2.8	RNA processing. pre-RNA splicing. U2-type-intron-specific major spliceosome. U2 small nuclear ribonucleoprotein particle (snRNP). pre-mRNA splicing factor (SF1)	cs9g15030.1	pre-mRNA splicing factor (SF1)	3UI-T
16.4.9.4	RNA processing. RNA homeostasis. mRNA stress granule formation. regulatory protein (UBA1/2) of UBP1 activity	cs6g16060.1	regulatory protein (UBA1/2) of UBP1 activity	3UI-T
		cs7g25330.3	regulatory protein (UBA1/2) of UBP1 activity	3UI-T
16.5.2.3.3	RNA processing. mRNA silencing. miRNA pathway. miRNA degradation. regulatory protein (HWS)	orange1.1t00443.1	regulatory protein (HWS) of miRNA degradation	5UI-T
16.6.1.1.11	RNA processing. organelle machinery. pre-RNA splicing. plastidial RNA splicing. splicing factor (mTERF4)	cs5g31960.1	mTERF4 plastidial RNA splicing factor	5UI-T
16.6.2.2.4.7	RNA processing. organelle machinery. RNA modification. C-to-U RNA editing. PPR-type RNA editing factor activities. RNA editing factor (MEF9)	cs4g13530.1	RNA editing factor (MEF9)	3UI-T
17.1.2.2.2.5	Protein biosynthesis. ribosome biogenesis. large ribosomal subunit (LSU). LSU processome. pre-60S ribosomal subunit nuclear export. export factor (NMD3)	cs6g17980.1	pre-60S subunit nuclear export factor (NMD3)	5UI-T
17.1.3.2.1.3.1	Protein biosynthesis. ribosome biogenesis. small ribosomal subunit (SSU). SSU processome. pre-40S ribosomal subunit nuclear assembly. UtpB module. assembly factor (UTP18)	cs5g30340.1	SSU processome assembly factor (UTP18)	5UI-T
17.3.1.1.2	Protein biosynthesis. translation initiation. Pre-Initiation Complex (PIC) module. eIF1 PIC assembly factor activity. assembly factor (eIF1A)	cs2g20280.1	assembly factor (eIF1A) of eIF1	5UI-T
17.3.1.2.2	Protein biosynthesis. translation initiation. Pre-Initiation Complex (PIC) module. eIF2 Met-tRNA binding factor activity. activating factor (eIF5) of eIF2-GTP hydrolysis	cs3g18950.1	activating factor (eIF5) of eIF2-GTP hydrolysis	5UI-T
18.2.4	Protein modification. acetylation. N-terminal acetylase (NatD)	cs1g20790.1	N-terminal acetylase (NatD)	5UI-T
18.3.4.1.1.2	Protein modification. lipidation. glycophosphatidylinositol (GPI) anchor addition. GPI pre-assembly. GPI N-acetylglucosamine transferase complex. component PIG-C	cs2g11690.2	component PIG-C of GPI N-acetylglucosamine transferase complex	5UI-T
18.4.1.16	Protein modification. phosphorylation. TKL protein kinase superfamily. protein kinase (CrlRLK1)	cs6g10250.2	protein kinase (CrlRLK1)	5UI-T
18.4.3.1.5	Protein modification. phosphorylation. CMGC protein kinase superfamily. CDK protein kinase families. protein kinase (CDKE/CDK8)	cs5g01900.2	protein kinase (CDKE/CDK8)	5UI-T
18.13.1	Protein modification. protein folding. protein folding catalyst (Cyclophilin)	cs8g12840.1	protein folding catalyst	5UI-T

Table 4. *Cont.*

BINcode	BinName	Gene ID	Description	Type
19.1.2.3	Protein homeostasis. protein quality control. ribosome-associated chaperone activities. co-chaperone (ZRF)	cs6g08770.1	Hsp40-chaperone ZRF ribosome-associated chaperone complex	5&3UI-T
19.2.1.3.1.2	Protein homeostasis. ubiquitin-proteasome system. N-degron pathways. Pro/N-degron pathway. GID ubiquitination complex. ubiquitin ligase component GID2	cs8g03080.1	ubiquitin ligase component GID2 of GID ubiquitination complex	5UI-T
19.2.2.1.4.3.3.2	Protein homeostasis. ubiquitin-proteasome system. ubiquitin-fold protein conjugation. ubiquitin conjugation (ubiquitylation). ubiquitin-ligase E3 activities. RING-domain E3 ligase activities. RING-H2-class ligase activities. BTL-subclass ligase	cs6g16300.2	RING-H2-class E3 BTL-subclass ubiquitin ligase	5UI-T
19.2.2.8.1.4.3	Protein homeostasis. ubiquitin-proteasome system. ubiquitin-fold protein conjugation. Cullin-based ubiquitylation complexes. SKP1-CUL1-FBX (SCF) E3 ubiquitin ligase complexes. F-BOX substrate adaptor activities. substrate adaptor (FBX)	orange1.1t00443.1	substrate adaptor FBX of SCF E3 ubiquitin ligase complex	5UI-T
19.2.5.2.2.3	Protein homeostasis. ubiquitin-proteasome system. 26S proteasome. 19S regulatory particle. lid subcomplex. regulatory component RPN6	cs4g04180.1	regulatory component RPN6 of 26S proteasome	5UI-T
21.4.1.1.3	Cell wall organisation. cell wall proteins. hydroxyproline-rich glycoprotein activities. arabinogalactan-protein activities. Fasciclin-type arabinogalactan protein (FLA)	cs2g20030.1	arabinogalactan protein (Fasciclin)	5UI-T
		cs2g20030.2	arabinogalactan protein (Fasciclin)	5UI-T
		cs2g20030.3	arabinogalactan protein (Fasciclin)	5UI-T
22.1.1.1.1	Vesicle trafficking. anterograde trafficking. Coat protein II (COPII) coatomer machinery. coat protein complex. scaffolding component Sec13	cs2g28780.1	scaffolding component Sec13 of coat protein complex	5UI-T
22.3.1.1.2	Vesicle trafficking. endocytic trafficking. ESCRT-mediated sorting. ESCRT-I complex. component VPS28	cs2g06750.1	component VPS28 of ESCRT-I complex	5UI-T
22.5.2.4.3.2	Vesicle trafficking. multi-pathway trafficking regulation. vesicle tethering. RAB-GTPase membrane association. RAB-GDI displacement factor (GDF) activities. B-G-class Rab-GDF protein	cs7g30390.2	B-G-class Rab-GDF protein	3UI-T
22.5.3.1.1.1	Vesicle trafficking. multi-pathway trafficking regulation. target membrane fusion. SNARE membrane fusion complexes. Qa-type SNARE components. SYP1-group component	cs5g26090.1	SYP1-group Qa-type SNARE component	5UI-T
23.1.2.2	Protein translocation. chloroplast. outer envelope TOC translocation system. receptor GTPase (Toc90/120/132/159)	cs8g12230.1	component Toc90/120/132/159 of outer envelope TOC translocation system	5UI-T
23.5.2.3.2	Protein translocation. nucleus. nucleocytoplasmic transport. RAN GTPase cycle. Ran-activating protein (Ran-GAP)	cs9g06440.1	Ran-activating protein of nucleocytoplasmic transport	5UI-T
24.2.5.2.2	Solute transport. carrier-mediated transport. BART superfamily. AEC family. auxin transporter (PILS)	cs2g13710.1	auxin transporter (PILS)	5UI-T
26.9.2.2.4	External stimuli response. pathogen. effector-triggered immunity (ETI) network. RIN4-RPM1 immune signalling. regulatory protein (GCN4) of RIN4 activity	cs1g13980.1	regulatory protein (GCN4) of RIN4 activity	5UI-T
27.6.1.4.5	Multi-process regulation. phosphatidylinositol and inositol phosphate system. biosynthesis. phosphatidylinositol kinase activities. phosphatidylinositol 4-kinase (PI4K-gamma)	cs7g12040.1	phosphatidylinositol 4-kinase (PI4K-gamma)	5UI-T
35.1	not assigned. annotated	cs1g11100.1	probable CCR4-associated factor 1 homolog 6	5UI-T
		cs1g11700.2	F-box/LRR-repeat protein 14	5UI-T
		cs2g11780.1	pentatricopeptide repeat-containing protein	3UI-T

Table 4. *Cont.*

BINcode	BinName	Gene ID	Description	Type
		cs2g18610.1	transmembrane 9 superfamily member 12	5UI-T
		cs3g20090.2	pentatricopeptide repeat-containing protein	5UI-T
		cs4g03945.1	pentatricopeptide repeat-containing protein	3UI-T
		cs5g03910.1	putative pentatricopeptide repeat-containing protein	5&3UI-T
		cs5g09740.1	zinc finger A20 and AN1 domain-containing stress-associated protein 1	5UI-T
		cs6g04970.2	serine/threonine-protein kinase ATM	5&3UI-T
		cs6g07760.1	pentatricopeptide repeat-containing protein	3UI-T
		cs6g08820.1	pentatricopeptide repeat-containing protein	3UI-T
		cs6g08820.2	pentatricopeptide repeat-containing protein	3UI-T
		cs6g15060.2	zinc finger A20 and AN1 domain-containing stress-associated protein 4	5UI-T
		cs6g15600.2	protein PSK SIMULATOR 1	5UI-T
		cs7g02570.1	calmodulin binding protein PICBP	3UI-T
		cs7g04050.2	F-box/Kelch-repeat protein SKIP11	5UI-T
		cs7g04620.1	pentatricopeptide repeat-containing protein	5&3UI-T
		cs7g15390.1	pentatricopeptide repeat-containing protein	3UI-T
		cs7g19080.2	FT-interacting protein 3	5UI-T
		cs8g04770.1	F-box/Kelch-repeat protein	3UI-T
		cs8g12590.2	pentatricopeptide repeat-containing protein	5UI-T
		cs8g16750.1	F-box/Kelch-repeat protein	5UI-T
		cs9g04270.1	chaperone protein dnaJ 49	3UI-T
		cs9g04270.2	chaperone protein dnaJ 49	5UI-T
		orange1.1t00345.2	FT-interacting protein 3	5UI-T
		orange1.1t04379.2	FT-interacting protein 3	5UI-T
35.2	not assigned. not annotated	cs1g26580.1	unknown	5UI-T
		cs1g26840.2	unknown	5UI-T
		cs2g13570.1	unknown	5UI-T
		cs2g19060.1	unknown	5UI-T
		cs2g21340.2	unknown	5UI-T
		cs3g11280.1	unknown	5UI-T
		cs3g16520.2	unknown	5UI-T
		cs4g16820.2	unknown	3UI-T
		cs4g17430.1	unknown	3UI-T

Table 4. *Cont.*

BINcode	BinName	Gene ID	Description	Type
		cs4g18220.2	unknown	5UI-T
		cs5g24290.1	unknown	5UI-T
		cs5g27890.1	unknown	5UI-T
		cs6g13200.1	unknown	5UI-T
		cs7g09380.2	unknown	5UI-T
		cs7g10870.2	unknown	5UI-T
		cs7g24520.1	unknown	5UI-T
		cs8g16190.1	unknown	5UI-T
		cs9g02430.1	unknown	5UI-T
		cs9g11810.2	unknown	5UI-T
		orange1.1t00607.1	unknown	5UI-T
		orange1.1t01667.1	unknown	5UI-T
		orange1.1t02210.1	unknown	3UI-T
		orange1.1t05845.1	unknown	5UI-T

Gene expression includes transcription and translation steps. According to the MapMan annotation results, many UI-Ts were found to be involved in gene-expression-related pathways, such as 'RNA biosynthesis', 'RNA processing', 'protein modification', 'protein homeostasis', 'protein biosynthesis', 'protein translocation', 'vesicle trafficking' and so on. Moreover, nine of these UI-Ts were genes encoding transcription factors.

Introns have been reported to function in controlling chromatin assembly [13]. In this study, genes encoding component SPT16 of the FACT histone chaperone complex (cs6g01200.1) and class-VI histone methyltransferase (SMYD) (cs8g12080.2) were classified into 'chromatin organisation'-related UI-Ts. Many UI-Ts were predicted to be involved in phytohormone metabolism and signaling, such as genes encoding the regulatory protein (EAR1) of abscisic acid signaling (cs8g14510.2), auxin efflux transporter (PILS) (cs2g13710.1), RALF-peptide receptor (CrRLK1L) (cs6g10250.2) and auxin transporter (PILS) (cs2g13710.1). Moreover, a UI-T (cs5g01900.2) encoding protein kinase (CDKE/CDK8) and a UI-T (cs5g26090.1) encoding the SYP1-group Qa-type SNARE component were classified into the 'cell division' pathway.

Among the 'not assigned' UI-Ts, there were 10 UI-Ts (including 7 3UI-Ts, 2 5UI-Ts and 1 5&3UI-Ts) encoding pentatricopeptide repeat-containing proteins, 3 UI-Ts (cs7g19080.2, orange1.1t00345.2 and orange1.1t04379.2) encoding FT-interacting proteins, 3 UI-Ts (cs1g11700.2, cs7g04050.2 and cs8g16750.1) encoding F-box/Kelch-repeat proteins, 2 UI-Ts (cs9g04270.1 and cs9g04270.2) encoding chaperone protein dnaJ 49 and 2 UI-Ts (cs5g09740.1 and cs6g15060.2) encoding zinc finger A20 and AN1 domain-containing stress-associated proteins. Moreover, genes encoding probable CCR4-associated factor 1 homolog 6 (cs1g11100.1) and protein PSK SIMULATOR 1 (cs6g15600.2) were also identified as UI-Ts.

4. Discussion

With the development of the high-throughput sequencing techniques, large-scale transcriptome data were obtained. By aligning the transcriptome data to the genome sequence, the genome data can be well annotated and the intron–exon structure within CDSs and UTRs can be directly determined [4,48]. In this study, based on the well-addressed citrus genome data, we identified and compared the UTR introns from six citrus species.

4.1. Characteristics of UIs in Different Citrus Species

The length, location and density characteristics of UIs differ greatly from that of the introns in the CDSs [28,47–49,57,58]. In our present study, we found that the UI sizes of

the six citrus species were all much less conserved compared with their corresponding CIs. Most CIs ranged from 100 to 300 bp. There were also many UIs within this range but not overwhelmingly so, unlike the CIs. Moreover, for *C. sinensis*, *C. medica* and *C. cavaleriei*, there were more 3UIs within this range compared to 5UIs. We also found that both the abundance and the density of UIs differed a lot among the six citrus species, suggesting that the UIs varied greatly during citrus evolution.

The UIs of the six citrus species tended to be located near the stop end of the 5′UTR or the start end of the 3′UTR, indicating that they preferentially located close to the CDSs. Thus, it was predicted that the position distribution would influence the regulatory roles of introns [28]. Moreover, the average intron number and density in each 5′UTR or 3′UTR were found to be significantly lower than those in the CDS. However, the mean lengths of the 5UIs and 3UIs were both higher than those for the CIs, and the mean 5UI length of *C. maxima* was even about 3.40-fold that of its CIs. This indicated that UTRs are more inclined to have longer introns than CDSs [28,47,48].

The splice site (SS) pair types greatly affect the efficiency of recruiting splicing machinery [56,59,60]. For *C. medica* and *C. sinensis*, we identified 11 and 9 types of SS pairs in the UIs, respectively. However, there were only three types of SS pairs ('GT-AG', 'GC-AG' and 'AT-AC') in the UIs of *C. hindsii*, *C. maxima*, *C. reticulata* and *C. cavaleriei*. This indicated that the SS pair types varied among different citrus species. Furthermore, it was found that 'GT-AG' was the most frequent SS pair used by both 5UIs and 3UIs.

4.2. Most Common UI-Containing Transcripts Were Involved in Gene Expression or Gene Expression Regulation

The UTR regions were thought to be under less stringent substitutional constraint than the CDSs [48]. As introns do not influence the encoded protein structure at all, mutation occurring in the introns will not affect protein sequences and functions, thus serving as mutational buffer in eukaryotic genomes [4]. In this study, through coexistence analysis, we found that most UIs were species specific, indicating that many mutations accumulated in UIs during citrus evolution. Totally, we identified 81 common 5UIs and 26 common 3UIs among all the six citrus species. In the study of Cenik et al. [58], they found that many genes with regulatory roles had 5UIs. Among the common UI-containing transcripts, many genes were found to be involved in gene transcription or gene expression regulation pathways, and more than 30 UI-Ts were involved in gene-expression-related pathways such as 'RNA biosynthesis', 'RNA processing', 'protein modification', 'protein homeostasis', 'protein biosynthesis' and so on. Moreover, nine of these UI-Ts were transcription factor genes, indicating again that the common UI-containing transcripts were closely related to gene expression.

Moreover, many UI-Ts containing common UIs were also found to be involved in gene transcription regulation [10,11]. Suppressor of Ty16 (Spt16), a component of the facilitates chromatin transcription (FACT) complex, is a histone chaperone involved in gene expression [61]. Protein methylation plays a pivotal role in the regulation of various cellular processes including chromatin remodeling and gene expression [62]. In this study, a component, SPT16, of the FACT histone chaperone complex gene (cs6g01200.1) containing common 5UIs and a class-VI histone methyltransferase (SMYD) gene (cs8g12080.2) carrying common 3UIs were identified as 'chromatin organisation'-related UI-Ts. The carbon catabolite repressor 4 complex is involved in the control of gene expression [63]. In this study, a CCR4-associated factor 1 homolog 6 gene (cs1g11100.1) was identified to contain common 5UIs. The pentatricopeptide repeat-containing protein is thought to be the main mediator of post-transcriptional regulation of organelles [28,64]. In this study, 10 genes (including 7 3UI-Ts, 2 5UI-Ts and 1 5&3UI-T) encoding pentatricopeptide repeat-containing proteins were found to contain common UIs. It was, thus, suggested that these common UIs contribute greatly to gene expression regulation. Additionally, we found that sweet orange shared the largest number of common UIs with its original parents, suggesting that UIs might have the potential to be used in the citrus origination and evolution studies.

4.3. UIs Might Function in Cell Development, Stress Responses and Phytohormone Metabolism and Signaling

In our study, several cell-development-related genes were identified to be UI-Ts. Fasciclin-like arabinogalactan protein 4 was a cell surface adhesion protein that is required for normal cell expansion [65]. In this study, three arabinogalactan protein (Fasciclin) transcripts (cs2g20030.1, cs2g20030.2 and cs2g20030.3) were found to harbor a common 5UI. Moreover, two 'cell division'-related genes, protein kinase (CDKE/CDK8) (cs5g01900.2) and SYP1-group Qa-type SNARE component (cs5g26090.1), were also identified as UI-Ts. The cyclin-dependent kinases play important roles in controlling cell division and modulating transcription in response to internal and external environmental changes [66]. CDK8 is one of the most widely studied components of eukaryotic mediator complexes. In arabidopsis, the regulatory functions of AtCDK8 in defense have been demonstrated [67]. *SNARE* genes have been proved to be involved in innate immunity [68]. In addition, in this study, we also found one 'external stimuli response'-related gene (regulatory protein (GCN4) of RIN4 activity, cs1g13980.1) and two zinc finger A20 and AN1 domain-containing stress-associated protein genes (cs5g09740.1 and cs6g15060.2) contain common 5UIs, which were plant defense related [69,70]. Therefore, the roles of these Uis in citrus stress resistance should be further focused on.

Phytohormones function greatly in regulating gene expression. The genes encoding regulatory protein (EAR1) of abscisic acid signaling (cs8g14510.2), auxin efflux transporter (PILS) (cs2g13710.1), RALF-peptide receptor (CrRLK1L) (cs6g10250.2) and auxin transporter (PILS) (cs2g13710.1) were identified to contain common 5UIs. Phytosulfokine (PSK) is a peptide phytohormone that acts as a growth factor [71], and PSK stimulator 1 (PSI1) was reported to be required during plant vegetative growth and reproduction [71]. In this study, a *PSI1* gene (cs6g15600.2) was found to contain a common 5UI. All these factors indicate that these 5UIs might play roles in regulating phytohormone metabolism and signaling in citrus.

Four 'vesicle trafficking'-related genes were found to bear common 5UIs. Among them, the expression of *VPS28 of ESCRT-I complex* (cs2g06750.1) and its 3UI has been identified to be significantly negatively correlated [28]. Additionally, three genes encoding FT-interacting proteins, three genes encoding F-box/Kelch-repeat proteins, two transcripts encoding chaperone protein dnaJ 49 and many genes with unknown functions were also identified to contain common UIs. The functions of these UIs need to be further investigated in the future research.

5. Conclusions

In this study, based on the genome data of six citrus species, the introns located in UTRs and CDSs were identified and characterized. Our study revealed that the UI length, abundance, density and SS pair types varied a lot among the six citrus species. Moreover, 81 5UIs and 26 3UIs were found to commonly exist in all the six citrus species, which seemed to contribute greatly to gene expression regulation and might have great potential to be used in the citrus origin and evolution studies. Moreover, many UIs were predicted to contribute to the cell development, stress responses and phytohormone metabolism and signaling processes in citrus. The results obtained from this study could provide evidence for further understanding the regulatory roles of UIs in citrus.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/horticulturae8050434/s1, Table S1: Information on the 3UIs in *Citrus hindsii*; Table S2: Information on the 5UIs in *Citrus hindsii*; Table S3: Information on the 3UIs in *Citrus maxima*; Table S4: Information on the 5UIs in *Citrus maxima*; Table S5: Information on the 3UIs in *Citrus reticulata*; Table S6: Information on the 5UIs in *Citrus reticulata*; Table S7: Information on the 3UIs in *Citrus cavaleriei*; Table S8: Information on the 5UIs in *Citrus cavaleriei*; Table S9: Information on the 3UIs in *Citrus medica*; Table S10: Information on the 5UIs in *Citrus medica*; Table S11: The coexistence analysis results for 3UIs in the six citrus species; Table S12: The coexistence analysis

results for 5UIs in the six citrus species; Table S13: MapMan annotation results of the transcripts containing common UIs.

Author Contributions: Conceptualization, C.C. and S.W.; formal analysis, X.S. and C.C.; data curation, C.C., S.W., X.S., B.W., J.W., S.Y. and Y.Z.; writing—original draft preparation, C.C. and S.W.; writing—review and editing, C.C. and S.W.; funding acquisition, C.C. and S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Fund for High-Level Talents of Shanxi Agricultural University (2021XG010); the National Natural Science Foundation of China (32060664) and Major Special Projects and Key R&D Projects in Yunnan Province (202102AE090054).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data generated or analyzed during this study are included in this published article and its Supplemental Data.

Acknowledgments: The authors would like to thank Zhenhua Zhuang of Chengdu Life Baseline Company for his guidance and assistance in the UTR intron identification and characterization analysis.

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

References

1. Berget, S.M.; Moore, C.; Sharp, P.A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 3171–3175. [CrossRef] [PubMed]
2. Chow, L.T.; Gelinas, R.E.; Broker, T.R.; Roberts, R.J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **1977**, *12*, 1–8. [CrossRef]
3. Vonk, J.; Shackelford, T. *Encyclopedia of Animal Cognition and Behavior*; Springer: New York, NY, USA, 2017.
4. Jo, B.S.; Choi, S.S. Introns: The Functional Benefits of Introns in Genomes. *Genom. Inform.* **2015**, *13*, 112–118. [CrossRef] [PubMed]
5. Lane, N.; Martin, W. The energetics of genome complexity. *Nature* **2010**, *467*, 929–934. [CrossRef] [PubMed]
6. Chorev, M.; Carmel, L. The function of introns. *Front. Genet.* **2012**, *3*, 55. [CrossRef] [PubMed]
7. Lynch, M. Intron evolution as a population-genetic process. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6118–6123. [CrossRef]
8. Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **2008**, *40*, 1413–1415. [CrossRef]
9. Xia, T.; Yang, Y.; Zheng, H.; Han, X.; Jin, H.; Xiong, Z.; Qian, W.; Xia, L.; Ji, X.; Li, G.; et al. Efficient expression and function of a receptor-like kinase in wheat powdery mildew defence require an intron-located MYB binding site. *Plant Biotechnol. J.* **2021**, *19*, 897–909. [CrossRef]
10. Baier, T.; Jacobebbinghaus, N.; Einhaus, A.; Lauersen, K.J.; Kruse, O. Introns mediate post-transcriptional enhancement of nuclear gene expression in the green microalga *Chlamydomonas reinhardtii*. *PLoS Genet.* **2020**, *16*, e1008944. [CrossRef]
11. Remy, E.; Cabrito, T.R.; Batista, R.A.; Hussein, M.A.; Teixeira, M.C.; Athanasiadis, A.; Sá-Correia, I.; Duque, P. Intron retention in the 5′UTR of the novel ZIF2 transporter enhances translation to promote zinc tolerance in arabidopsis. *PLoS Genet.* **2014**, *10*, e1004375. [CrossRef]
12. Gage, J.L.; Mali, S.; Mcloughlin, F.; Khaipho-Burch, M.; Monier, B.; Bailey-serres, J.; Vierstra, R.; Buckler, E.S. Variation in upstream open reading frames contributes to allelic diversity in protein abundance. *Proc. Natl. Acad. Sci. USA* **2021**, *119*, e2112516119. [CrossRef] [PubMed]
13. Schwartz, S.; Meshorer, E.; Ast, G. Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* **2009**, *16*, 990–995. [CrossRef]
14. Lewis, B.P.; Green, R.E.; Brenner, S.E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 189–192. [CrossRef] [PubMed]
15. Kalyna, M.; Simpson, C.G.; Syed, N.H.; Lewandowska, D.; Marquez, Y.; Kusenda, B.; Marshall, J.; Fuller, J.; Cardle, L.; McNicol, J.; et al. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.* **2012**, *40*, 2454–2469. [CrossRef] [PubMed]
16. Carvunis, A.R.; Rolland, T.; Wapinski, I.; Calderwood, M.A.; Yildirim, M.A.; Simonis, N.; Charloteaux, B.; Hidalgo, C.A.; Barbette, J.; Santhanam, B.; et al. Proto-genes and de novo gene birth. *Nature* **2012**, *487*, 370–374. [CrossRef] [PubMed]
17. Rearick, D.; Prakash, A.; McSweeny, A.; Shepard, S.S.; Fedorova, L.; Fedorov, A. Critical association of ncRNA with introns. *Nucleic Acids Res.* **2011**, *39*, 2357–2366. [CrossRef]
18. Wong, A.C.H.; Rasko, J.E.J. Splice and dice: Intronic microRNAs, splicing and cancer. *Biomedicines* **2021**, *9*, 1268. [CrossRef]
19. Wahl, M.C.; Will, C.L.; Luhrmann, R. The spliceosome: Design principles of a dynamic RNP machine. *Cell* **2009**, *136*, 701–718. [CrossRef]

20. Le Hir, H.; Nott, A.; Moore, M.J. How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.* **2003**, *28*, 215–220. [CrossRef]
21. Oswald, A.; Oates, A.C. Control of endogenous gene expression timing by introns. *Genome Biol.* **2011**, *12*, 107. [CrossRef]
22. Gottlieb, L.D.; Ford, V.S. The 5′ leader of plant PgiC has an intron: The leader shows both the loss and maintenance of constraints compared with introns and exons in the coding region. *Mol. Biol. Evol.* **2002**, *19*, 1613–1623. [CrossRef] [PubMed]
23. Parra, G.; Bradnam, K.; Rose, A.B.; Korf, I. Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nucleic Acids Res.* **2011**, *39*, 5328–5337. [CrossRef] [PubMed]
24. Laxa, M.; Muller, K.; Lange, N.; Doering, L.; Pruscha, J.T.; Peterhansel, C. The 5′UTR Intron of Arabidopsis GGT1 Aminotransferase Enhances Promoter Activity by Recruiting RNA Polymerase II. *Plant Physiol.* **2016**, *172*, 313–327. [CrossRef] [PubMed]
25. Shaul, O. How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* **2017**, *91 Pt B*, 145–155. [CrossRef]
26. You, H.; Sun, B.; Li, N.; Xu, J.W. Efficient expression of heterologous genes by the introduction of the endogenous glyceraldehyde-3-phosphate dehydrogenase gene intron 1 in Ganoderma lucidum. *Microb. Cell Fact.* **2021**, *20*, 164. [CrossRef] [PubMed]
27. Bicknell, A.A.; Cenik, C.; Chua, H.N.; Roth, F.P.; Moore, M.J. Introns in UTRs: Why we should stop ignoring them. *Bioessays* **2012**, *34*, 1025–1034. [CrossRef]
28. Shi, X.; Wu, J.; Mensah, R.A.; Tian, N.; Liu, J.; Liu, F.; Chen, J.; Che, J.; Guo, Y.; Wu, B.; et al. Genome-wide identification and characterization of UTR-introns of Citrus sinensis. *Int. J. Mol. Sci.* **2020**, *21*, 3088. [CrossRef]
29. Matsumoto, K.; Wassarman, K.M.; Wolffe, A.P. Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. *EMBO J.* **1998**, *17*, 2107–2121. [CrossRef]
30. Furger, A.; O'Sullivan, J.M.; Binnie, A.; Lee, B.A.; Proudfoot, N.J. Promoter proximal splice sites enhance transcription. *Genes Dev.* **2002**, *16*, 2792–2799. [CrossRef]
31. Masuda, S.; Das, R.; Cheng, H.; Hurt, E.; Dorman, N.; Reed, R. Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev.* **2005**, *19*, 1512–1517. [CrossRef]
32. Salimonti, A.; Carbone, F.; Romano, E.; Pellegrino, M.; Benincasa, C.; Micali, S.; Tondelli, A.; Conforti, F.L.; Perri, E.; Ienco, A.; et al. Association study of the 5′UTR Intron of the *FAD2-2* gene with oleic and linoleic acid content in *Olea europaea* L. *Front. Plant Sci.* **2020**, *11*, 66. [CrossRef] [PubMed]
33. Barrou, B.M.; Gruessner, A.C.; Sutherland, D.E.; Gruessner, R.W. A potent enhancer element in the 5′-UTR intron is crucial for transcriptional regulation of the human ubiquitin C gene. *Gene* **2009**, *448*, 88–101.
34. Grant, T.N.; De La Torre, C.M.; Zhang, N.; Finer, J.J. Synthetic introns help identify sequences in the 5′ UTR intron of the Glycine max polyubiquitin (Gmubi) promoter that give increased promoter activity. *Planta* **2017**, *245*, 849–860. [CrossRef] [PubMed]
35. David-Assael, O.; Berezin, I.; Shoshani-Knaani, N.; Saul, H.; Mizrachy-Dagri, T.; Chen, J.X.; Brook, E.; Shaul, O. AtMHX is an auxin and ABA-regulated transporter whose expression pattern suggests a role in metal homeostasis in tissues with photosynthetic potential. *Funct. Plant Biol.* **2006**, *33*, 661–672. [CrossRef]
36. Akua, T.; Shaul, O. The Arabidopsis thaliana MHX gene includes an intronic element that boosts translation when localized in a 5′ UTR intron. *J. Exp. Bot.* **2013**, *64*, 4255–4270. [CrossRef]
37. Cenik, C.; Chua, H.N.; Zhang, H.; Tarnawsky, S.P.; Akef, A.; Derti, A.; Tasan, M.; Moore, M.J.; Palazzo, A.F.; Roth, F.P. Genome analysis reveals interplay between 5′UTR introns and nuclear mRNA export for secretory and mitochondrial genes. *PLoS Genet.* **2011**, *7*, e1001366. [CrossRef]
38. Chang, Y.F.; Imam, J.S.; Wilkinson, M.F. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **2007**, *76*, 51–74. [CrossRef]
39. Barrett, L.W.; Fletcher, S.; Wilton, S.D. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* **2012**, *69*, 3613–3634. [CrossRef]
40. Kurilla, A.; Szőke, A.; Auber, A.; Káldi, K.; Silhavy, D. Expression of the translation termination factor eRF1 is autoregulated by translational readthrough and 3′UTR intron-mediated NMD in *Neurospora crassa*. *FEBS Lett.* **2020**, *594*, 3504–3517. [CrossRef]
41. Zhang, J.; Sun, X.; Qian, Y.; Maquat, L.E. Intron function in the nonsense-mediated decay of beta-globin mRNA: Indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. *RNA* **1998**, *4*, 801–815. [CrossRef]
42. Weischenfeldt, J.; Damgaard, I.; Bryder, D.; Theilgaard-Monch, K.; Thoren, L.A.; Nielsen, F.C.; Jacobsen, S.E.; Nerlov, C.; Porse, B.T. NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev.* **2008**, *22*, 1381–1396. [CrossRef] [PubMed]
43. Pan, Q.; Saltzman, A.L.; Kim, Y.K.; Misquitta, C.; Shai, O.; Maquat, L.E.; Frey, B.J.; Blencowe, B.J. Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev.* **2006**, *20*, 153–158. [CrossRef] [PubMed]
44. Saltzman, A.L.; Pan, Q.; Blencowe, B.J. Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev.* **2011**, *25*, 373–384. [CrossRef] [PubMed]
45. Baskerville, S.; Bartel, D.P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **2005**, *11*, 241–247. [CrossRef] [PubMed]
46. Bartel, D.P. MicroRNAs: Target recognition and regulatory functions. *Cell* **2009**, *136*, 215–233. [CrossRef] [PubMed]
47. Chung, B.Y.; Simons, C.; Firth, A.E.; Brown, C.M.; Hellens, R.P. Effect of 5′UTR introns on gene expression in Arabidopsis thaliana. *BMC Genom.* **2006**, *7*, 120. [CrossRef]

48. Hong, X.; Scofield, D.G.; Lynch, M. Intron size, abundance, and distribution within untranslated regions of genes. *Mol. Biol. Evol.* **2006**, *23*, 2392–2404. [CrossRef]
49. Cheng, C.; Shi, X.; Wu, J.; Zhang, Y.; Lü, P. Genome-scale computational identification and characterization of UTR introns in *Atalantia buxifolia*. *Horticulturae* **2021**, *7*, 556. [CrossRef]
50. Xu, Q.; Chen, L.L.; Ruan, X.; Chen, D.; Zhu, A.; Chen, C.; Bertrand, D.; Jiao, W.B.; Hao, B.H.; Lyon, M.P.; et al. The draft genome of sweet orange (Citrus sinensis). *Nat. Genet.* **2013**, *45*, 59–66. [CrossRef]
51. Wu, G.A.; Prochnik, S.; Jenkins, J.; Salse, J.; Hellsten, U.; Murat, F.; Perrier, X.; Ruiz, M.; Scalabrin, S.; Terol, J.; et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat. Biotechnol.* **2014**, *32*, 656–662. [CrossRef]
52. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef] [PubMed]
53. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef] [PubMed]
54. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]
55. Thimm, O.; Blasing, O.; Gibon, Y.; Nagel, A.; Meyer, S.; Kruger, P.; Selbig, J.; Muller, L.A.; Rhee, S.Y.; Stitt, M. MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **2004**, *37*, 914–939. [CrossRef] [PubMed]
56. Kumari, A.; Sedehizadeh, S.; Brook, J.D.; Kozlowski, P.; Wojciechowska, M. Differential fates of introns in gene expression due to global alternative splicing. *Hum. Genet.* **2022**, *141*, 31–47. [CrossRef] [PubMed]
57. Pesole, G.; Mignone, F.; Gissi, C.; Grillo, G.; Licciulli, F.; Liuni, S. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* **2001**, *276*, 73–81. [CrossRef]
58. Cenik, C.; Derti, A.; Mellor, J.C.; Berriz, G.F.; Roth, F.P. Genome-wide functional analysis of human 5′ untranslated region introns. *Genome Biol.* **2010**, *11*, R29. [CrossRef]
59. Goguel, V.; Rosbash, M. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* **1993**, *72*, 893–901. [CrossRef]
60. Brown, J.W.; Simpson, C.G.; Thow, G.; Clark, G.P.; Jennings, S.N.; Medina-Escobar, N.; Haupt, S.; Chapman, S.C.; Oparka, K.J. Splicing signals and factors in plant intron removal. *Biochem. Soc. Trans.* **2002**, *30*, 146–149. [CrossRef]
61. Yang, L.; Wang, X.; Jiao, X.; Tian, B.; Zhang, M.; Zhou, C.; Wang, R.; Chen, H.; Wang, B.; Li, J.; et al. Suppressor of Ty 16 promotes lung cancer malignancy and is negatively regulated by miR-1227-5p. *Cancer Sci.* **2020**, *111*, 4075–4087. [CrossRef]
62. Tracy, C.; Warren, J.S.; Szulik, M.; Wang, L.; Garcia, J.; Makaju, A.; Russell, K.; Miller, M.; Franklin, S. The Smyd Family of Methyltransferases: Role in Cardiac and Skeletal Muscle Physiology and Pathology. *Curr. Opin. Physiol.* **2018**, *1*, 140–152. [CrossRef] [PubMed]
63. Fang, J.C.; Tsai, Y.C.; Chou, W.L.; Liu, H.Y.; Chang, C.C.; Wu, S.J.; Lu, C.A. A CCR4-associated factor 1, OsCAF1B, confers tolerance of low-temperature stress to rice seedlings. *Plant Mol. Biol.* **2021**, *105*, 177–192. [CrossRef] [PubMed]
64. Manna, S. An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* **2015**, *113*, 93–99. [CrossRef] [PubMed]
65. Costa, M.; Pereira, A.M.; Pinto, S.C.; Silva, J.; Pereira, L.G.; Coimbra, S. In silico and expression analyses of fasciclin-like arabinogalactan proteins reveal functional conservation during embryo and seed development. *Plant Reprod.* **2019**, *32*, 353–370. [CrossRef] [PubMed]
66. Malumbres, M. Cyclin-dependent kinases. *Genome Biol.* **2014**, *15*, 122. [CrossRef]
67. Zhu, Y.F.; Schluttenhoffer, C.M.; Wang, P.C.; Fu, F.Y.; Thimmapuram, J.; Zhu, J.K.; Lee, S.Y.; Yun, D.J.; Mengiste, T. Cyclin-dependent kinase8 differentially regulates plant immunity to fungal pathogens through kinase-dependent and -independent functions in arabidopsis. *Plant Cell* **2014**, *26*, 4149–4170. [CrossRef]
68. Slane, D.; Reichardt, I.; El Kasmi, F.; Bayer, M.; Jurgens, G. Evolutionarily diverse SYP1 Qa-SNAREs jointly sustain pollen tube growth in Arabidopsis. *Plant J.* **2017**, *92*, 375–385. [CrossRef]
69. Giri, J.; Vij, S.; Dansana, P.K.; Tyagi, A.K. Rice A20/AN1 zinc-finger containing stress-associated proteins (SAP1/11) and a receptor-like cytoplasmic kinase (OsRLCK253) interact via A20 zinc-finger and confer abiotic stress tolerance in transgenic Arabidopsis plants. *New Phytol.* **2011**, *191*, 721–732. [CrossRef]
70. Kim, G.D.; Cho, Y.H.; Yoo, S.D. Regulatory functions of evolutionarily conserved AN1/A20-like Zinc finger family proteins in Arabidopsis stress responses under high temperature. *Biochem. Biophys. Res. Commun.* **2015**, *457*, 213–220. [CrossRef]
71. Stuhrwohldt, N.; Hartmann, J.; Dahlke, R.I.; Oecking, C.; Sauter, M. The PSI family of nuclear proteins is required for growth in arabidopsis. *Plant Mol. Biol.* **2014**, *86*, 289–302. [CrossRef]