Abstract: Accumulated evidence has shown that CDS introns (CIs) play important roles in regulating gene expression. However, research on UTR introns (UIs) is limited. In this study, UIs (including 5′UTR and 3′UTR introns (5UIs and 3UIs)) were identified from the *Atalantia buxifolia* genome. The length and nucleotide distribution characteristics of both 5UIs and 3UIs and the distributions of cis-acting elements and transcription factor binding sites (TFBSs) in 5UIs were investigated. Moreover, PageMan enrichment analysis was applied to show the possible roles of transcripts containing UIs (UI-Ts). In total, 1077 5UIs and 866 3UIs were identified from 897 5UI-Ts and 670 3UI-Ts, respectively. Among them, 765 (85.28%) 5UI-Ts and 527 (78.66%) 3UI-Ts contained only one UI, and 94 (6.38%) UI-Ts contained both 5UI and 3UI. The UI density was lower than that of CDS introns, but their mean and median intron sizes were ~2 times those of the CDS introns. The *A. buxifolia* 5UIs were rich in gene-expression-enhancement-related elements and contained many TFBSs for BBR-BPC, MIKC_MADS, AP2 and Dof TFs, indicating that 5UIs play a role in regulating or enhancing the expression of downstream genes. Enrichment analysis revealed that UI-Ts involved in ‘not assigned’ and ‘RNA’ pathways were significantly enriched. Noteworthily, 119 (85.61%) of the 3UI-Ts were genes encoding pentatricopeptide (PPP) repeat-containing proteins. These results will be helpful for the future study of the regulatory roles of UIs in *A. buxifolia*.

Keywords: *Atalantia buxifolia*; genome-wide identification; UTR intron; gene expression regulation; pathway-enrichment analysis

1. Introduction

Introns, the removed genomic sequences from corresponding RNA transcripts, have been intensely studied since their first discovery [1,2]. They can be generally divided into three groups (Groups I–III). Group I and II introns are of self-splicing activity and are both widely identified in some bacterial and organelar genomes [3,4], while group III introns are spliceosomal introns mainly found in the nuclear genomes of eukaryotes, and the excision of this kind of introns is spliceosome-dependent [5]. Accumulated evidence has shown that the presence of introns and the behaviors of spliceosomes affect almost every step of gene expression [6,7]. Some introns have the effect of boosting gene expression [8–11], and this intronic effect is called intron-mediated enhancement (IME). The addition of *alcohol dehydrogenase* (Adh1) first intron increased the expression of a maize chimeric *chloramphenicol acetyltransferase (CAT)* gene for 100-fold [12]. The *Shrunken-1* (*Sh1*) intron 1 could enhance chimeric gene expression by approximately 100-fold, and the combined *Sh1* first exon and intron 1 could enhance report gene expression by more than 1000-fold [13]. The expression of petunia *small subunit of ribulose bisphosphate carboxylase (rbcS)* in transgenic tobacco overexpressing its gDNA expressed about five-fold higher than that in a transgenic plant overexpressing its cDNA [14]. The first intron of Arabidopsis *elongation factor 1 beta* gene (*AtEF-1β*) was proved to be required for the gene’s high expression due to the enhancer-like element that existed in this coding sequence intron (CI) [15].
Except in the CDS, introns are also located within the untranslated regions (UTRs) of a gene [16]. Much evidence has proved that 5'UTs function in regulating gene expression [17–21]. The 5'UT of the soybean *polyubiquitin* (*Gmubi*) gene seems to act as a promoter of regulatory elements and is closely associated with the gene's high expression [22]. The 5'UT of Arabidopsis *glyoxylate aminotransferase 1* (*GGT1*) has been proved to contribute to maximum transcript abundance [10,23]. The addition of *Gladiolus polyubiquitin* (*GUBQ1*) 5'UT to its promoter led to enhanced *GUS* expression in transgenic Gladiolus and Arabidopsis plants compared with transgenic plants only overexpressing its promoter [24]. The sesame *FAD2* 5'UT was enhanced to 100-fold of *GUS* expression as compared with 5'UT-less controls in transgenic Arabidopsis tissues [25].

Transcripts containing 3'UT were generally considered non-functional because they could stimulate mRNA degradation by nonsense-mediated decay (NMD) [26–30]. However, numerous studies have verified that 3'UTs also play a part in normal gene expression modulation [29,31–35]. For example, the yeast *HAC1* transcript containing a retained 3'UTR intron could block its mRNA translation [32]. Moreover, many functional transcripts containing 3'UT either boosted expression or differential exon usage upon NMD inhibition [36,37].

Over the past decade, whole genome-wide structure and sequence characterizations of UIs have been successfully identified from several species. In the genomes of *A. thaliana*, *Drosophila melanogaster*, human, and mouse, CIs and UIs were identified and the 5'UTs sizes were roughly twice of CIs and 3'UTs [38]. In addition, 5'UTs could significantly enhance gene expression, and the intron length greatly influences the gene expression level in Arabidopsis [39]. Roy et al. [40] conducted the evolutionary conservation analysis of UTR splicing in *Cryptococcus neoformans* and found that the splicing boundaries in 5'UTRs were more conserved than in 3'UTRs. Cenik et al. [41] discovered that human genes with regulatory roles were surprisingly rich in 5'UT. Shi et al. [42] analyzed the UIs characteristics and revealed UIs' differential expression in different organs of sweet orange. These findings can provide information for evolution and functional significances of UIs.

Chinese box orange (*Atalantia buxifolia*, once named as *Severinia buxifolia*) is an evergreen citrus plant native to China and some Asian countries [43]. Its roots and branches are rich in various bioactive compounds [44,45]. Its genome has been published with accurate UTR information (http://citrus.hzau.edu.cn/, accessed on 1 March 2020), which facilitated the genome-wide identification of UIs. In this study, based on the *A. buxifolia* genome data, we identified and characterized the introns, including 5'UTs, 3'UTs and CIs, in *A. buxifolia*. Furthermore, transcripts containing UIs were further bioinformatically analyzed to show their functions. The results obtained in this study will be helpful for understanding the regulatory roles of UIs in gene expression in *A. buxifolia*.

2. Materials and Methods

2.1. Data Preparation

The *A. buxifolia* genome data file was downloaded from the Orange Annotation Project (http://citrus.hzau.edu.cn/orange/download/index.php/, accessed on 1 March 2020) [46]. All the *A. buxifolia* CDS sequences were submitted to Mercator v.3.6 (https://plabipd.de/portal/mercator-sequence-annotation, accessed on 2 March 2020) to obtain the mapping file used for Mapman analysis.

2.2. Genome-Wide Identification of *A. buxifolia* UIs

Based on the annotation information of the *A. buxifolia* genome, introns in CDSs, 5'UTR and 3'UTR were separately extracted. Then, the 5'UTs and 3'UTs were identified from the *A. buxifolia* genome according to the method described by Shi et al. [42]. Introns between UTR exons were extracted according to the genome annotation file, and introns showed retention in the exon region of any transcripts were excluded to ensure the identified UIs strictly exist in the intron regions. Information of the identified 5'UTs and 3'UTs were shown in Additional file Tables S1 and S2, respectively. UI density, position preference, length and
nucleotide composition statistical analysis were performed using Perl, and figures were drawn using ggplot2 [42].

2.3. Gene Pathway-Enrichment Analysis of UI-Ts

To annotate and illustrate the *A. buxifolia* genes containing UIs, we conducted PageMan pathway-enrichment analysis for all the UI-Ts, 5UI-Ts and 3UI-Ts, respectively. Briefly, the genes containing 5UI and/or 3UI were first subjected to Mapman analysis based on the abovementioned *A. buxifolia* mapping file. Then, pathway-enrichment analysis was performed using PageMan embedded in MapMan [47]. By applying the Benjamini and Hochberg adjustment, pathways with corrected *p* value < 0.05 were considered significantly enriched by UI-Ts.

2.4. Cis-Acting Element and Transcription Factor Binding Sites (TFBS) Prediction Analysis of 5UI Sequences

PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, accessed on 27 March 2020) was applied to predict the possible cis-acting elements in the sequences of 1077 *A. buxifolia* 5UIs [48]. The TFBSs in all the 5UI sequences were categorized and matched with the *Citrus sinensis* TFs library using PlantTFDB (http://plantregmap.cbi.pku.edu.cn/, accessed on 27 March 2020). Additionally, the parameters used for TFBSs prediction were set as *q*-value ≤ 0.05 and *p*-value ≤ 10^-7..

3. Results

3.1. Identification of Introns in *A. buxifolia* CDSs and 5′ and 3′ Untranslated Regions (UTRs)

In total, we identified 16,218 5'UTRs, 16,337 3'UTRs and 28,412 CDSs from the *A. buxifolia* genome. Additionally, 597 5'UTRs (accounting for 3.68% of 5'UTRs), 452 3'UTRs (2.77% of 3'UTRs) and 21,005 CDSs (73.93% of CDSs) were found to contain introns. The intron-harboring ratios for 5'UTRs and 3'UTRs were significantly lower than those in CDSs, which may well explain why the UIs were often overlooked. After normalizing the intron density to the average number of introns per nucleotide of each gene transcript sequence, we found that the intron density followed the order: CDS > 5'UTR > 3'UTR, which was 3.11 × 10^-3, 8.42 × 10^-5 and 3.08 × 10^-5, respectively (Table 1). The 5UI density is ~2.7 times higher than 3UI's, which is only ~2.7% of the CIs.

In total, 1077 5UIs and 866 3UIs were identified from 897 5UTR-containing gene transcripts (corresponding to 597 genes) and 670 3UTR-containing gene transcripts (452 genes), respectively. Among these UI-containing gene transcripts (UI-Ts), 94 transcripts (91 genes) contained both 5UI and 3UI. Approximately 85.28% of the 5UI-Ts and 78.66% of the 3UI-Ts contain only one 5UI or 3UI (Table 2). The proportion of UI-Ts containing two or more UIs dropped dramatically. Only 11.82% of the 5UI-Ts contain two 5UIs, and 13.43% of the 3UI-Ts contain two 3UIs (Table 2). The proportion of UI-Ts containing more than two 5UIs or 3UIs accounted for about 2.98% and 5.23%, respectively (Table 2). Some UI-Ts were found to contain several UIs (Table 2), for example, one cysteine-rich receptor-like kinase 8 gene (CRK8, sb36468.1) had eight 5UIs; an unknown gene transcript (sb14804.1) had seven 5UIs; one PPRP gene (sb26751.1) had six 5UIs; a PPRP gene (sb25651.1) and a LRR and NB-ARC domains-containing disease resistance protein gene (sb31448.1) had six 3UIs; an unknown gene transcript (sb31747.1) contained five 5UIs and five 3UIs; a AP2/ERF and bHLH DNA-binding superfamily protein (sb28046.1) both had five 5UIs; a bromodomain transcription factor (sb13718.1), an extra-large G-protein 1 (sb31879.1) and an unknown gene transcript (sb10627.1) had five 3UIs; a TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 2 (sb19593.1), a Leucine-rich repeat protein kinase family protein (sb20478.1), a putative clathrin assembly protein (sb12332.1) and an unknown gene (sb28722.1) had four 5UIs; and three PPRP genes (sb17114.1, sb26143.1 and sb25688.1) and a Class II aminoacyl-IRNA and biotin synthetases superfamily protein gene (SYNC3, sb31386.1) had four 3UIs.
Table 1. Statistics information of 5′ UTR, CDS and 3′ UTR in A. buxifolia.

| No. of Sequences | Sequences with Introns | Total Bases (Genomic) | Intron/Sequence | No. of Introns/Nucleotide (mRNA) |
|------------------|------------------------|-----------------------|-----------------|----------------------------------|
| 5′UTR            | 16,218                 | 597                   | 1.28 × 10^7     | 0.07                             | 8.42 × 10^−5 |
| CDS              | 28,412                 | 21,005                | 3.64 × 10^7     | 3.99                             | 3.11 × 10^−3 |
| 3′UTR            | 16,337                 | 452                   | 2.81 × 10^7     | 0.05                             | 3.08 × 10^−5 |

Table 2. The UI numbers in the A. buxifolia gene transcripts containing 5UI (5UI-Ts) and 3UI (3UI-Ts). There are 897 and 670 gene transcripts containing 5UI and 3UI, respectively. NS: not shown.

| UI Number | 5UI-Ts Number/Percentage (Gene ID) | 3UI-Ts Number/Percentage (Gene ID) |
|-----------|-----------------------------------|-----------------------------------|
| 1 UI      | 765/85.28% (NS)                   | 527/78.66% (NS)                   |
| 2 UIs     | 106/11.82% (NS)                   | 90/13.43% (NS)                    |
| 3 UIs     | 16/1.78% (sb11639.1, sb32006.1, sb10151.1, sb18930.1, sb18698.1, sb31744.1, sb37205.1, sb20998.1, sb14514.1, sb29193.1, sb21675.1, sb20177.1, sb34492.1, sb36282.1, sb12383.1) | 25/3.73% (sb13064.1, sb32006.1, sb37841.1, sb18145.1, sb18385.1, sb34575.1, sb33241.1, sb37886.1, sb24116.1, sb10905.1, sb26419.1, sb26292.1, sb35222.1, sb37027.1, sb32422.1, sb12195.1, sb26143.1, sb10905.1, sb20663.1, sb32886.1, sb22221.1, sb35222.1, sb26419.1, sb36282.1, sb12383.1) |
| 4 UIs     | 4/0.45% (sb19593.1, sb20478.1, sb12332.1, sb28722.1) | 4/0.60% (sb17114.1, sb31386.1, sb26143.1, sb25688.1) |
| 5 UIs     | 3/0.33% (sb27690.1, sb31747.1, sb28046.1) | 4/0.60% (sb33187.1, sb16027.1, sb17174.1) |
| 6 UIs     | 1/0.11% (sb26751.1) | 2/0.30% (sb25651.1, sb31448.1) |
| 7 UIs     | 1/0.11% (sb14004.1) | 0 |
| 8 UIs     | 1/0.11% (sb36468.1) | 0 |

3.2. Intron Sizes and Distributions within UTRs and CDS

The introns within 5′UTRs, CDSs and 3′UTRs of A. buxifolia varied greatly in amounts and lengths (Figure 1). The average length distributions of A. buxifolia 5UIs and 3UIs were more similar (5′UTR: n = 1077, the mean, median, LQ, UQ and SD length correspond to 823.74, 409, 165, 851 and 2599.52 nucleotides, respectively; 3′UTR: n = 866, the mean, median, LQ, UQ and SD length is 838.85, 452, 141, 932.5 and 1823.39 nucleotides, respectively). Their mean and median intron sizes were ~2 times higher than the CIs (n = 319,907, mean = 430.4 nucleotides, median = 175 nucleotides, LQ = 102 nucleotides, UQ = 464 nucleotides and SD = 1698.36 nucleotides). The frequency of CIs with lengths ranging from 100 to 300 nucleotides was significantly higher than 5UIs and 3UIs. However, the relative frequencies of short introns < 50 nucleotides and introns > 300 nucleotides of 3UIs and 5UIs were higher than CIs. Similar to sweet orange [42], the A. buxifolia 5UIs and 3UIs are more preferentially located at the stop ends of 5′UTRs and at the beginning of 3′UTRs, respectively.

Figure 1. Length distributions of A. buxifolia 5UIs, 3UIs and CIs. The horizontal axis and the vertical axis represent the size and relative frequencies of introns, respectively.
3.3. Nucleotide Conservation around the Splice Junctions

To show the nucleotide bias around the donor and acceptor sites of 5UIs, CIs and 3UIs, the sequence logos were used [49]. Results showed that both the A. buxifolia 5UIs and CIs possess A/T-rich element around both donor and receptor sites [42]. Moreover, GT-AG was found to be the major splice site pair in both A. buxifolia 5'UTRs (98.32%) and 3'UTRs (98.26%), followed by GC-AG splice site pair (accounting for 1.67% of 5’UTRs and 1.73% of 3’UTRs) (Figure 2).

Figure 2. Nucleotide bias around the donor sites (A) and acceptor sites (B) of A. buxifolia 5UIs, CIs and 3UIs. The x-axis refers to nucleotides around the beginnings or ends of introns, and the letter height reflects the nucleotide deviation.
3.4. Cis-Acting Elements and TFBS Prediction Analysis of 5UIs

In total, 46,543 cis-acting elements belonging to 47 element types were identified from all 5UI sequences; each 5UI contains 43.22 elements on average (Table 3, Additional file Table S3). About 82.92% and 88.02% of 5UIs contained ‘core promoter element around −30 of transcription start’ and ‘common cis-acting element in promoter and enhancer regions’, respectively. Additionally, these two kinds of elements took the largest part, respectively, accounting for 26.01% and 11.47% of the total elements. More than 73.00% 5UI-Ts contain both the two elements. Besides, many light-related elements were identified in 5UI sequences.

TFBS prediction analysis identified 1,092 binding sites of 21 TFs from 90 (44.78%) input 5UI sequences (Table 4, Additional file Table S4). Additionally, TFBSs for BARLEY B RECOMBINANT/ BASIC PENTACYSTEINE (BBR-BPC) transcription factors took up the largest part (24.54%, the matched sequences are GAGA/ CTCT), followed by MADS intervening keratin-like and C-terminal -type MADS (MIKC_MADS) (24.35%), APETALA2/Ethylene-Responsive factor (AP2) (16.12%), DNA binding with one finger (Dof) (13.64%), etc. Among these UIs, the first 5UI (175 bp in length) of homeobox protein 24 gene (HB24, sb16080.1.5UTR_intron.1) contained the largest TFBSs (88), among which the BBR-BPC BSs accounted for 39.08%.

### Table 3. Predicted cis-acting elements in 5UI sequences. Only the elements with number higher than 300 were shown.

| Function                                                                 | Elements Number | 5UI Number (Ratio) |
|-------------------------------------------------------------------------|----------------|--------------------|
| core promoter element around −30 of transcription start                 | 12,106         | 893 (82.92%)       |
| common cis-acting element in promoter and enhancer regions             | 5538           | 948 (88.02%)       |
| light responsive element                                                | 1297           | 271 (25.16%)       |
| part of a conserved DNA module involved in light responsiveness         | 1003           | 489 (45.40%)       |
| part of a light responsive element                                      | 904            | 491 (45.59%)       |
| cis-acting regulatory element essential for the anaerobic induction     | 768            | 426 (39.55%)       |
| cis-acting regulatory element involved in the MeJA-responsiveness       | 678            | 252 (23.40%)       |
| cis-acting regulatory element involved in light responsiveness          | 486            | 270 (25.7%)        |
| cis-acting element involved in the abscisic acid responsiveness         | 430            | 246 (22.84%)       |
| light responsive element                                                | 392            | 271 (25.16%)       |

### Table 4. Predicted transcription factor binding sites (TFBSs) in 5UI sequences. Only the TFBSs with number higher than 10 were shown.

| Family                                                                 | Number (Ratio) |
|-----------------------------------------------------------------------|----------------|
| BARLEY B RECOMBINANT/BASIC PENTACYSTEINE (BBR-BPC)                    | 268 (24.54%)   |
| MADS intervening keratin-like and C-terminal -type MADS (MIKC_MADS)  | 266 (24.35%)   |
| APETALA2/Ethylene-Responsive factor (AP2)                             | 176 (16.12%)   |
| DNA binding with one finger (Dof)                                     | 149 (13.64%)   |
| The three-amino-acid-loop-extension (TALE)                            | 53 (4.85%)     |
| GIBBERELLIC-ACID INSENSITIVE, REPRESSOR of GAI and SCARECROW (GRAS)  | 50 (4.58%)     |
| GATA                                                                  | 25 (2.29%)     |
| The Cys 2 His 2-like fold group (C2H2)                                | 24 (2.20%)     |
| NODULE-INCEPTION-like (Nin-like)                                      | 22 (2.01%)     |
| Leafy (LFY)                                                           | 10 (0.92%)     |

3.5. Gene Pathway-Enrichment Analysis of UI-Containing Transcripts (UI-Ts)

1171 (79.50%) of the UI-Ts were successfully annotated by Mapman. PageMan enrichment analysis of these annotated UI-Ts showed that they were significantly enriched in six pathways (p-value ≤ 0.05), i.e., ‘not assigned. no ontology. pentatricopeptide (PPR) repeat-containing protein’, ‘not assigned. no ontology’, ‘not assigned’, ‘RNA. regulation of transcription’ and ‘RNA. RNA binding’ (Table 5). 5UI-Ts and 3UI-Ts were also respectively subjected to PageMan enrichment analysis. The ‘not assigned’, ‘RNA’, ‘RNA. regulation of
transcription’, ‘protein’, ‘protein degradation’, ‘signaling’ and ‘stress’ pathways ranked top 10 enriched pathways for both 5UI-Ts and 3UI-Ts (Table 6).

Intriguingly, 85.61% (119/139) of the 3UI-Ts were found to be members of PPRP gene family. There are, respectively, 108 (corresponding to 65 genes), 11 (corresponding to 11 genes) and 20 (corresponding to 14 genes) PPRPs containing only 3UI, both 5UI and 3UI, and only 5UI. PPRP gene family can be further divided into two subfamilies: P subfamily and PLS subfamily [50]. Among the 119 PPRP 3UI-Ts, there are 85 P subfamily members and 34 PLS subfamily members. About 48.24% (41/85) of the P subfamily PPRP 3UI-Ts and 64.71% (22/34) PLS subfamily PPRP 3UI-Ts possess only one UI. However, all the UI containing PPRPs have no intron in CDS regions.

Table 5. Six significantly enriched pathways of UI-Ts in A. buxifolia. BIN is the unit used in Mapman graphs denoting a pathway, and each BIN has been assigned a specific ID in Mapman. UI-Ts: UI-containing gene transcripts.

| BIN    | Pathway Name                                      | Number | p-Value     |
|--------|---------------------------------------------------|--------|-------------|
| 35.1.5 | not assigned. no ontology. pentatricopeptide (PPR) repeat-containing protein | 139    | 2.45 × 10^{-22} |
| 35.1   | not assigned. no ontology                        | 217    | 3.40 × 10^{-16} |
| 35     | not assigned                                     | 733    | 2.57 × 10^{-4}   |
| 27.3   | RNA. regulation of transcription                  | 199    | 2.77 × 10^{-4}   |
| 27.4   | RNA. RNA binding                                 | 11     | 4.14 × 10^{-3}   |

Table 6. The top 10 enriched pathways for 5UI-Ts and 3UI-Ts. Enriched pathways ranked top 10 for both 5UI-Ts and 3UI-Ts were underlined. The pathway ‘not assigned’ is not shown.

| BIN    | Pathway Name                                      | 5UI Number | BIN    | Pathway Name                                      | 3UI Number |
|--------|---------------------------------------------------|------------|--------|---------------------------------------------------|------------|
| 35.2   | not assigned. unknown                             | 328        | 35.2   | not assigned. unknown                             | 321        |
| 27     | RNA                                               | 160        | 35.1   | not assigned. no ontology                         | 157        |
| 27.3   | RNA. regulation of transcription                  | 151        |        | not assigned. no ontology. pentatricopeptide (PPR) repeat-containing protein | 119        |
| 29     | protein                                           | 120        | 27     | RNA                                               | 85         |
| 29.5   | protein. degradation                              | 77         | 27.3   | RNA. regulation of transcription                  | 62         |
| 29.5.11| protein. degradation. ubiquitin                   | 60         | 30     | signaling                                         | 52         |
| 30     | signaling                                         | 48         | 20     | stress                                            | 31         |
| 29.5.11| protein. degradation. ubiquitin                   | 45         | 29.5   | protein. degradation                              | 30         |
| 20     | stress                                            | 43         |        | 29.5.11 protein. degradation. ubiquitin            | 25         |

3.6. ‘RNA’ Related UI-Ts

In this study, a total of 230 ‘RNA’ related UI-Ts, including 159 5UI-Ts and 85 3UI-Ts, were identified. The ‘RNA’ pathway of Mapman consisted of ‘RNA. processing’, ‘RNA. transcription’, ‘RNA. regulation of transcription’ and ‘RNA. RNA binding’ (Table 7), among which ‘RNA. regulation of transcription’ and ‘RNA. RNA binding’ were significantly enriched pathways for UI-Ts as mentioned above (Table 5). In total, 199 UI-Ts (13.51% of all UI-Ts), including 137 5UI-Ts (9.30% of all UI-Ts) and 48 3UI-Ts (3.26% of all UI-Ts), were ‘RNA. regulation of transcription’-related. Most of these UI-Ts were transcription factor genes such as genes encoding bHLH, SET-domain transcriptional regulator, TCP, C2H2, GRAS, Trihelix and AP2 transcription factors, and so on. The significantly enriched ‘RNA. RNA binding’-related UI-Ts were all 3UI-Ts, including two genes encoding UBP1-associated protein 2A (UBA2A) (sb20247.1 and sb20247.2), two genes encoding U12-type spliceosomal protein U11/U12-31K (sb31320.1 and sb31320.3), five genes encoding D111/G-patch domain-containing protein (sb32422.1, sb32422.2, sb32422.3, sb32422.4 and sb32422.5) and two genes encoding RNA-binding (RRM/RBD/RNP motifs) family protein (sb37383.1 and sb37383.2).
Table 7. ‘RNA’ pathway related UI-Ts identified in this study. 5UI-Ts represent transcripts containing only 5’UI; 3UI-Ts represent transcripts containing only 3’UIs; and 5+3UI-Ts represent transcripts containing both 5’UI and 3’UI.

| BIN Pathway Name | 5UI-Ts | 3UI-Ts | 5+3UI-Ts |
|-----------------|--------|--------|----------|
| 27.3.99 RNA. regulation of transcription. unclassified | 30 | 12 | 2 |
| 27.3.6 RNA. regulation of transcription. bHLH, basic Helix-Loop-Helix family | 15 | 4 | 1 |
| 27.3.69 RNA. regulation of transcription. SET-domain transcriptional regulator family | 12 | 8 | 3 |
| 27.3.29 RNA. regulation of transcription. TCP transcription factor family | 11 | 6 | 2 |
| 27.3.11 RNA. regulation of transcription. C2H2 zinc finger family | 11 | 6 | 2 |
| 27.3.21 RNA. regulation of transcription. GRAS transcription factor family | 10 | 3 | 1 |
| 27.4 RNA. RNA binding | 0 | 11 | 0 |
| 27.3.67 RNA. regulation of transcription. putative transcription regulator | 9 | 3 | 2 |
| 27.3.30 RNA. regulation of transcription. Trihelix, Triple-Helix transcription factor family | 4 | 6 | 0 |
| 27.3 RNA. regulation of transcription. AP2/EREBP, APETALA2/Ethylene-responsive element binding protein family | 5 | 3 | 0 |
| 27.3.12 RNA. regulation of transcription. C3H zinc finger family | 6 | 1 | 1 |
| 27.1 RNA. processing | 4 | 3 | 0 |
| 27.1.1 RNA. processing. splicing | 0 | 5 | 0 |
| 27.3.25 RNA. regulation of transcription. MYB domain transcription factor family | 5 | 0 | 0 |
| 27.3.37 RNA. regulation of transcription. AS2, Lateral Organ Boundaries Gene Family | 5 | 0 | 0 |
| 27.3.16 RNA. regulation of transcription. CCAAT box binding factor family, HAP5 | 0 | 4 | 0 |
| 27.3.19 RNA. regulation of transcription. EIN3-like (EIL) transcription factor family | 4 | 0 | 0 |
| 27.3.68 RNA. regulation of transcription. PWWP domain protein | 4 | 1 | 1 |
| 27.3.8 RNA. regulation of transcription. CC2 (Zn) DOF zinc finger family | 3 | 1 | 0 |
| 27.2 RNA. transcription | 3 | 0 | 0 |
| 27.1.2 RNA. processing, RNA helicase | 0 | 3 | 0 |
| 27.3.1 RNA. regulation of transcription. ABID/VP1-related B3-domain-containing transcription factor family | 3 | 0 | 0 |
| 27.3.20 RNA. regulation of transcription. C2-like transcription factor family, GARP | 3 | 0 | 0 |
| 27.3.24 RNA. regulation of transcription. MADS box transcription factor family | 3 | 0 | 0 |
| 27.3.52 RNA. regulation of transcription. Global transcription factor group | 3 | 0 | 0 |
| 27.3.80 RNA. regulation of transcription. zf-HD | 3 | 0 | 0 |
| 27.1.19 RNA. processing, ribonucleases | 1 | 1 | 1 |
| 27.3.35 RNA. regulation of transcription. bZIP transcription factor family | 2 | 0 | 0 |
| 27.3.15 RNA. regulation of transcription. CCAAT box binding factor family, HAP3 | 0 | 1 | 0 |
| 27.3.26 RNA. regulation of transcription. MYB-related transcription factor family | 0 | 1 | 0 |
| 27.3.49 RNA. regulation of transcription. C2C2 (Zn) DOF zinc finger family | 0 | 1 | 0 |
| 27.3.73 RNA. regulation of transcription. Zinc-finger (CCCH) | 1 | 0 | 0 |
| 27.3.75 RNA. regulation of transcription. GRP | 0 | 1 | 0 |
| 27.3.84 RNA. regulation of transcription. BBR/BPC | 1 | 0 | 0 |
| 27.3.89 RNA. regulation of transcription. ovate family OFP | 1 | 0 | 0 |

4. Discussion

In the present study, based on the A. buxifolia genome data, we identified the introns existing in the UTR and CDS regions. Similar to sweet orange [42], more than 70% of A. buxifolia CDSs were found to contain introns. Unlike the CDSs, few UTRs were found to own intron. Only 3.68% of 5’UTRs and 2.77% of 3’UTRs were intron-containing. This might well explain why UTR introns are often neglected. Bioinformatic analysis of UIs and UI-containing transcripts (UI-Ts) was then performed. Additionally, the results obtained in this study were shown as follows.

4.1. The Lengths of A. buxifolia UIs Were Less Conserved Than CIs, and Most UI-Ts Contain Only One UI

Although the density of UIs was significantly lower than that of CIs, their mean and median intron sizes were ~2 times those of CIs. Similar to Arabidopsis and C. sinensis [39,42], the frequency of A. buxifolia CIs with 100–300 nucleotides was significantly higher than that of 5UIs and 3UIs, but the relative frequencies of short introns <50 nucleotides and long introns >300 nucleotides in 5’UTRs and 3’UTRs were higher than those in CDSs, indicating that the UI lengths were less conserved. Moreover, in accordance with previous studies [16,38,41], we found that most 5UI-Ts and 3UI-Ts contain only one UI.

4.2. A/T-Rich Elements around Both Donor Sites and Receptor Sites of A. buxifolia UTRs Are Important for U1 Recognition and Removal

Accumulated evidence showed that there were many splicing signals and factors influencing intron removal and mRNA transcription [51]. Among these factors, splice site pairs greatly influenced the effectivity of recruiting splicing machinery [52]. Consistent with previous studies [40,42], the 5’ donor sites and 3’ acceptor sites in A. buxifolia UTRs were...
also very conserved, and GT-AG and GC-AG were the two major splice site pairs for both A. buxifolia 5'UTRs and 3'UTRs. Moreover, it has been suspected that the A/T-rich elements play a role in intron recognition [42,53–57]. In the present study, an A/T-rich element was found around both donor sites and receptor sites of A. buxifolia UTRs, indicating that they might function in UI recognition and removal.

4.3. A. buxifolia 5UIs Were Rich of Gene-Expression-Enhancement-Related Elements and TFBSs, Indicating That They Might Contribute Greatly to Gene Expression Regulation

Currently, evidence has shown that UTR introns, especially the 5UIs, contribute greatly to gene expression regulation [58–60]. Some 5UIs and other 5'-proximal introns with cis-elements were proved to have the ability of enhancing gene transcription [6,61,62]. Kamo et al. [24] demonstrated that the 5UI of GUBQ1 acted as the promoter core sequence to increase GUS translation efficiency in transgenic plants. Lu et al. [58] and Samadder et al. [63] reported that the 5UI of rice rubi3 gene could improve the gene’s expression at both transcriptional and post transcriptional levels. In our present study, the A. buxifolia 5UIs were found to be rich of gene-expression-enhancement-related elements (such as ‘core promoter element around −30 of transcription start’ and ‘common cis-acting element in promoter and enhancer regions’ elements), indicating that these 5UIs may play roles in enhancing the expression of their corresponding genes.

Cenik et al. [41] reported the particular enrichment of 5UIs in genes with regulatory roles. The eukaryotic genes’ transcription is regulated by transcription factor (TFs), which could interact specifically with sequences in the promoter regions of the genes they regulate [53,54]. Consistently, regulatory genes tend to have more transcription factor binding sites (TFBSs) in their 5UIs [21]. In our present study, we found that the A. buxifolia 5UIs contain many TFBSs for BBR-BPC, MIKC_MADS, AP2 and Dof TFs. Functional analyses have revealed the indispensable role of BBR/BPC proteins in the gene expression control of TF genes [55–57]. Several Dof proteins have also been proved to contribute to gene expression activation by interacting with some other regulatory proteins [64,65]. The enrichment of these TFBSs in A. buxifolia 5UIs indicated that they play roles in the expression regulation of their corresponding downstream genes.

4.4. Many UI-Ts Are Involved in RNA Metabolism

Pathway-enrichment analysis revealed that UI-Ts involved in ‘RNA’ pathways were significantly enriched. Moreover, these ‘RNA’-pathway-related UI-Ts are more inclined to contain 5UI, which supported the findings of Cenik et al. [41] that genes with regulatory roles were more prone to possess 5UI. In this study, 40.00% (12/30) of the ‘RNA. regulation of transcription. unclassified’-pathway-related 5UI-Ts were genes encoding A20/AN1-like zinc finger (Znf) family proteins. Plant A20/AN1-ZnF family proteins were implicated in the plant responses to various abiotic and biotic stresses [66], suggesting that UIs might function in A. buxifolia stress responses. The bHLH TFs contributed greatly to regulating multiple plant cellular and biological processes [67]. In this study, genes encoding bHLH transcription factors were also enriched by 5UI-Ts. In addition, the ‘RNA binding’ pathway was significantly enriched by 3UI-Ts. The correlations between 3UI-Ts and RNA binding have been proved in previous studies. The expression of 3UI-Ts was more likely to be inhibited by NMD than gene transcripts with no 3UI, and the most significantly enriched NMD-affected gene transcripts were those encoding RNA-binding proteins [36,68]. Eukaryotes RNA-binding proteins (RBPs) play crucial roles in almost all aspects of post-transcriptional gene expression regulation [69]. In this study, two UBP1-associated protein 2A (UBA2A) genes were identified to be 3UI-Ts. It was reported that UBA2A could bind to RNA molecules containing U-rich sequences in 3'UTRs and might make contributions to the stabilization of mRNAs in the nucleus [70]. Thus, it is hypothesized that the 3UIs of the two UBA2As contribute to the mRNA stabilization status in A. buxifolia.
4.5. Most A. buxifolia 3UI-Ts Are Members of PPRP Gene Family, and Many UI-Ts Are Stress-Response-Related or with Unknown Function

Noteworthily, 119 (85.61%) of the 3UI-Ts belong to genes encoding pentatricopeptide (PPR) repeat-containing proteins. In Arabidopsis, rice and *Populus trichocarpa*, 441, 477 and 626 PPRP gene members were identified, respectively [50,71,72]. PPRP gene family widely exists in plants, especially terrestrial plants, and plays a crucial role in plant growth and development, and stress response processes [73–75]. Usually, the PPRP genes contain no intron in their CDSs [76,77]. In our present study, we also found that all the UI-containing PPRPs contained no intron in their CDSs, indicating that alternative splicing events of PPRPs mainly occurred in UTRs, and UIs contributed greatly to the tissue- or organ-specific expression and the regulatory functions of PPRPs.

Stress can affect the efficiency or patterns of splicing and intron retention for stabilizing the transcripts or serving to modify its biological functions [78–80]. The 3UI-Ts of sweet orange were significantly enriched in stress pathway [42]. In this study, we also identified many UI-Ts related to ‘stress’ and ‘signaling’, suggesting that UIs may play an extremely important role in plant defense responses [42,78]. Additionally, we found many UI-Ts, which were categorized into the ‘unclassified’ pathway. The functions of these UIs and their corresponding UI-Ts need to be further studied.

5. Conclusions

In this study, we performed a genome-scale computational analysis of UIs in *A. buxifolia*, investigated their size and nucleotide distribution characteristics, explored the regulatory cis-elements and TFBSs in the 5UI sequences, and explicated the possible functions of UI-Ts. In total, we identified 1077 5UIs and 866 3UIs from the *A. buxifolia* genome data. The density of 5UIs and 3 UIs was lower than that of CIs, but they were twice as big as CIs. The 5′ donor sites and 3′ acceptor sites in *A. buxifolia* UTRs were very conserved, and GT-AG was the mostly commonly splice site pair for both 5′UTRs and 3′UTRs. A/T-rich elements functioning in intron recognition and removal were found around both donor sites and receptor sites of *A. buxifolia* UTRs. Most UI-Ts contained one 5UI or 3UI. Many gene-expression-enhancement-related elements and TFBSs were discovered in the *A. buxifolia* 5UIs, indicating that 5UIs play a role in regulating or enhancing the expression of their corresponding downstream genes. Consistently, pathway-enrichment analysis revealed that UI-Ts involved in ‘RNA’ pathways were significantly enriched. Notably, more than 85% of the 3UI-Ts were genes encoding pentatricopeptide (PPR) repeat-containing proteins. The results obtained in this study are of great significance for further understanding the regulation of UIs in *A. buxifolia* gene expression.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/horticulturae7120556/s1. Additional file Table S1: Information of the 5UIs identified in the Atalantia buxifolia genome and the gene transcripts containing them; Additional file Table S2: Information of the 3UIs identified in the Atalantia buxifolia genome and the gene transcripts containing them; Additional file Table S3: Information of the identified cis-acting elements in 5UI sequences; Additional file Table S4: TFBS prediction results of regulatory 5UI-Ts of transcription factor.

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

Funding: This research was funded by the Fund for high-level talents of Shanxi Agricultural University (2021XG010) and the Construction of Plateau Discipline of Fujian Province (102/71201801104).

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 and Chengdu Life Baseline Company for their assistance in bioinformatics analysis.

Conflicts of Interest: The authors declare that they have no competing interest.

References
1. Roy, S.W.; Gilbert, W. The evolution of spliceosomal introns: Patterns, puzzles and progress. Nat. Rev. Genet. 2006, 7, 211–221.
2. Sambrook, J. Adenovirus amazes at Cold Spring Harbor. Nature 1977, 268, 101–104. [CrossRef]
3. Cannone, J.J.; Subramanian, S.; Schnare, M.N.; Collett, J.R.; D’Souza, L.M.; Du, Y.; Feng, B.; Lin, N.; Madabusi, L.V.; Müller, K.M.; et al. The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinform. 2002, 3, 2.
4. Wahl, M.C.; Will, C.L.; Lührmann, R. The spliceosome: Design principles of a dynamic RNP machine. Cell 2009, 136, 701–718. [CrossRef]
5. Le Hir, H.; Nott, A.; Moore, M.J. How introns influence and enhance eukaryotic gene expression. Trends Biochem. Sci. 2003, 28, 215–220. [CrossRef]
6. Moore, M.J.; Proudfoot, N.J. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 2009, 136, 688–700. [CrossRef]
7. Chorev, M.; Carmel, L. The function of introns. Front. Plant Sci. 2012, 3, 55. [CrossRef]
8. Clark, A.J.; Archibald, A.L.; McLenaghan, M.; Simons, J.P.; Wallace, R.; Whitelaw, C.B. Enhancing the efficiency of transgene expression. Philos. Trans. R. Soc. Lond. B Biol. Sci. 1993, 339, 225–232.
9. Wahl, M.C.; Will, C.L.; Lührmann, R. The spliceosome: Design principles of a dynamic RNP machine. Cell 2009, 136, 701–718. [CrossRef]
10. Mascarenhas, D.; Mettler, I.J.; Pierce, D.A.; Lowe, H.W. Intron-mediated enhancement of heterologous gene expression in maize. Plant Mol. Biol. 1990, 15, 913–920. [CrossRef] [PubMed]
11. Callis, J.; Fromm, M.; Walbot, V. Introns increase gene expression in cultured maize cells. Genes Dev. 1987, 1, 1183–1200. [CrossRef]
12. 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]
13. Gidekel, M.; Jimenez, B.; Herrera-Estrella, L. The first intron of the Arabidopsis thaliana gene with the following intron 1 enhances reporter gene expression up to 1000-fold. Plant Mol. Biol. 1991, 16, 199–207. [CrossRef]
14. Dean, C.; Favreau, M.; Bond-Nutter, D.; Bedbrook, J.; Dunsmuir, P. Sequences downstream of translation start regulate quantitative expression of two petunia rbcS genes. Plant Cell 1989, 1, 201–208. [PubMed]
15. 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] [PubMed]
16. Clark, A.J.; Archibald, A.L.; McLenaghan, M.; Simons, J.P.; Wallace, R.; Whitelaw, C.B. Enhancing the efficiency of transgene expression. Philos. Trans. R. Soc. Lond. B Biol. Sci. 1993, 339, 225–232.
17. Majewski, J.; Ott, J. Distribution and characterization of regulatory elements in the human genome. Genome Res. 2002, 12, 1827–1836. [CrossRef]
18. 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]
19. Matsumoto, K.; Wassarman, K.M.; Wolfe, A.P. Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. EMBO J. 1998, 17, 2107–2121. [CrossRef] [PubMed]
20. Grant, T.N.L.; 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]
21. Laxa, M.; Müller, K.; Lange, N.; Doering, L.; Pruscha, J.T.; Peterhänsel, C. The 5′UTR Intron of Arabidopsis GGT1 Aminotransferase Enhances Promoter Activity by Recruiting RNA Polymerase II. Plant Physiol. 2016, 172, 313–327. [CrossRef]
22. Weischenfeldt, J.; Damgaard, I.; Bryder, D.; Theilgaard-Mönch, K.; Thoren, L.A.; Nielsen, F.C.; Jacobsen, S.E.W.; 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]
23. Ni, J.Z.; Grate, L.; Donohue, J.P.; Preston, C.; Nobida, N.; O’Brien, G.; Shiue, L.; Clark, T.A.; Blume, J.E.; Ares, M., Jr. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. 2007, 21, 708–718. [CrossRef] [PubMed]
28. Chang, Y.F.; Imam, J.S.; Wilkinson, M.F. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* 2007, 76, 51–74. [CrossRef] [PubMed]

29. Mendell, J.T.; Sharriff, N.A.; Meyers, J.L.; Martinez-Murilla, F.; Dietz, H.C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* 2004, 36, 1073–1078. [CrossRef] [PubMed]

30. 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 RNA translation in the cytoplasm. *RNA* 1998, 4, 801–815. [CrossRef]

31. Yepiskoposyan, H.; Aeschimann, F.; Nilsson, D.; Okoniewski, M.; Mühlemann, O. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* 2011, 17, 2108–2118. [CrossRef] [PubMed]

32. Rüegsegger, U.; Leber, J.H.; Walter, P. Block of translation upon induction of the unfolded protein response. *Cell* 2001, 107, 103–114. [CrossRef]

33. Bruno, I.G.; Karam, R.; Huang, L.; Bhardwaj, A.; Lou, C.H.; Shum, E.Y.; Song, H.W.; Corbett, M.A.; Gifford, W.D.; Geicz, J.; et al. Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. *Mol. Cell* 2011, 42, 500–510. [CrossRef] [PubMed]

34. McIlwain, D.R.; Pan, Q.; Reilly, P.T.; Elia, A.J.; McCracken, S.; Wakeham, A.C.; Itie-Youten, A.; Blencowe, B.J.; Mak, T.W. Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. *Proc. Natl. Acad. Sci. USA* 2010, 107, 12186–12191. [CrossRef]

35. Wittmann, J.; Hol, E.M.; Jäck, H.M. hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol. Cell* 2006, 26, 1272–1287. [CrossRef] [PubMed]

36. 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]

37. Pan, Q.; Saltzman, A.L.; Yoon, K.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]

38. Hong, X.; Scafidi, D.G.; Lynch, M. Intron size, abundance, and distribution within untranslated regions of genes. *Mol. Biol. Evol.* 2006, 23, 2392–2404. [CrossRef]

39. Chung, B.Y.W.; 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] [PubMed]

40. Roy, S.W.; Penny, D.; Neafsey, D.E. Evolutionary conservation of UTR intron boundaries in *Cryptococcus*. *Mol. Biol. Evol.* 2007, 24, 1140–1148. [CrossRef]

41. Cenik, C.; Derti, A.; Mellor, J.C.; Bertz, G.F.; Roth, F.P. Genome-wide functional analysis of human 5′UTR introns on gene expression in *Atalantia buxifolia*. *Genome Biol.* 2010, 11, R29. [CrossRef] [PubMed]

42. 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] [PubMed]

43. Liang, H.X.; Sun, J.J.; Shen, Z.B.; Yu, B.W.; Cui, H.H.; Yin, Y.Q. A novel alkaloid glycoside isolated from *Atalantia buxifolia*. *Nat. Prod. Res.* 2020, 34, 3042–3047. [CrossRef]

44. Chang, F.R.; Li, P.S.; Huang Liu, R.; Hu, H.C.; Hwang, T.L.; Lee, J.C.; Chen, S.L.; Wu, Y.C.; Cheng, Y.B. Bioactive Phenolic Components from the Twigs of *Atalantia buxifolia*. *J. Nat. Prod.* 2018, 81, 1534–1539. [CrossRef]

45. Yang, Y.Y.; Yang, W.; Zuo, W.J.; Zeng, Y.B.; Liu, S.B.; Mei, W.L.; Dai, H.F. Two new acridone alkaloids from the branch of *Atalantia buxifolia* and their biological activity. *J. Asian Nat. Prod. Res.* 2013, 15, 899–904. [CrossRef] [PubMed]

46. 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] [PubMed]

47. Thimm, O.; Bläsing, O.; Gibon, Y.; Nagel, A.; Meyer, S.; Krüger, P.; Selbig, J.; Müller, 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]

48. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van De Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002, 30, 325–327. [CrossRef]

49. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* 2004, 14, 1188–1190. [CrossRef] [PubMed]

50. Lin, C.; Andrés, C.; Aubourg, S.; Bellalou, M.; Bitton, F.; Bruyère, C.; Caboche, M.; Debast, C.; Gualberto, J.; Hoffmann, B.; et al. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 2004, 16, 2089–2103. [CrossRef] [PubMed]

51. Brown, J.W.; Simpson, C.G.; Thow, G.; Clark, G.P.; Jennings, S.N.; Medina-Escobar, N.; Haupt, S.; Chapman, S.C.; Opara, K.J. Splicing signals and factors in plant intron removal. *Biochem. Soc. Trans.* 2002, 30, 146–149. [CrossRef] [PubMed]

52. 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]

53. Rois, R.; West, S.M.; Sosinsky, A.; Liu, P.; Mann, R.S.; Honig, B. The role of DNA shape in protein-DNA recognition. *Nature* 2009, 461, 1248–1253. [CrossRef]
