

Article

Global Profiling of lncRNAs Expression Responsive to Allopolyploidization in *Cucumis*

Panqiao Wang [†] , Xiaqing Yu [†] , Zaobing Zhu, Yufei Zhai, Qinzheng Zhao, Ya Meng, Ji Li, Qunfeng Lou and Jinfeng Chen ^{*} 

State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China; 2016204024@njau.edu.cn (P.W.); xqyu@njau.edu.cn (X.Y.); 2016104051@njau.edu.cn (Z.Z.); zhaiyufei2012@163.com (Y.Z.); 2017204024@njau.edu.cn (Q.Z.); 2018104046@njau.edu.cn (Y.M.); liji1981@njau.edu.cn (J.L.); qflou@njau.edu.cn (Q.L.)

* Correspondence: jfchen@njau.edu.cn; Tel.: +86-25-84396279

† These authors contributed equally to this work.

Received: 9 October 2020; Accepted: 10 December 2020; Published: 12 December 2020



Abstract: Long non-coding RNAs (lncRNAs) play critical regulatory roles in various biological processes. However, the presence of lncRNAs and how they function in plant polyploidy are still largely unknown. Hence, we examined the profile of lncRNAs in a nascent allotetraploid *Cucumis hytivous* (S₁₄), its diploid parents, and the F₁ hybrid, to reveal the function of lncRNAs in plant-interspecific hybridization and whole genome duplication. Results showed that 2206 lncRNAs evenly transcribed from all 19 chromosomes were identified in *C. hytivous*, 44.6% of which were from intergenic regions. Based on the expression trend in allopolyploidization, we found that a high proportion of lncRNAs (94.6%) showed up-regulated expression to varying degrees following hybridization. However, few lncRNAs (33, 2.1%) were non-additively expressed after genome duplication, suggesting the significant effect of hybridization on lncRNAs, rather than genome duplication. Furthermore, 253 cis-regulated target genes were predicted for these differentially expressed lncRNAs in S₁₄, which mainly participated in chloroplast biological regulation (e.g., chlorophyll synthesis and light harvesting system). Overall, this study provides new insight into the function of lncRNAs during the processes of hybridization and polyploidization in plant evolution.

Keywords: lncRNAs; allopolyploid; *Cucumis*; transcriptome; interspecific hybridization

1. Introduction

Long noncoding RNAs (lncRNAs) are transcripts of more than 200 nucleotides in length that have no capacity to code for proteins [1]. lncRNAs typically originate from the intergenic regions of the genome and the antisense strands of coding genes [2]. Some lncRNAs can also be transcribed from the sense strands or intronic regions of coding genes [3]. In the past decade, thousands of lncRNAs have been widely identified in both animal and plant genomes because of rapid progress in high-throughput sequencing technologies [4–9]. In animals, increasing evidence supports that lncRNAs play critical roles in epigenetic regulation [10], for example, X chromosome inactivation [11,12], disease occurrence [13], genomic imprinting [14], developmental regulation [15], etc. In plants, lncRNAs are involved in many biological processes, including flowering time regulation [16,17], biotic and abiotic stimuli responses [18–21], photomorphogenesis [22,23], and regulation in other developmental processes [24,25]. In contrast to extensive studies of lncRNAs elucidated in mammals to date, the functions of plant lncRNAs are still obscure, especially in the process of polyploidization.

Interspecific hybridization and polyploidization are important in genome evolution [26]. Because the vigor traits created by hybridization are well stabilized during whole genome duplication (WGD), allopolyploids show their evolutionary advantages in development and adaptation [27–29]. The genomic interactions of two divergent subgenomes in a single nucleus trigger extensive and rapid genome restructuring, along with changes in DNA methylation, histone modification, and chromosomal rearrangements [30–32]. The combined results of these genomic and epigenetic changes lead to altered expression of about 20–50% of the mRNAs, providing a molecular basis for the viability of hybridization and polyploidization [3,33]. In nascent allohexaploid wheat, a high proportion of microRNAs showed nonadditive expression upon polyploidization [3]. However, the impact of hybridization and polyploidization on the other important types of noncoding RNAs are rarely reported [34].

The allotetraploid *Cucumis hytivus* (HHCC, $2n = 4x = 38$) was synthesized through chromosome doubling of a hybrid between cultivated cucumber (*C. sativus*, CC, $2n = 2x = 14$) and wild species *C. hystrix* (HH, $2n = 2x = 24$) [35–37]. According to the molecular evidence from cytoplasmic and nuclear genomes, *C. hystrix* is classified as the closest wild relative of cucumber, diverging approximately 4.6 million years ago [38]. This *Cucumis* allotetraploid with a clear genetic background could facilitate the study of genetic, functional, and epigenetic changes during hybridization and genome duplication. Thus, transcriptomic analysis of synthetic allotetraploid *Cucumis* can shed light on the expression changes of transcripts following allopolyploidization. Previously, cDNA-AFLP and reverse Northern blot studies in early generations of *C. hytivus* (S_1 – S_2) showed rapid activation or silencing of gene expression [39]. Recently, we have completed the high-quality assembly of *C. hytivus* genome, which make it possible to identify and characterize lncRNAs at whole-genome level in this newly synthetic allopolyploid specie in *Cucumis*. In this study, high-throughput RNA-seq was performed for genome-wide identification and characterization of lncRNAs in *C. hytivus*. To further investigate the potential roles of lncRNAs in interspecific hybridization and whole genome duplication in *Cucumis*, we examined the profile of lncRNAs in nascent allotetraploid *C. hytivus*, its progenitors and interspecific F_1 hybrid. Study on the expression patterns of lncRNAs and their target genes will help us to better understand the epigenetic regulation mode of lncRNAs in allopolyploidization, and provide a molecular theoretical basis for distant hybrid breeding.

2. Materials and Methods

2.1. Plant Materials and Growth Conditions

The following species were employed: cultivated cucumber *Cucumis sativus* L. “BeijingJietou” (*C. sativus*, $2n = 14$, genome CC), wild species *C. hystrix* ($2n = 24$, genome HH), and their interspecific F_1 homoploid hybrid and synthetic allotetraploid *C. hytivus* ($2n = 4x = 38$, genome HHCC). The highly inbred synthetic allotetraploid *C. hytivus* (S_{14}) were obtained by self-pollinating 14 generations from the original chromosome doubling of interspecific F_1 (designated as S_0) (Figure S1a,b). All planting materials were supplied from the state key laboratory of Cucurbit Genetics and Germplasm Enhancement of Nanjing Agricultural College. Seeds of five samples were sown on plastic pots and grown under controlled conditions (14 h light and temperature range of 22–28°C), with a relative humidity of 70–80%.

2.2. RNA Isolation, cDNA Library Preparation, and Illumina Sequencing

The first young leaf of each seedling with three biological replicates were collected at ten-leaf stage, immediately frozen in liquid nitrogen, and then stored at -80°C for further use. The extraction of total RNA and synthesis of first-strand cDNA were performed using Trizol reagent (TaKaRa Bio Inc., Kusat su, Shiga, Japan) and the PrimeScript™ RTPCR kit (TaKaRa) according to the manufacturer’s instructions, respectively.

After RNA extraction, the purity, concentration, and integrity of RNA were tested using NanoDrop 2000 (Thermo Fisher Scientific, Wilmington, DE, USA). The poly(A)⁺-type cDNA libraries for the

leaf samples collected from different species were constructed and subjected to high-throughput sequencing provided by 1 Gene Ltd. (Hangzhou, China). The libraries were sequenced on Illumina HiSeq™ 2500/MiSeq platform with 150-bp paired-end reads. The datasets generated during the current study are available in the National Center for Biotechnology Information (NCBI) under the bioproject accession number PRJNA658669.

2.3. Transcriptome Assembling and Identification of lncRNAs

Raw reads obtained from RNA-Seq were pre-processed. The 5' and 3' adapters were trimmed; low-quality (Phred Score ≤ 20) and shorter reads (length ≤ 30 bp) were removed by FASTX-toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). The clean reads were mapped to the reference genome (S₁₄ genome assembled in our lab using SMRT sequencing technology, Cucumis Genome Database: <http://www.cucumisgdb.cn/>) using Hisat2 v2.1.0 (-p 20 -dta) [40], and were assembled using StringTie (v1.3.4d) [41]. The length of all transcripts should be no less than 200 bp. The assembled transcripts were annotated using the Cuffcompare program from the Cufflinks package (v2.2.1) (<http://cufflinks.cbc.umd.edu/>). According to Cuffcompare, only transcripts class-coded as “u” (unknown intergenic transcript), “i” (a transfrag falling entirely within a reference intron), and “x” (exonic overlap with reference on the opposite strand) were selected as novel transcripts including long non-coding RNA (lncRNA).

Then, the candidate lncRNA was identified based on the coding potentials score, which was calculated by using the coding potential calculator (CPC2) [42], PLEK [43], CPAT [44], and CNCI [45]. The intersection with CPC2 score < 0 , PLEK score < 0 , CNCI score < 0 , and CPAT score < 0.36 were the initial predicted lncRNAs [42–45]. The Venn diagram of the lncRNA list from each software were prepared using the function `vennDiagram` in R. The predicted lncRNAs with at least two exons were used for the next analysis. The lncRNAs were classified into several categories based on their genomic location, according to description in previous studies [46]. RepeatModeler v1.0.11 (-LTRstruct -engine ncbi -pa 18) (<http://www.repeatmasker.org/RepeatModeler.html>) was used to construct de-novo transposable element (TE) libraries based on the *C. hytivus* reference genome using default parameters. The initial repeat database was then merged with the known Repbase database (20170127), which was used to distinguish the *C. hytivus* TEs by using RepeatMasker v4.0.9 (-parallel 18 -engine ncbi -nolow -no_is -norna -gff) (<http://www.repeatmasker.org>). TE-derived lncRNAs were identified by the determining overlapping genomic coordinates of TEs or TE fragments of at least 10 bp using the intersect program from BEDTools v2.17.0 (<https://github.com/arq5x/bedtools2>) with reference to previous studies [34,47]. The comparative analysis of syntenic orthologous gene and lncRNA pairs were performed using blast (blastn -evalue 1e-5 -max_target_seq 1) and MCScanX (-b 2, -s 4) [48]. The co-linear pairs were determined by one to one optimal matching results from BLASTn. The longest transcripts (mRNA) of 46,845 protein coding genes (PCGs) annotated in the *C. hytivus* reference genome were used to comparing sequence characteristics of lncRNA. The sequence characteristics of the lncRNAs were visualized in ggplot2 R-package (v3.1.0) and Circos v0.69-3 [49].

2.4. Differential Expression Analysis

Expression levels of genes and lncRNAs were measured with normalized counts of reads by their respective lengths using kallisto v0.44.0 (quant -t 20 -b 100) [50]. Then we collapsed the expression of lncRNA into gene level by an in-house R script. Transcripts per million mapped reads (TPM) was applied to represent the normalized expression value. Heatmap of samples was generated using the R package pheatmap v1.0.10 (<https://cran.r-project.org>) and the Pearson correlation between biological replicates was calculated using the function `cor` in R based on the TPM values. The average values of the two parents' expression levels (reads count and TPM) were calculated as mid-parent value (MPV), which were used for subsequent analysis. The expression profiles of lncRNAs were clustered and visualized by Java program Short Time-series Expression Miner (STEM v1.3.11) [51] under the default parameters. DEseq2 R package (v1.20.0) was used to estimate variance-mean dependence in count

data from kallisto and test for differential expression based on a model using the negative binomial distribution [52]. Because lncRNAs exhibited more expression variance than PCGs, lncRNAs with a p -adjust < 0.0001 (Benjamini-Hochberg [BH] multiple test correction) and $\log_2(\text{fold change}) > 2$ or < -2 were set as the threshold of the significantly differential expression [34,53]. The species specific expression was classified as a type of differential expression.

2.5. Prediction of lncRNA Function

The *cis*-acting mode of lncRNA is the main regulating function on their neighboring genes [54,55]. Thus, the protein-coding genes in 100k upstream and downstream of lncRNA were searched and predicted as potential target genes (PTGs). A Pearson correlation of the expression of lncRNA and its target gene pair was calculated using the function `cor` in R. Protein-coding gene function was annotated using BLAST 2.6.0+ (`blastp -evalue 1e-6 -outfmt 7 -num_threads 24`) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), Gene Ontology (GO) classify (`blast2go.jar -in alignment_result.xml -a -v`) (<http://geneontology.org/>) and in-house R scripts based on the following databases: NCBI Non-Redundant Protein Sequences Database, Swiss-Prot for a manually annotated and reviewed protein sequence database and GO. GO enrichment analysis of the target genes for differentially expressed (DE) lncRNA was conducted by clusterProfiler R package (v3.16.0) [56].

2.6. qRT-PCR Analysis for DE lncRNAs and Their Targets

The relative quantity of lncRNAs and their targets were evaluated by Real-time fluorescent quantitative polymerase chain reaction (qRT-PCR). All primer sequences were designed by software Primer Premier 6.0 software and listed in Table S1. qRT-PCRs were performed with the SYBR Premix Ex Taq™ II kit (TaKaRa) and CFX96 real-time detection system (Bio-Rad Laboratories Inc., Hercules, CA, USA) according to the manufacturer's protocol. Three biological replicates and three technical replicates were assigned. The actin gene was used as an internal control for data normalization, and quantitative changes in different replicates were calculated using the delta-delta threshold cycle relative quantification method according to Yu et al. [57]. Significant differences were determined by Duncan's multiple range test at $P = 0.01$ level in SPSS. The fitting and visualization of the regression curve for qPCR and RNA-seq quantitative results were performed in R 3.4.1 (<http://www.r-project.org/>).

2.7. Photosynthetic Parameters Measurements

The net photosynthetic rates (P_n) and stomatal conductance (G_s) were measured using a Li-6400 portable photosynthesis assay apparatus (LI-COR Biosciences Inc., Lincoln, NE, USA). The leaf photo-synthetically active radiation (PAR) was controlled at 12 levels from 0 to 1500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (0, 25, 50, 75, 100, 150, 200, 300, 600, 900, 1200, and 1500). The maximum net photosynthesis rate under saturation light (P_{max}) was obtained by using Photosynthesis Work Bench (Li-Cor, Lincoln, NE, USA) according to the method of Yu et al. [58].

3. Results

3.1. Identification of the Putative lncRNAs in *C. hytivus* Leaves

To identify lncRNAs in *C. hytivus*, we established a pipeline (https://github.com/wangpanqiao/lncRNA_Pipeline) consisting of four main steps: transcript assembly, novel long RNA extraction, protein-coding capacity prediction, and filtering, as shown in Figure 1a. We first obtained 70,701,732 clean reads from the library in leaves of *C. hytivus* (S_{14}), with 93.15% of which were successfully mapped to reference genome. Then, mapped data were used to assemble transcripts by StringTie. In total, 70,180 assembled transcripts were obtained. Assembled annotation file gtf was compared with the reference to distinguish novel long RNAs by filtering transcripts location not overlapping with the known gene and excluding length smaller than 200 nt. As a result, 22,640 transcripts were obtained after extracting. Then, 2870 transcripts without coding ability were singled out from the intersection of

the CPC2, CNCI, CPAT, and PLEK results (Figure 1b). Finally, these non-coding potential transcripts were retained under two criteria: (1) exon number ≥ 2 ; (2) TPM > 0.3 , and were defined as credible lncRNA with total 2206 transcripts transcribed from 1593 unigenes (lncRNA-gene). CPAT was used to test the coding probability of final discriminated lncRNAs and protein coding transcripts, and almost all of the lncRNAs were less than 0.36 (Figure S2).

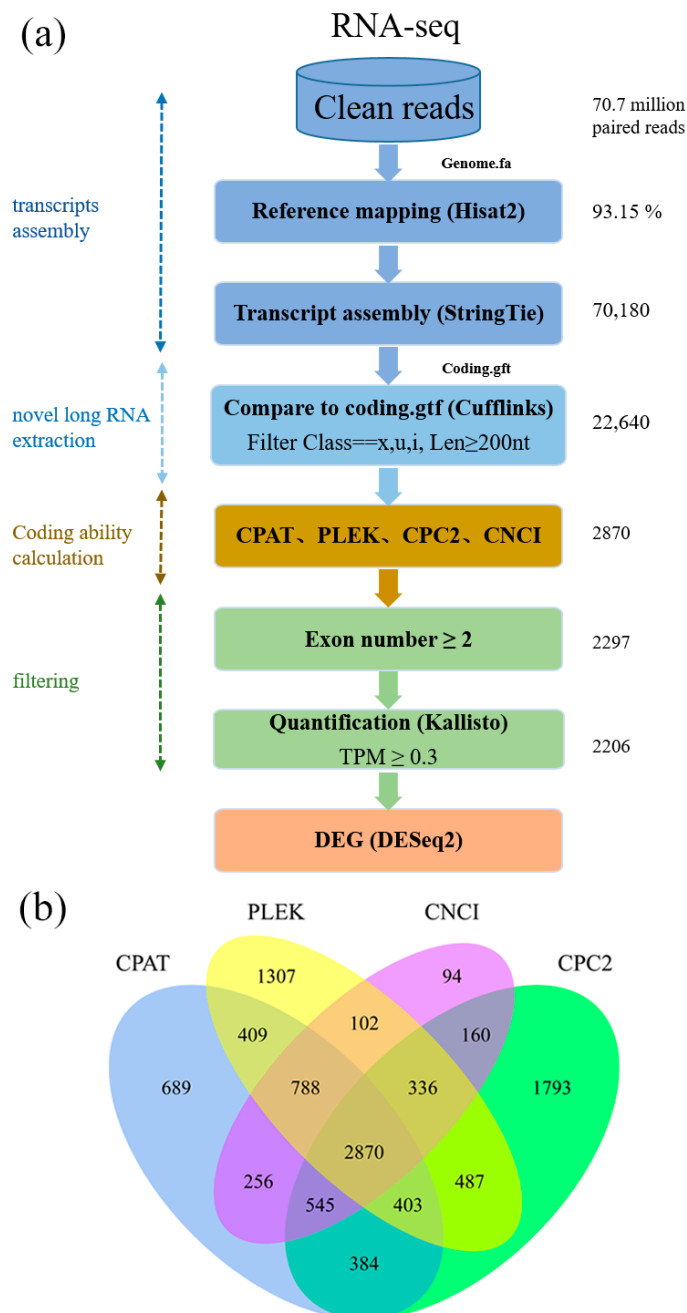


Figure 1. Complete computational pipeline for the systematic identification of long non-coding RNA (lncRNAs) in allotetraploid *Cucumis hytivus* leaves. (a) Detailed schematic diagram of the bioinformatic pipeline for the identification of leaf lncRNAs. (b) Venn diagram of the number of candidate lncRNAs filtered by CPC2, CNCI, CPAT, and PLEK. CPC2: coding potential calculator; CNCI: coding-noncoding index; CPAT: coding potential assessment tool; PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved *k-mer* scheme.

3.2. Sequence Characteristics of the lncRNAs

To gain a comprehensive understanding of the distribution and structure of lncRNAs in *C. hytivus*, we then investigated their exon numbers, transcript length, and overlapping with transposable elements (TEs). The lncRNAs that could be assigned to a chromosome location were more evenly distributed across the 19 chromosomes than TE and PCGs in *C. hytivus* with no obvious preferences of locations. The number of lncRNAs in two subgenomes were 1406 (subH) and 675 (subC) respectively (Figure 2a). The remaining 125 lncRNAs were located on the scaffold fragments. Between the two subgenomes, 145 pairs of lncRNA-genes showed good collinearity, which was consistent with the collinearity of the coding genes (Figure 2a; Figure S3a,b). Especially, H01 and C07 maintained good collinearity in two subgenomes. However, some lncRNAs from the chromosome C05 of the subC genome showed better collinearity with the lncRNAs on the chromosomes H12 and H04 of the subH genome, which was absent in the collinearity of the encoding genes in two subgenomes (Figure S3a,b).

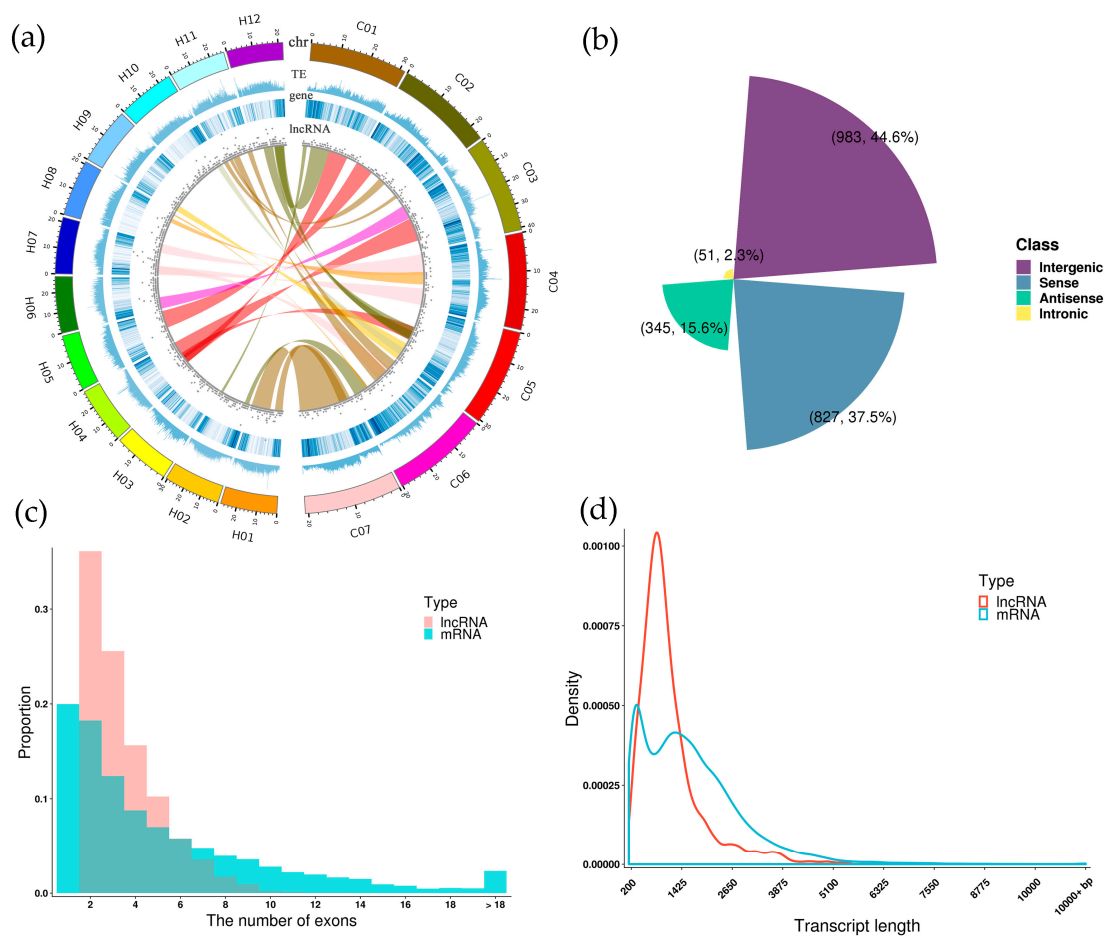


Figure 2. Description of the sequence characteristics of lncRNAs in *C. hytivus* leaves. **(a)** Analysis of the lncRNAs along each chromosome, showing density of genes, transposon elements, and lncRNAs in physical bins of 100 kb for each chromosome. The circles from inside to outside represent lncRNA, gene, transposon elements (TE), and chromosomes (chr). The internal collinear bands represent 145 lncRNAs in 23 synteny blocks. The scale for the chromosomes (outer bars) is Mb. **(b)** Numbers of intergenic, intronic, antisense, and sense lncRNAs in *C. hytivus* leaves. **(c)** Numbers of exons of lncRNAs in comparison with protein-coding transcripts of *C. hytivus*. **(d)** Comparison of the transcript length distributions of lncRNAs and mRNAs.

Among these lncRNAs, 983 intergenic (lincRNAs), 827 sense, 345 antisense, and 51 intronic lncRNAs were identified according to the locations of lncRNAs in the genome (Figure 2b). With respect to exon numbers, we found that the range of exon numbers in lncRNA transcript was more concentrated than mRNA, and most of the lncRNAs (61.7%) contained two or three exons (Figure 2c). Only a few intergenic lncRNAs (1.9%) and sense lncRNAs (1.5%) contained more than nine exons (Figure S4a). The mean transcript length (1016 bp) of the lncRNAs was obviously shorter than that of protein-coding genes (1490 bp). Most lncRNAs (84.1%) had length shorter than 1500 nucleotides (Figure 2d). The majority of the antisense lncRNAs (85.5%) were shorter than 1000 nucleotides (Figure S4b). When using ≥ 10 bp as an overlapping criterion, the proportion of lncRNAs overlapping with TEs (76.7%) was much more than that of mRNAs (51.2%) (Table S2) possibly because of the less conservativeness of ncRNAs than coding genes.

3.3. Profiling of lncRNA Expression in Interspecific Hybridization and Genome Duplication

To identify the dynamical model of lncRNA expression in three different stages (hybridization, duplication, and diploidization), we constructed four RNA-seq libraries from leaves of two diploid parents (CC, HH), their interspecific F_1 hybrid, and synthetic allotetraploid (S_{14}) for high-throughput sequencing (Figure S1a–c). From CC, HH, and F_1 , we respectively obtained 72,823,501, 72,036,788, and 70,755,534 clean reads. For interlibrary comparison, read numbers were normalized to relative abundance transcripts per million mapped base pairs (TPM) [50]. Gene expression correlations between three biological replicates were high, with average Pearson correlation coefficients (lncRNAs and mRNAs) of 0.99, 0.98, 0.98, and 0.98 for CC, F_1 , HH, and S_{14} , respectively (Figure S5a,b). In stark contrast to PCGs, we found that the expression level of lncRNAs (\log_2 TPM) exhibited much lower than mRNAs across the four samples (Figure S6; Table S3). The lncRNAs expression levels of hybrid F_1 and synthetic tetraploid were between two diploid ancestors (Table S3). To identify significant temporal expression profiles across allopolyploidization process, 1593 lncRNA-genes expression data from two progenitors, hybrid F_1 and S_{14} progeny were clustered in the Java program STEM. The expression of lncRNAs changed regularly along with allopolyploidization. Here, lncRNAs were classified into seven bins (as shown in Figure 3) according to their expression patterns among the mid-parent value (MPV) and the progeny (MPV- F_1 - S_{14}). We found that 1328 lncRNA-genes showed varying degrees of change during the allopolyploidization, of which 1256 (94.6%) were activated during interspecific hybridization. Following, there were 348 lncRNA-genes increased and 450 lncRNA-genes decreased during polyploidization, respectively (Figure 3; Table S4). After genome duplication, up-regulated lncRNA accounted for 66.1% (excluding expression profile 10 and 14). The expression profile 11 contained the most genes (530), which were up-regulated in F_1 and maintained the similar expression level in S_{14} . These results suggested that hybridization may have caused more lncRNAs of up-regulation.

3.4. Different Expression of lncRNA during *Cucumis* Allopolyploidization

To study the effect of allopolyploidization on lncRNA expression in the *Cucumis* synthetic allopolyploid, we first compared lncRNA expression levels between the two progenitors. Among 145 collinear lncRNA-genes expressed in leaves, 27.6% (40) differentially expressed ($p < 0.0001$) between *C. hystrix* and *C. sativus* parental lines (HH-CC_{dif}) (Figure 4a; Figure S7a). Furthermore, 19 (13.1%) were expressed at a higher level in the wild species (HH > CC), with the remaining more highly expressed in the cultivated cucumber (HH < CC, 14.5%, 21).

The differential expression of lncRNAs between the S_{14} and each parental line was then studied. There were 202 (18.8%) differentially expressed lncRNAs (DE-lncRNAs) between *C. hystrix* S_{14} and *C. hystrix* (S_{14} -HH_{dif}) which is more than comparison between S_{14} and *C. sativus* (S_{14} -CC_{dif}). In S_{14} -CC_{dif} lncRNAs, 55 lncRNA-genes expressed higher in S_{14} , accounting for 8.9% of the total expressed lncRNA-genes (617, subC). It was more than the ratio (2.3%, 14) of lncRNAs highly expressed in parental *C. sativus* (Figure 4a). Similar patterns were found for S_{14} -HH_{dif} (Figures 4a and S7c),

the number of higher expressed lncRNAs was smaller in progenitor HH (28; 2.6%) than S₁₄ (174; 16.2% of the total 1073 subH lncRNA-genes). However, the difference ratio between S₁₄ and female parent (HH; 18.8%) was larger than the male parent (CC; 12.2%), indicating that allotetraploid *Cucumis* is indeed more highly diverged from *C. sativus* than from *C. hystrix* in terms of lncRNAs transcriptome profiles. To determine the non-additively expressed lncRNAs, the average of parental lncRNAs expression levels (mid-parent value; MPV) were compared with nascent allotetraploid progeny S₁₄. Results showed that non-additively expressed lncRNAs-genes ($p < 0.0001$) only represented a small percentage (2.1%; 33) of the total lncRNAs.

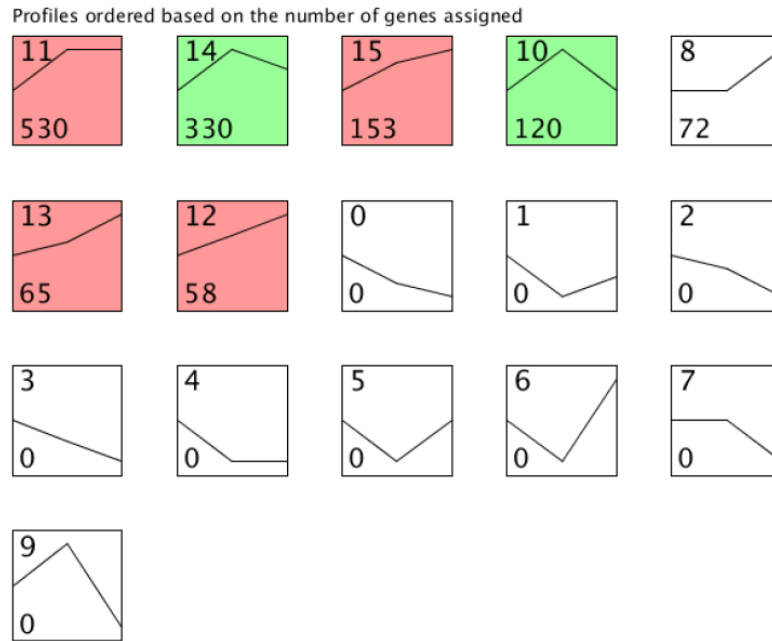


Figure 3. Expression profile clusters of lncRNAs across MPV, F₁, allotetraploid S₁₄. The number in the top left-hand corner of a profile box is the profile ID number; bottom number represent number of genes assigned to the expression profile. The colored profiles had a statistically significant number of genes assigned. Non-white profiles of the same color represent profiles grouped into a single cluster. MPV: mid-parent value.

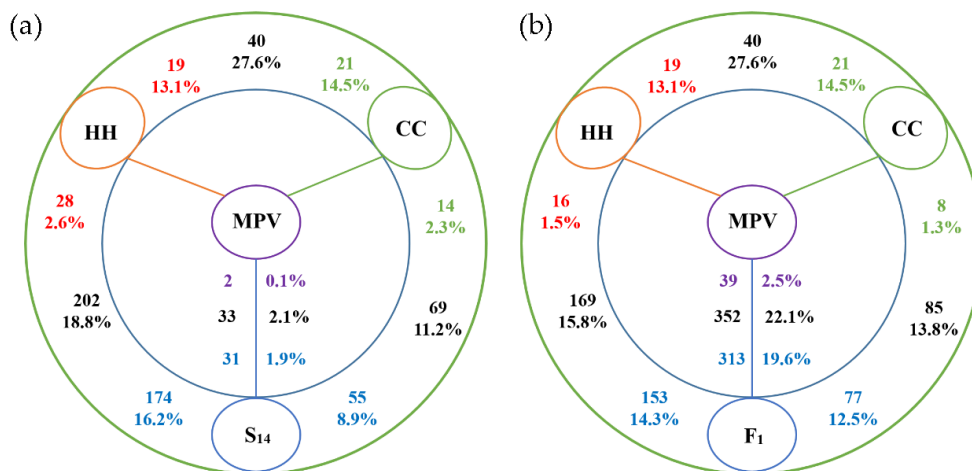


Figure 4. lncRNA genes differentially expressed in S₁₄ (a)/F₁ (b) progeny and their diploid (HH, CC) progenitors. Numbers close to the species (colored) represent upregulated lncRNAs compared with the neighboring species. Percentages indicate those among all expressed lncRNAs. The total number of lncRNAs differentially expressed between the two species is given (black).

In order to distinguish the effects of hybridization and genome doubling on lncRNA expression, the interspecific F_1 hybrid was used to compare with its parents and MPV. The results of differential expression analysis showed that the lncRNAs expressed higher in F_1 were 9.56 (153:16) and 9.62 (77:8) times more than that in HH and CC, respectively (Figure 4b). Contrary to the result in allotetraploid progeny S_{14} , a high portion (22.1%, 352) of the expressed genes were identified as nonadditively expressed lncRNA genes in F_1 when compared with MPV derived from the two parental lines. 88.9% (313) of these non-additively expressed lncRNA-genes (MPV- F_{1dif}) were highly expressed in interspecific F_1 hybrid progeny (Figure 4b), of which 65 lncRNA-genes were highly expressed in F_1 hybrid when compared with its parents (F_1 -CC $_{dif}$, F_1 -HH $_{dif}$) (Figure S8). Of the differentially expressed lncRNA-genes between parents, only 2 (10.5% of HH highly expressed; 19) and 1 (4.8% in CC highly expressed; 21) lncRNAs were retained in the F_1 hybrid (Figure S8). The number of non-additively expressed genes in F_1 to S_{14} reduced from 352 to 33. Among 352 non-additively expressed genes, only three lncRNA-genes (LINC-chh12G002100-1, LINC-chh10G005690-1, LINC-chc05G015470-1) were transgressively heritable across F_1 and S_{14} generations (Figure S9). After interspecific hybridization, a total of 148 lncRNA-genes showed significantly differential expression in leaves of *C. hytivus* S_{14} compared with interspecific F_1 hybrid, including 69 upregulated genes and 79 downregulated genes (Figure S7f).

Overall, these results indicated that the expression of lncRNAs were up-regulated after the distant hybridization was completed. Then, most of the lncRNAs remained stable in polyploidization, and another batch of lncRNAs had up-regulated or down-regulated differentiation. Moreover, the number and difference in polyploidization were much lower than in the hybridization process, suggesting these allopolyploidy-related lncRNAs might play different roles in the two important stages.

3.5. Functional Analysis of Allopolyploidization-Related lncRNAs

To reveal the potential functions of the allopolyploidization-related lncRNAs, the union of DE-lncRNAs between *C. hytivus* S_{14} and its diploid ancestors (or MPV) was obtained. We predicted 253 genomic co-location protein-coding (*cis*-regulated) genes as potential target genes (PTGs) of the DE-lncRNAs to participate in allopolyploidization. Gene Ontology (GO) functional categories of PTGs revealed enrichment for photosynthesis (Figure 5; p -value < 0.05), including four chloroplast cellular component genes (GO:0044434) in the cellular components category, seven iron-ion-binding genes (GO:0005506) and eight oxidoreductase activity genes (GO:0016705) in the molecular function category. In addition, there were two PTGs involved in sulfate transport and fatty acid metabolism, respectively.

Hybridization and polyploidization are two crucial processes in genome evolution. To exploit more possible functions of the hybridization-related lncRNAs, the union of DE-lncRNAs between interspecific F_1 hybrid and its diploid ancestors (or MPV) was obtained. We predicted 501 *cis*-regulated protein-coding genes as PTGs of the DE-lncRNAs to participate in interspecific hybridization. GO enrichment analysis of PTGs revealed enrichment for translation function (Figure S10; p -value < 0.05), including 15 genes related to structural constituent of ribosome (such as ribonucleoprotein complex) and seven constituent genes of mitochondria. Furthermore, target genes of DE-lncRNAs also correspond to chloroplast composition and photosynthesis-related processes, which are mainly involved in stomatal complex development (two), chloroplast organization (nine), light-harvesting (two), chlorophyll biosynthetic process (two).

The results of functional enrichment of target genes suggested that allopolyploidization-related lncRNAs (differentially expressed in allopolyploidy) plays an important role in the regulation of biological processes of organelles, especially in photosynthesis. Among these lncRNAs, abundance expression of 43 lncRNAs related to chloroplast were visualized in heatmap (Figure S11; Table S5). The absolute value of correlation coefficient between lncRNAs abundance and corresponding PTGs ranged from 0.62 to 0.99 according to p -value at 5% level, of which chc03G019620-INside-1, chc03G030700-AS-1, chc06G011860-AS-1, chh06G012250-INside-1, chh10G006290-AS-1, chh10G012720-AS-1 and their candidate target genes reached 0.99 (Table S6).

These results implied the possibility of molecular connection among lncRNAs and PTGs induced by polyploidization. Given the variation of leaf color of hybrid and allotetraploid, the chlorophyll content was measured in each species. As shown in our previous findings, the leaf chlorophyll (Chl) content of *C. sativus* (CC), *C. hystrix* (HHCC), and *C. hystrix* (HH) were 19.67, 15.38, and 29.06, respectively [58]. The target gene of LINC-chc01G00070-1 down-regulated in *C. hystrix* (\log_2FC : -1.63) was the key gene (*CHLM*) in chlorophyll biosynthetic process, which was also downregulated (\log_2FC : -1.60 ; Table S6) when compared with the paternal parent (CC). In order to study the final effect of polyploidy on chloroplast activity and photosynthesis, photosynthetic parameters were measured in each species. The net photosynthetic rates (Pn) in *C. hystrix* were in the middle level between parents with the light change ($0\text{--}1500 \mu\text{mol m}^{-2} \text{s}^{-1}$) (Figure S12a). The stomatal conductance (Gs) of *C. hystrix* and *C. hystrix* was lower than that of *C. sativus*, even when the light intensity was increased (Figure S12b). Initially the maximum net photosynthesis rate (Pmax) of *C. hystrix* was similar to that of *C. sativus* but lower than in *C. hystrix* (Figure S12c). However, the maximum quantum efficiency of photosystem II (F_v/F_m) of *C. \times hystrix* showed over-parent genetics [58], suggesting that the photosystem II might be regulated by lncRNA associated with polyploidy. Similar to *CHLM*, lncRNAs (chc06G011860-AS-1, chc05G021510-AS-1) and their target genes, chlorophyll-binding protein gene (*LHCP*), were downregulated in S₁₄, and the target gene was downregulated by 2.46 and 2.36 times respectively. It was also found that the DE-lncRNAs were associated with the changes of key genes in photorespiration (*CRR6*), stomatal complex development (*MEPF2*, *MEPFL3*), and carbon fixation (*RuBisCO* small chain) (Tables S5 and S6). The allotetraploid maintained higher level of photosynthesis may be through improving the thylakoid membrane lipid components and function of protein complexes under genome shock.

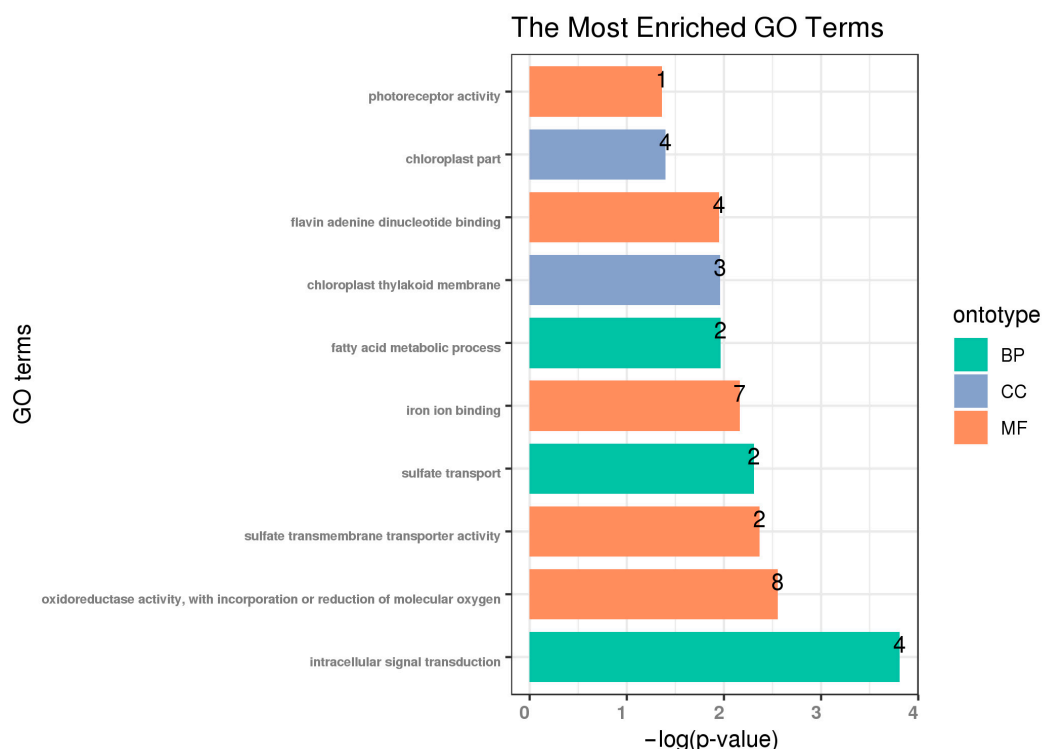


Figure 5. Gene Ontology (GO) enrichment analysis of potential target genes (PTGs) of differently expressed lncRNAs in allotetraploid S14 compared with parents/MPV. Shown are significantly enriched GO terms (top10, p -value < 0.05). The label at the top of the bar represents the number of genes. BP, biological process; MF, molecular function; CC, cellular component.

3.6. qPCR Verification for DE-lncRNAs and Their Target Genes

To confirm the validity of RNA-seq data, 14 differentially expressed genes (including 8 lncRNAs and 6 PCGs) involved in chloroplastic fractions were selected for qRT-PCR analysis (Table S1). The gene expression trend of qRT-PCR data was in accordance with RNA-seq data (Figure 6). Pearson correlation coefficient between TPM (Transcripts per Million) values and relative transcript abundance estimated from qRT-PCR was found to be significant ($R = 0.91$, p -value < 0.01), indicating the reliability of our RNA-seq data.

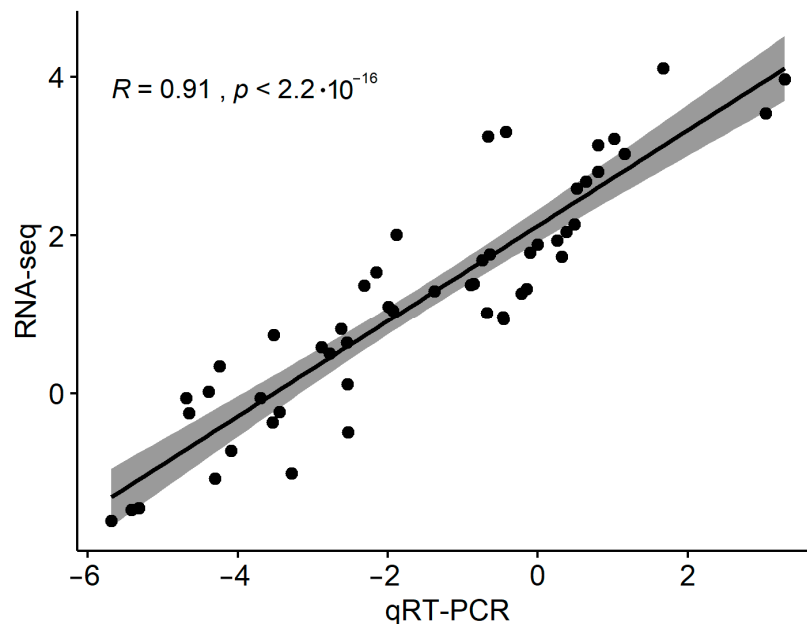


Figure 6. Transcript abundance estimated from qPCR analysis and from RNA-Seq data. The y axis represents $\log_{10}(\text{TPM})$, and the x axis represents $\log_{10}(2^{-\Delta\text{Ct}})$. The shadow band shows the confidence interval.

4. Discussion

4.1. Genome-Wide Identification, Characterization of lncRNAs in Allotetraploid *C. hytivus*

The prevalence of polyploid plants demonstrates the adaptive advantages of polyploid in evolution [59–61]. *C. hytivus* (HHCC, $2n = 4x = 38$) is the first synthetic crop plant in *Cucumis* with increased ploidy level and showing growth vigor and improved adaptability [58]. *C. hytivus* has become one of the most important resources for introgression of needed traits from wild relatives into cucumber (*Cucumis hystrix* Chakr., $2n = 2x = 24$) [62–65]. Taking advantage of its small genome size and clear genetic background, the newly synthesized allotetraploid has been examined as a model for many studies of allopolyploidization, such as genomic changes [66–71], gene expression patterns [39,72], epigenetic modification [73], chromosomal behavior [67,74], and miRNA regulation [75]. Like miRNAs, long non-coding RNA (lncRNAs) are crucial “regulatory” ncRNAs in the post-transcriptional and transcriptional regulation in eukaryotes [76]. However, the role of lncRNAs in the formation of *C. hytivus* remains undiscovered.

The advent of next-generation sequencing technology and the availability of the newly sequenced *C. hytivus* genome provided unprecedented opportunities to examine the impact of this important process on lncRNAs expression patterns. In the present study, a total of 1593 lncRNA loci (2206 isoforms) were identified in *C. hytivus* based on stringent criteria. lncRNAs were also identified in other polyploid plants, example given bread wheat (44,698) [77], cotton (8514) [34], tobacco (7423) [78], and *Brassica napus* (549) [79]. The identification efficiency of lncRNAs may be affected by their reference genome,

RNA sequencing methods, and different identification strategies. Because the full-length cDNA library was enriched from Oligo (dT) magnetic beads, these lncRNAs without poly(A) tails have not been identified in the present study. However, most of the lncRNAs (>84.5%) [34] are transcribed by RNA polymerase II (Pol II), which are always capped, polyadenylated, and frequently spliced, much like coding mRNAs [55]. In the present study, we focus on the RNA pol II transcribed nuclear lncRNAs.

The distribution of lncRNA throughout the chromosome can provide clues about possible functions and mechanisms of action. These lncRNAs located near protein-coding genes may regulate the expression of its neighbors by *cis*-acting [55]. Other lncRNAs transcribed from centromeric regions may play roles in centromere maintenance and cellular division [80]. The activated lncRNAs in polyploid cotton genome are predominantly transcribed from demethylated TE regions, especially from long interspersed nuclear elements (LINEs) [34]. Centromeric regions of cucumber genome are enriched with transposable elements (TEs) and lack of protein-coding gene sites [81]. In *C. hytivus*, lncRNA locus display an even distribution across the chromosomes (Figure 2a), with active transcription from centromeric regions. Our results suggested that lncRNAs located near centromeres or in gene deserts are implicated to act distally or have other effects in genome duplication.

Furthermore, approximately 19,252 gene pairs were one-to-one syntenic in two subgenomes (23,108 genes in subH; 22,535 genes in subC) (Figure S2b). These highly syntenic genomes helped identify the homologous lncRNA loci. Sequences of the lncRNA pairs with positional similarity within *C. hytivus* genome were compared, which allowed identification of 145 pairs of homeologous loci (Figure S2a). Using the collinear method, 103 homologous lincRNA loci with sequence similarities were identified in soybean, a diploidized ancient tetraploid crop [82]. In general, the primary model for rapid lncRNA emergence and decay in plants whole genome duplications and genome rearrangements may result in the lncRNA non-conservativeness. Although lncRNAs show less conservation at the primary sequences compared with mRNAs, some lncRNAs show striking conservation at the level of synteny and function [83].

According to the location relative to the nearest protein-coding gene, lncRNA can be further grouped into sense, antisense, intronic, and intergenic lncRNA [11]. In most plants, lncRNAs that come from the intergenic regions constitute the highest proportion, such as 85.1% in tomato [84], 79.7% in *Brassica napus* [47], and about 90% in cotton [34]. Similar result was also observed in our study, the majority of lncRNAs were located in intergenic regions (44.6%). However, divergent results were also reported in some other plants. For example, antisense lncRNAs are abundant in *Arabidopsis thaliana* [30], sense lncRNAs have a higher proportion in *Populus tomentosa* (75.6%) [85] and *Ananas comosus* [86]. Therefore, the location of lncRNAs may vary significantly from one species to another.

Candidate transcripts with only one exon have been removed in this study, because transcripts with two exons or more are more stable than unspliced (contain single-exon) transcripts. Despite this, most of the *C. hytivus* lncRNAs are relatively shorter and contain fewer exons than the protein-coding genes, consistent with the results of studies on tomato [84], *P. tomentosa* [85], and cotton [34]. Meanwhile, the transcript levels of the lncRNAs were also significantly lower than those of the mRNAs. Similar results were observed for lncRNAs in other plant species, such as *Arabidopsis* [87], rice [88] and *P. tomentosa* [85]. *C. hytivus* as a nascent polyploid plant, its lncRNAs showed both conventional characteristics and new special aspects in polyploid. It is very likely that these lncRNAs enable *C. hytivus* with greater physiological and ecological plasticity. Thus, it is necessary to compare the expression levels of lncRNAs in different samples and explore their regulation roles in the allopolyploidization process.

4.2. Non-Additive lncRNA Expression under Intergeneric Hybridization and Polyploidization

Polyploidy is a major driver of genome evolution in dicotyledons. Genome-wide changes including variations in mRNA expression and alteration of epigenetic modifications, such as DNA methylation and histone modification, are frequently observed in polyploids. Most PCGs in synthetic allopolyploid are expressed at mid-parent level [8,9]. In our study, the hybridization and subsequent genome

duplication also resulted in many intermediate morphological traits in *C. hystrix*, including plant height and stem diameter (Table S7), which was consistent with our previous results [58]. By contrast, we found that lncRNA expression changed dramatically by the genomic shock of interspecific hybridization in hybrid; several lncRNAs are related to agronomic traits. The differently expressed lncRNAs in hybrid/allopolyploid compared with maternal parent (HH) were more than with paternal parent (CC), suggesting parental-biased expression under hybridization and allopolyploidization, which were consistent with PCGs [89–91]. However, contrary to PCGs, most differentially expressed (DE) lncRNAs showed up-regulation in the hybrid/allopolyploid relative to the parents. It implied that these DE lncRNAs had a potential regulatory role in polyploidization.

Based on the GO enrichment analyses of potential target gene (PTGs) of differently expressed lncRNAs, we found that many thylakoid membrane, energy metabolism, and photosynthesis related genes were differentially expressed in the allopolyploidization. Chlorophylls are of great importance in plant light-harvesting and play essential roles in responding to changes in environmental conditions [92]. Chloroplast, as the organelle for photosynthesis, its photosynthetic membrane is considered the most complex and ingenious structure in all biofilms. When the light and temperature conditions change, it will capture the light quantum and drive a series of redox reactions [93]. The primary reaction of photosynthesis occurs on thylakoid membrane in chloroplast, which contains photosynthetic pigments and electron transport chain components, and the conversion of light energy to chemical energy is carried out on thylakoids. [94]. Some lncRNAs of the wild banana may regulate thylakoid membrane to maintain photosynthesis in low temperature stress [95]. The biological metabolism of plant chlorophyll is a complex and precise process, which takes 15 steps and involves multiple genes such as glutamine tRNA synthetase gene (*glxt*) [96]. Recently, many lncRNAs related to chlorophyll synthesis have been found in *Ananas comosus* [86]. In previous studies, we found that the chloroplast development of allotetraploid *C. hystrix* was hindered, and the genes *HEMA1*, *HEMA2*, and *POR* related to chlorophyll biosynthesis were inhibited, resulting in a decrease in chlorophyll content [57]. Meanwhile, the expression of lncRNA near the genes of chlorophyll-binding protein and heme-binding protein changed, which might further affect the lineage-specific emergence of photosynthesis. More subsequent experiments concerning the fluorescence in situ hybridization (RNA-FISH) and silencing of *C. hystrix* antisense lncRNAs *chc06G011860-AS-1* etc. may provide more information for revealing the regulation mechanism of leaf yellowing mediated by allopolyploidization.

5. Conclusions

In this study, we identified 2206 lncRNAs in allotetraploid *C. hystrix* by high throughput sequencing, among which most lncRNAs (~66.1%) were found to be active in allopolyploidization. GO enrichments of genes that either overlap with or are immediate neighbors of these significantly differentially expressed lncRNAs compared with two ancestors and MPV suggested involvement of many biological processes, including chloroplast components and photosynthesis biological process. Moreover, investigations on photosynthetic parameters provide necessary information for possible molecular mechanisms that LINC-*chc01G000070-1*, *chc01G006010-AS-1*, *chc05G021510-AS-1*, *chc06G011860-AS-1*, and other lncRNAs may be the key epigenetic regulators of leaf chlorosis and photosynthetic enhancement in *C. hystrix*. These results present a comprehensive atlas of lncRNAs in *Cucumis* allopolyploidization and paves the way for future understanding of the function of lncRNAs in polyploidization.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/11/12/1500/s1>. Figure S1: Materials used for RNA sequencing. Figure S2: Cumulative distribution function of coding probability of candidate lncRNA and mRNA against CPAT. Figure S3: Dot plot for lncRNA/gene collinear blocks on two subgenomes of allopolyploid *C. hystrix* (HHCC). Figure S4: Description of the sequence characteristics of lncRNAs in *C. hystrix* leaves. Figure S5: Gene expression correlations between three biological replicates of lncRNA and mRNA in *C. sativus* (CC), *C. hystrix* (HH), interspecific hybrid F₁, allotetraploid *C. hystrix* (S₁₄). Figure S6: Comparison of the expression levels of lncRNAs and protein coding transcripts. Figure S7: Volcano plot of differentially expressed lncRNAs. Figure S8: UpSetR plot of variants across eight differentially expressed lncRNA set

queries. Figure S9: The number of differentially expressed lncRNA genes detected in F₁ hybrid and allotetraploid S₁₄ progeny compared with diploid progenitors (MPV). Figure S10: GO enrichment analysis of PTGs of differently expressed lncRNAs in F₁ hybrid compared with parents/MPV. Figure S11: Heat maps of lncRNAs and PTGs expression in F₁ hybrid, allotetraploid S₁₄ and their parents (HH, CC). Figure S12: Photosynthetic characterization of the three species about allotetraploid S₁₄ and their parents (HH, CC). Table S1: Sequences of qRT-PCR primers. Table S2: The number of lncRNAs and gene overlapping with or without TEs. Table S3: Basic summary of the expression levels of lncRNAs and mRNAs. Table S4: Seven bins of MPV-F₁-S₁₄ expression profile of lncRNAs. Table S5: Annotation of potential target genes of lncRNAs involved in photosynthesis. Table S6: Effect of allotetraploidization on the expression of lncRNAs and their targets involved in photosynthesis. Table S7: Plant growth characteristics among the *C. hytivus* and diploid parents.

Author Contributions: Formal analysis, P.W., Z.Z., and Y.Z.; funding acquisition, J.C. and X.Y.; investigation, Q.Z.; methodology, P.W., Q.L., and J.L. resources, J.C.; visualization, P.W.; writing—original draft, P.W.; writing—review and editing, X.Y., Y.M., and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (NSFC 31902006), the Fundamental Research Funds for the Central Universities (KJQN202029), Natural Science Foundation of Jiangsu Province (BK20180536), the National Key Research and Development Program of China (2018YFD1000804) and Jiangsu Belt and Road innovation cooperation project: Cooperative development of overseas application of molecular breeding technology in horticultural crops (BZ2019012).

Acknowledgments: We wish to thank Yuhui Wang for her English language assistance.

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

References

- Rinn, J.L.; Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **2012**, *81*, 145–166. [[CrossRef](#)] [[PubMed](#)]
- Ponting, C.P.; Oliver, P.L.; Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **2009**, *136*, 629–641. [[CrossRef](#)] [[PubMed](#)]
- Li, A.L.; Liu, D.C.; Wu, J.; Zhao, X.B.; Hao, M.; Geng, S.F.; Yan, J.; Jiang, X.X.; Zhang, L.Q.; Wu, J.Y.; et al. mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Plant Cell* **2014**, *26*, 1878–1900. [[CrossRef](#)] [[PubMed](#)]
- Iyer, M.K.; Niknafs, Y.S.; Malik, R.; Singhal, U.; Sahu, A.; Hosono, Y.; Barrette, T.R.; Prensner, J.R.; Evans, J.R.; Zhao, S.; et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **2015**, *47*, 199–208. [[CrossRef](#)] [[PubMed](#)]
- Pauli, A.; Valen, E.; Lin, M.F.; Garber, M.; Vastenhouw, N.L.; Levin, J.Z.; Fan, L.; Sandelin, A.; Rinn, J.L.; Regev, A.; et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **2012**, *22*, 577–591. [[CrossRef](#)] [[PubMed](#)]
- Zhang, Y.C.; Liao, J.Y.; Li, Z.Y.; Yu, Y.; Zhang, J.P.; Li, Q.F.; Qu, L.H.; Shu, W.S.; Chen, Y.Q. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* **2014**, *15*, 512. [[CrossRef](#)] [[PubMed](#)]
- Li, L.; Eichten, S.R.; Shimizu, R.; Petsch, K.; Yeh, C.T.; Wu, W.; Chettoo, A.M.; Givan, S.A.; Cole, R.A.; Fowler, J.E.; et al. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* **2014**, *15*, R40. [[CrossRef](#)]
- Liu, J.; Jung, C.; Xu, J.; Wang, H.; Deng, S.L.; Bernad, L.; Arenas-Huertero, C.; Chua, N.H. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* **2012**, *24*, 4333–4345. [[CrossRef](#)]
- Necsulea, A.; Kaessmann, H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat. Rev. Genet.* **2014**, *15*, 734–748. [[CrossRef](#)]
- Chen, L.L.; Carmichael, G.G. Decoding the function of nuclear long noncoding RNAs. *Curr. Opin. Cell Biol.* **2010**, *22*, 357–364. [[CrossRef](#)]
- Duret, L.; Chureau, C.; Samain, S.; Weissenbach, J.; Avner, P. The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **2006**, *312*, 1653–1655. [[CrossRef](#)] [[PubMed](#)]
- Zhao, J.; Sun, B.K.; Erwin, J.A.; Song, J.J.; Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **2008**, *322*, 750–756. [[CrossRef](#)] [[PubMed](#)]
- Prensner, J.R.; Iyer, M.K.; Balbin, O.A.; Dhanasekaran, S.M.; Cao, Q.; Brenner, J.C.; Laxman, B.; Asangani, I.A.; Grasso, C.S.; Kominsky, H.D.; et al. Transcriptome sequencing across a prostate cancer cohort identifies

- PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* **2011**, *29*, 742–749. [[CrossRef](#)] [[PubMed](#)]
14. Sleutels, F.; Zwart, R.; Barlow, D.P. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **2002**, *415*, 810–813. [[CrossRef](#)]
 15. Rinn, J.L.; Kertesz, M.; Wang, J.K.; Squazzo, S.L.; Xu, X.; Bruggmann, S.A.; Goodnough, L.H.; Helms, J.A.; Farnham, P.J.; Segale, E.; et al. Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **2007**, *129*, 1311–1323. [[CrossRef](#)]
 16. Liu, F.; Marquardt, S.; Lister, C.; Swiezewski, S.; Dean, C. Targeted 3' processing of antisense transcripts triggers *Arabidopsis* *FLC* chromatin silencing. *Science* **2010**, *327*, 94–97. [[CrossRef](#)]
 17. Heo, J.B.; Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **2011**, *331*, 76–79. [[CrossRef](#)]
 18. Mach, J. The long-noncoding RNA ELENA1 functions in plant immunity. *Plant Cell* **2017**, *29*, 916. [[CrossRef](#)]
 19. Yatusевич, R.; Fedak, H.; Ciesielski, A.; Krzyczmonik, K.; Kulik, A.; Dobrowolska, G.; Swiezewski, S. Antisense transcription represses *Arabidopsis* seed dormancy QTL *DOG1* to regulate drought tolerance. *EMBO Rep.* **2017**, *18*, 2186–2196. [[CrossRef](#)]
 20. Amor, B.B.; Wirth, S.; Merchan, F.; Laporte, P.; d'Aubenton-Carafa, Y.; Hirsch, J.; Maizel, A.; Mallory, A.C.; Lucas, A.; Deragon, J.M.; et al. Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res.* **2009**, *19*, 57–69. [[CrossRef](#)]
 21. Matsui, A.; Ishida, J.; Morosawa, T.; Mochizuki, Y.; Kaminuma, E.; Endo, T.A.; Okamoto, M.; Nambara, E.; Nakajima, M.; Kawashima, M.; et al. *Arabidopsis* transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol.* **2008**, *49*, 1135–1149. [[CrossRef](#)] [[PubMed](#)]
 22. Wang, H.; Chung, P.J.; Liu, J.; Jang, I.C.; Kean, M.J.; Xu, J.; Chua, N.H. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res.* **2014**, *24*, 444–453. [[CrossRef](#)] [[PubMed](#)]
 23. Wang, Y.; Fan, X.; Lin, F.; He, G.; Terzaghi, W.; Zhu, D.; Deng, X.W. *Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10359–10364. [[CrossRef](#)] [[PubMed](#)]
 24. Franco-Zorrilla, J.M.; Valli, A.; Todesco, M.; Mateos, I.; Puga, M.I.; Rubio-Somoza, I.; Leyva, A.; Weigel, D.; Garcia, J.A.; Paz-Ares, J. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **2007**, *39*, 1033–1037. [[CrossRef](#)]
 25. Ding, J.H.; Lu, Q.; Ouyang, Y.D.; Mao, H.L.; Zhang, P.B.; Yao, J.L.; Xu, C.; Li, X.; Xiao, J.; Zhang, Q. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2654–2659. [[CrossRef](#)]
 26. Jiao, Y.; Wickett, N.J.; Ayyampalayam, S.; Chanderbali, A.S.; Landherr, L.; Ralph, P.E.; Tomsho, L.P.; Hu, Y.; Liang, H.; Soltis, P.S.; et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* **2011**, *473*, 97–100. [[CrossRef](#)]
 27. Chen, Z.J. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* **2007**, *58*, 377–406. [[CrossRef](#)]
 28. Leitch, A.R.; Leitch, I.J. Genomic plasticity and the diversity of polyploid plants. *Science* **2008**, *320*, 481–483. [[CrossRef](#)]
 29. Soltis, D.E.; Albert, V.A.; Leebens-Mack, J.; Bell, C.D.; Paterson, A.H.; Zheng, C.; Sankoff, D.; Depamphilis, C.W.; Wall, P.K.; Soltis, P.S. Polyploidy and angiosperm diversification. *Am. J. Bot.* **2009**, *96*, 336–348. [[CrossRef](#)]
 30. Gaeta, R.T.; Pires, J.C.; Iniguez-Luy, F.; Leon, E.; Osborn, T.C. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* **2007**, *19*, 3403–3417. [[CrossRef](#)]
 31. Matsushita, S.C.; Tyagi, A.P.; Thornton, G.M.; Pires, J.C.; Madlung, A. Allopolyploidization lays the foundation for evolution of distinct populations: Evidence from analysis of synthetic *Arabidopsis* allohexaploids. *Genetics* **2012**, *191*, 535–547. [[CrossRef](#)] [[PubMed](#)]
 32. Song, Q.; Chen, Z.J. Epigenetic and developmental regulation in plant polyploids. *Curr. Opin. Plant Biol.* **2015**, *24*, 101–109. [[CrossRef](#)] [[PubMed](#)]
 33. Yoo, M.J.; Szadkowski, E.; Wendel, J.F. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity (Edinb)* **2013**, *110*, 171–180. [[CrossRef](#)] [[PubMed](#)]

34. Zhao, T.; Tao, X.Y.; Feng, S.L.; Wang, L.Y.; Hong, H.; Ma, W.; Shang, G.D.; Guo, S.S.; He, Y.X.; Zhou, B.L.; et al. LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. *Genome Biol.* **2018**, *19*, 195. [[CrossRef](#)] [[PubMed](#)]
35. Chen, J.F.; Kirkbride, J.H. A new synthetic species of *Cucumis* (*Cucurbitaceae*) from interspecific hybridization and chromosome doubling. *Brittonia* **2000**, *52*, 315–319. [[CrossRef](#)]
36. Chen, J.F.; Staub, J.E.; Tashiro, Y.; Isshiki, S.; Miyazaki, S. Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. *Euphytica* **1997**, *96*, 413–419. [[CrossRef](#)]
37. Chen, J.F.; Jeffrey, W.A.; Staub, J.E.; Skorupska, H.T.; Rhodes, B.B. A new synthetic amphidiploid in *Cucumis* from a *C. sativus* × *C. hystrix* F₁ interspecific hybrid. *Cucurbitaceae* **1998**, *1*, 336–339.
38. Sebastian, P.; Schaefer, H.; Telford, I.R.; Renner, S.S. Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14269–14273. [[CrossRef](#)]
39. Zhuang, Y.; Chen, J.F. Changes of gene expression in early generations of the synthetic allotetraploid *Cucumis* × *hytivus* Chen et Kirkbride. *Genet. Resour. Crop. Evol.* **2009**, *56*, 1071–1076. [[CrossRef](#)]
40. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)]
41. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295.
42. Kang, Y.J.; Yang, D.C.; Kong, L.; Hou, M.; Meng, Y.Q.; Wei, L.P.; Gao, G. Cpc2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **2017**, *45*, W12–W16. [[CrossRef](#)] [[PubMed](#)]
43. Li, A.; Zhang, J.Y.; Zhou, Z.Y. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved *k-mer* scheme. *BMC Bioinform.* **2014**, *15*, 311. [[CrossRef](#)] [[PubMed](#)]
44. Wang, L.G.; Park, H.J.; Dasari, S.; Wang, S.Q.; Kocher, J.P.; Li, W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **2013**, *41*, e74. [[CrossRef](#)] [[PubMed](#)]
45. Sun, L.; Luo, H.T.; Bu, D.C.; Zhao, G.G.; Yu, K.T.; Zhang, C.H.; Liu, Y.N.; Chen, R.S.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **2013**, *41*, e166. [[CrossRef](#)] [[PubMed](#)]
46. Roberts, A.; Pimentel, H.; Trapnell, C.; Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **2011**, *27*, 2325–2329. [[CrossRef](#)]
47. Shen, E.; Zhu, X.T.; Hua, S.J.; Chen, H.Y.; Ye, C.Y.; Zhou, L.H.; Liu, Q.; Zhu, Q.H.; Fan, L.J.; Chen, X. Genome-wide identification of oil biosynthesis-related long non-coding RNAs in allopolyploid *Brassica napus*. *BMC Genom.* **2018**, *19*, 1–13. [[CrossRef](#)]
48. Wang, Y.; Tang, H.; DeBarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)]
49. Krzywinski, M.; Schein, J.I. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [[CrossRef](#)]
50. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **2016**, *34*, 525–527. [[CrossRef](#)]
51. Ernst, J.; Bar-Joseph, Z. STEM: A tool for the analysis of short time series gene expression data. *BMC Bioinform.* **2006**, *7*, 191. [[CrossRef](#)] [[PubMed](#)]
52. Love, M.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
53. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, R106. [[CrossRef](#)] [[PubMed](#)]
54. Liao, Q.; Liu, C.N.; Yuan, X.Y.; Kang, S.L.; Miao, R.Y.; Xiao, H.; Zhao, G.G.; Luo, H.T.; Bu, D.H.; Zhao, H.T. Large-scale prediction of long noncoding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res.* **2011**, *39*, 3864–3878. [[CrossRef](#)] [[PubMed](#)]
55. Gil, N.; Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* **2020**, *21*, 102–117. [[CrossRef](#)]
56. Yu, G.C.; Wang, L.G.; Han, Y.Y.; He, Q.Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS.* **2012**, *16*, 284–287. [[CrossRef](#)] [[PubMed](#)]

57. Yu, X.Q.; Wang, X.X.; Hyldgaard, B.; Zhu, Z.B.; Zhou, R.; Kjaer, K.H.; Ouzounis, T.; Lou, Q.F.; Li, J.; Cai, Q.S.; et al. Allopolyploidization in *Cucumis* contributes to delayed leaf maturation with repression of redundant homoeologous genes. *Plant J.* **2018**, *94*, 393–404. [[CrossRef](#)]
58. Yu, X.Q.; Hyldgaard, B.; Rosenqvist, E.; Ottosen, C.O.; Chen, J.F. Interspecific hybridization in *Cucumis* leads to the divergence of phenotypes in response to low light and extended photoperiods. *Front. Plant. Sci.* **2015**, *6*, 802. [[CrossRef](#)]
59. Comai, L. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **2005**, *6*, 836–846. [[CrossRef](#)]
60. Wood, T.E.; Takebayashi, N.; Barker, M.S.; Mayrose, I.; Greenspoon, P.B.; Rieseberg, L.H. The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13875–13879. [[CrossRef](#)]
61. Mayrose, I.; Zhan, S.H.; Rothfels, C.J.; Magnuson-Ford, K.; Barker, M.S.; Rieseberg, L.H.; Otto, S.P. Recently formed polyploid plants diversify at lower rates. *Science* **2011**, *333*, 1257. [[CrossRef](#)] [[PubMed](#)]
62. Wang, Y.Z.; Zhang, Z.T.; Jia, L.; Li, Z.A.; Li, J.; Lou, Q.F.; Chen, J.F. Molecular and cytogenetic analyses provide evidence of the introgression of chromosomal segments from the wild *Cucumis hystrix* into the cultivated cucumber through the bridge of a synthetic allotetraploid. *Mol. Breed.* **2017**, *37*, 89. [[CrossRef](#)]
63. Zhang, K.J.; Wang, X.; Zhu, W.W.; Qin, X.D.; Xu, J.; Cheng, C.Y.; Lou, Q.F.; Li, J.; Chen, J.F. Complete resistance to powdery mildew and partial resistance to downy mildew in a *Cucumis hystrix* introgression line of cucumber were controlled by a co-localized locus. *Theor. Appl. Genet.* **2018**, *131*, 2229–2243. [[CrossRef](#)] [[PubMed](#)]
64. Cheng, C.Y.; Wang, X.; Liu, X.J.; Yang, S.Q.; Yu, X.Q.; Qian, C.T.; Li, J.; Lou, Q.F.; Chen, J.F. Candidate genes underlying the quantitative trait loci for root-knot nematode resistance in a *Cucumis hystrix* introgression line of cucumber based on population sequencing. *J. Plant Res.* **2019**, *132*, 813–823. [[CrossRef](#)] [[PubMed](#)]
65. Wang, X.; Cheng, C.Y.; Li, Q.R.; Zhang, K.J.; Lou, Q.F.; Li, J.; Chen, J.F. Multi-omics analysis revealed that MAPK signaling and flavonoid metabolic pathway contributed to resistance against *Meloidogyne incognita* in the introgression line cucumber. *J. Proteom.* **2020**, *220*, 103675. [[CrossRef](#)] [[PubMed](#)]
66. Chen, J.F.; Zhuang, F.Y.; Liu, X.A.; Qian, C.T. Reciprocal differences of morphological and DNA characters in interspecific hybridization in *Cucumis*. *Can. J. Bot.* **2004**, *82*, 16–21. [[CrossRef](#)]
67. Chen, L.Z.; Lou, Q.F.; Zhuang, Y.; Chen, J.F.; Zhang, X.Q.; Wolukau, J.N. Cytological diploidization and rapid genome changes of the newly synthesized allotetraploids *Cucumis* × *hytivus*. *Planta* **2007**, *225*, 603–614. [[CrossRef](#)]
68. Jiang, B.; Lou, Q.F.; Chen, J.F. Genetic variation analysis of the newly synthesized allotetraploids *Cucumis hytivus* using retrotransposon-based SSAP markers. *Acta Hortic.* **2010**, *871*, 567–572. [[CrossRef](#)]
69. Jiang, B.; Lou, Q.F.; Wu, Z.M.; Zhang, W.P.; Wang, D.; Mbira, K.G.; Weng, Y.Q.; Chen, J.F. Retrotransposon- and microsatellite sequence-associated genomic changes in early generations of a newly synthesized allotetraploid *Cucumis*. *Plant. Mol. Biol.* **2011**, *77*, 225–233. [[CrossRef](#)]
70. Jiang, B.; Lou, Q.F.; Wang, D.; Wu, Z.M.; Zhang, W.P.; Chen, J.F. Allopolyploidization induced the activation of Ty1-copia retrotransposons in *Cucumis hytivus*, a newly formed *Cucumis* allotetraploid. *Bot. Stud.* **2011**, *52*, 145–152.
71. Shen, J.; Kere, M.G.; Chen, J.F. Mitochondrial genome is paternally inherited in *Cucumis* allotetraploid (*C.* × *hytivus*) derived by interspecific hybridization. *Sci. Hortic.* **2013**, *155*, 39–42. [[CrossRef](#)]
72. Zhuang, Y.; Chen, J.F.; Jahn, M. Expression and sequence variation of the cucumber *Por* gene in the synthesized allotetraploid *Cucumis* × *hytivus*. *Mol. Biol. Rep.* **2009**, *36*, 1725–1731. [[CrossRef](#)] [[PubMed](#)]
73. Chen, L.Z.; Chen, J.F. Changes of cytosine methylation induced by wide hybridization and allopolyploidy in *Cucumis*. *Genome* **2008**, *51*, 789–799. [[CrossRef](#)] [[PubMed](#)]
74. Wang, Y.Z.; Zhao, Q.Z.; Qin, X.D.; Yang, S.Q.; Li, Z.A.; Li, J.; Lou, Q.F.; Chen, J.F. Identification of all homoeologous chromosomes of newly synthetic allotetraploid *Cucumis* × *hytivus* and its wild parent reveals stable subgenome structure. *Chromosoma* **2017**, *126*, 713–728. [[CrossRef](#)] [[PubMed](#)]
75. Yu, X.Q.; Zhu, Z.B.; Zhang, T.; Li, J.; Cheng, C.Y.; Lou, Q.F.; Ottosen, C.O.; Chen, J.F. High-throughput sequencing reveals the change of microRNA expression caused by allopolyploidization in *Cucumis*. *Biol. Plant.* **2020**, *64*, 104–109. [[CrossRef](#)]
76. Kim, E.D.; Sung, S. Long noncoding RNA: Unveiling hidden layer of gene regulatory networks. *Trends Plant Sci.* **2012**, *17*, 16–21. [[CrossRef](#)]

77. Shumayla; Sharma, S.; Taneja, M.; Tyagi, S.; Singh, K.; Upadhyay, S.K. Survey of high throughput RNA-Seq data reveals potential roles for lncRNAs during development and stress response in bread wheat. *Front. Plant Sci.* **2017**, *6*, 1019.
78. Chen, X.; Sun, S.; Liu, F.J.; Shen, E.H.; Liu, L.; Ye, C.Y.; Xiao, B.G.; Timko, M.; Zhu, Q.H.; Fan, L.J.; et al. A transcriptomic profile of topping responsive non-coding RNAs in tobacco roots (*Nicotiana tabacum*). *BMC Genom.* **2019**, *20*, 856. [[CrossRef](#)]
79. Yao, S.L.; Liang, F.; Gill, R.A.; Huang, J.Y.; Cheng, X.H.; Liu, Y.Y.; Tong, C.B.; Liu, S.Y. A global survey of the transcriptome of the allopolyploid *Brassica napus* based on single molecule long-read isoform sequencing and Illumina-based RNA-seq data. *Plant J.* **2020**, *103*, 843–857. [[CrossRef](#)]
80. Rošić, S.; Erhardt, S. No longer a nuisance: Long non-coding RNAs join CENP-A in epigenetic centromere regulation. *Cell Mol. Life Sci.* **2016**, *73*, 1387–1398. [[CrossRef](#)]
81. Li, Q.; Li, H.B.; Huang, W.; Xu, Y.C.; Zhou, Q.; Wang, S.H.; Ruan, J.; Huang, S.W.; Zhang, Z.H. A chromosome-scale genome assembly of cucumber (*Cucumis sativus* L.). *GigaScience* **2019**, *8*, 1–10. [[CrossRef](#)] [[PubMed](#)]
82. Golicz, A.; Singh, M.B.; Bhalla, P.L. The long intergenic non-coding RNA (lincRNA) landscape of the soybean genome. *Plant Physiol.* **2017**, *176*, 01657.
83. Vandivier, L.E.; Andeson, S.J.; Folev, S.W.; Gregory, B.D. The conservation and function of RNA secondary structure in plants. *Annu. Rev. Plant Biol.* **2016**, *67*, 463–488. [[CrossRef](#)] [[PubMed](#)]
84. Zhu, B.Z.; Yang, Y.F.; Li, R.; Fu, D.Q.; Wen, L.W.; Luo, Y.B.; Zhu, H.L. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. *J. Exp. Bot.* **2015**, *66*, 4483–4495. [[CrossRef](#)]
85. Zhou, D.L.; Du, Q.Z.; Chen, J.H.; Wang, Q.S.; Zhang, D.Q. Identification and allelic dissection uncover roles of lncRNAs in secondary growth of *Populus tomentosa*. *DNA Res.* **2017**, *24*, 473–486. [[CrossRef](#)]
86. Lin, Z.; Xiong, Y.Y.; Xue, Y.B.; Mao, M.Q.; Xiang, Y.X.; He, Y.H.; Rafique, F.; Hu, H.; Liu, J.W.; Li, X.; et al. Screening and characterization of long noncoding RNAs involved in the albinism of *Ananas comosus* var. *bracteatus* leaves. *PLoS ONE* **2019**, *14*, e0225602. [[CrossRef](#)]
87. Henriques, R.; Wang, H.; Liu, J.; Boix, M.; Huang, L.F.; Chua, N.H. The antiphase regulatory module comprising *CDF5* and its antisense RNA *FLORE* links the circadian clock to photoperiodic flowering. *N. Phytol.* **2017**, *216*, 854–867. [[CrossRef](#)]
88. Yuan, J.P.; Li, J.R.; Yang, Y.; Tan, C.; Zhu, Y.M.; Hu, L.; Qi, Y.J.; Lu, Z. Stress-responsive regulation of long noncoding RNAs polyadenylation in *Oryza sativa*. *Plant J.* **2017**, *93*, 814–827. [[CrossRef](#)]
89. Xu, Y.H.; Zhong, L.; Wu, X.M.; Fang, X.P.; Wang, J.B. Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta* **2009**, *229*, 471–483. [[CrossRef](#)]
90. Cheng, S.P.; Huang, Z.; Suo, Y.J.; Wang, J.; Kang, X.Y. Gene expression differences associated with growth vigor in *Populus* full-sib allotriploid progeny following manipulated first division restitution of the diploid maternal parent. *Euphytica* **2015**, *203*, 683–700. [[CrossRef](#)]
91. Qi, X.Y.; Wang, H.B.; Song, A.P.; Jiang, J.F.; Chen, S.M.; Chen, F.D. Genomic and transcriptomic alterations following intergeneric hybridization and polyploidization in the *Chrysanthemum nankingense* × *Tanacetum vulgare* hybrid and allopolyploid (Asteraceae). *Hortic. Res.* **2018**, *5*, 5. [[CrossRef](#)] [[PubMed](#)]
92. Tewari, A.K.; Tripathy, B.C. Temperature-stress-induced impairment of chlorophyll biosynthetic reactions in cucumber and wheat. *Plant Physiol.* **1998**, *117*, 851–858. [[CrossRef](#)] [[PubMed](#)]
93. Albertsson, P.Å. The structure and function of the chloroplast photosynthetic membrane—A model for the domain organization. *Photosynth. Res.* **1995**, *46*, 141–149. [[CrossRef](#)] [[PubMed](#)]
94. Hippler, M.; Klein, J.; Fink, A.; Allinger, T.; Hoerth, P. Towards functional proteomics of membrane protein complexes: Analysis of thylakoid membranes from *Chlamydomonas reinhardtii*. *Plant J.* **2010**, *28*, 595–606. [[CrossRef](#)] [[PubMed](#)]
95. Liu, W.H.; Cheng, C.Z.; Lin, Y.L.; Xu, X.H.; Lai, Z.X. Genome-wide identification and characterization of mRNAs and lncRNAs involved in cold stress in the wild banana (*Musa itinerans*). *PLoS ONE* **2018**, *13*, e0200002. [[CrossRef](#)] [[PubMed](#)]

96. Chang, T.E.; Wegmann, B.; Wang, W.Y. Purification and characterization of glutamyl-tRNA synthetase: An enzyme involved in chlorophyll biosynthesis. *Plant Physiol.* **1990**, *93*, 1641–1649. [[CrossRef](#)]

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



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).