Genomics 99 (2012) 299–307

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Transcriptome analysis of rosette and folding leaves in Chinese cabbage using high-throughput RNA sequencing

Fengde Wang [a], Libin Li [a], Huayin Li [a], Lifeng Liu [a], Yihui Zhang [a], Jianwei Gao [a,*], Xiaowu Wang [b,**]

[a] Institute of Vegetables, Shandong Academy of Agricultural Sciences/Shandong Key Laboratory of Greenhouse Vegetable Biology/Shandong Branch of National Vegetable Improvement Center, Jinan 250100, China

[b] Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China

ARTICLE INFO

Article history:
Received 7 December 2011
Accepted 15 February 2012
Available online 23 February 2012

Keywords:
Chinese cabbage
Rosette leaves
Folding leaves
Transcriptome
High-throughput RNA sequencing

ABSTRACT

In this study, we report the first use of RNA-sequencing to gain insight into the wide range of transcriptional events that are associated with leafy head development in Chinese cabbage. We generated 53.5 million sequence reads (90 bp in length) from the rosette and heading leaves. The sequence reads were aligned to the recently sequenced Chiifu genome and were analyzed to measure the gene expression levels, to detect alternative splicing events and novel transcripts, to determine the expression of single nucleotide polymorphisms, and to refine the annotated gene structures. The analysis of the global gene expression pattern suggests two important concepts, which govern leafy head formation. Firstly, some stimuli, such as carbohydrate levels, light intensity and endogenous hormones might play a critical role in regulating the leafy head formation. Secondly, the regulation of transcription factors, protein kinases and calcium may also be involved in this developmental process.

1. Introduction

Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) is a widely cultivated and economically important vegetable crop in Asia. Chinese cabbage originated in China, and it has now become increasingly popular in other countries. Chinese cabbage has a tight leafy head, which is the storage organ for the cabbage and is the part that is commonly eaten. The leafy head is composed of a large number of heading leaves, which usually initiate after the rosette stage. The rosette leaves differentiate at the rosette stage and serve as the photosynthetic organs during the vegetative growth, whereas the heading leaves surrounding the shoot apexes are compact enough to form a tight head or heart. The production of Chinese cabbage is usually hampered by the poor heading of the leaves [1]. The understanding of the leafy head development at a molecular level will greatly facilitate the genetic improvement of the Chinese cabbage yield, its nutritional value and the quality of its appearance. The initiation and developmental process of the leafy head may be influenced by many factors, including the uneven distribution of auxin levels in the leaves, temperature, weak light, short days and the carbohydrate nutrition level [2]. However, because the results of many physiological experiments have not been explained or verified completely, how the plant senses environmental signals and subsequently turns on downstream gene regulation networks is largely unknown [1].

During cultivation, the vegetative growth of Chinese cabbage is divided into five stages, including the germination, seedling, rosette, heading and dormancy stages [3]. At the end of rosette stage, the young leaves, named folding leaves (FL), begin to fold inward, and the heading leaf differentiation is initiated. It has been reported that spraying auxin onto the dorsal side of the FL induces them to bend inward, whereas the juvenile leaves (JL) and rosette leaves (RL) do not react in the same way [2]. In addition, transcripts of the *BcpLH* gene have been detected in the FL but not in the JL or in the RL, indicating that this gene may play an important role in regulating the leafy head development in Chinese cabbage [1]. Compared with the other leaf types, the FL are unique because they accept external environmental stimuli and transmit these signals into morphological responses. The initiation of the leafy head is characterized by the bending inward of the leaves and is marked by the differentiation of the FL. Therefore, the development of the FL is an excellent model system to investigate the regulation, differentiation and development of the leafy head.

Recently, the mRNA differential display technique [4], the analysis of expressed sequence tags [5] and the differential hybridization method [1] have been used to examine the expression patterns of the RL and FL genes and have provided the first illustration of transcriptome dynamics during leafy head development. However, these data are far from being complete due to the limitations of these approaches.

RNA-sequencing (RNA-Seq) is a novel, high-throughput, deep-sequencing technology that is widely used for genomics research

*Corresponding author. Fax: +86 531 83179060.
**Corresponding author. Fax: +86 10 62146163.
E-mail addresses: jianweigao3@yahoo.com (J. Gao), wangxw@mail.caas.net.cn (X. Wang).

0888-7543/$ – see front matter © 2012 Published by Elsevier Inc.
doi:10.1016/j.ygeno.2012.02.005

and provides new strategies to analyze the functional complexity of transcriptomes. In particular, the Solexa/Illumina sequencing technology has many advantages as a revolutionary tool for transcriptome analysis, such as high coverage at a relatively low cost [6,7], and it has been used to investigate transcriptomes in plants, such as *Arabidopsis*, rice and berry [8–10].

To understand further the complexity of the transcriptome during leafy head development at the level of the whole genome, we performed the first global analysis of the transcriptomes from RL and FL in Chinese cabbage using the Solexa/Illumina RNA-Seq platform. This comprehensive analysis of the transcriptome dynamics serves as a blueprint for the gene expression profile of leafy head development. The data may substantially improve the global view of the Chinese cabbage transcriptome during leafy head development and pave the way for its further analysis and application to breeding practices. In addition, our analysis revealed the expression of numerous novel transcriptional units (TU), single nucleotide polymorphisms (SNPs) and alternative splicing (AS) events. We also refined the gene structures and found that compared to the existing gene annotations, a large number of the genes could be extended at the 5′ end, the 3′ end or at both ends.

2. Results

2.1. RNA-Seq and mapping of the sequence reads

To obtain a global view of transcriptome relevant to leafy head development in Chinese cabbage inbred line Fushanbaotou, the high-throughput RNA-Seq analysis on poly(A)-enriched RNAs from RL and FL libraries, respectively, was performed using the Solexa/Illumina platform. After filtering out the low-complexity reads, the low-quality reads and the repetitive reads, 27,088,098 usable reads for the RL and 26,386,316 usable reads for the FL were obtained (Table 1). To identify the gene expression patterns in the RL and the FL of Chinese cabbage, we mapped the reads against the sequenced Chinese cabbage genome (http://brassicadb.org/brad/) using the SOAP2 software [11], which was set to allow two base mismatches. Of the total reads, 70.6% of the reads matched to unique (67.7%) or multiple (2.9%) genome locations (Table 1). In addition, the annotated exons from the Chinese cabbage genome were used as reference genes to assign each read to a specific gene. As shown in Table 1, approximately 60% of the reads were mapped to unique genes, and 2.83% of the reads were mapped to multiple reference genes.

2.2. Global gene expression pattern analysis

The analysis of differentially expressed genes (DEGs) between the FL and RL libraries should aid our understanding of the molecular events involved in leafy head development. To confirm that the differences in the gene expression patterns observed among the developmental stages are statistically significant, we compared the RPKM-derived read count using Audic's method [12] with some modifications. A threshold value of FDR $\leq$ 0.001 and an absolute value of $\log_2$Ratio $\geq$ 1 were used to judge the significance of the differences in gene expression. Based on these criteria, 2255 genes were differentially expressed between the RL and FL including 953 genes that were up-regulated and 1302 genes that were down-regulated in the FL tissues (Supplementary Table S1). To understand their biological function, these DEGs were subjected to an on-line (http://brassicadb.org/brad/searchAll.php) BlastX search against the *Arabidopsis* genome.

The Gene Ontology (GO) database includes information on biological processes, molecular functions and cellular components and is an international standardized gene functional classification system, which offers a dynamically updated controlled vocabulary and a strictly defined concept that comprehensively describes the properties of genes and their products in any organism. To facilitate the global analysis of gene expression, a GO analysis was performed by mapping each differentially expressed gene into the records of the GO database (http://www.geneontology.org/). Under the biological process category, 20 GO categories were significantly enriched (corrected p-value $\leq$ 0.05) in the DEGs (Fig. 1A; Supplementary Table S2). Of these, a large number of gene responses to various stimuli, including carbohydrate, light intensity, endogenous and other biotic and abiotic stimuli, were prominently represented, suggesting that these stimuli are needed for leafy head development. Under the category of molecular function (Fig. 1B; Supplementary Table S2), the main functional groups of the DEGs (151 genes) were genes with nucleic acid binding transcription factor (TF) activity, including the MYB protein family, the zinc finger protein family, the AUX/IAA protein family, the WRKY protein family, and the bHLH protein family, which are involved in various plant developmental processes and stimuli responses. For the cellular component category (Fig. 1C; Supplementary Table S2), most of the DEGs were associated with the external encapsulating structure, the non-membrane-bounded organelle, the intracellular non-membrane-bound organelle and the cell periphery, whereas only a few DEGs were assigned to the chromosomal components, the chromatin, and the chromosome.

Because leafy head formation is affected by the auxin distribution in the leaves, a lower temperature, a greater day–night temperature difference, weak light, a short day length and sufficient carbohydrate nutrition [2], four subgroups were characterized from the group of DEGs that were related to plant development and the stimuli response (Table 2). The first subgroup of the DEGs was shown to be involved in transcriptional regulation. A number of the transcription factors were identified, such as *MYB*, *LBD*, *HD-ZIPIII*, *TCP*, *MADS*, *zf-HD*, Zinc finger protein, *NAM*, *WRKY*, *bZIP*, *bHLH*, *ANT*, *GRF*, *HB*, *AP2-EREBP* and *ERF*. The second subgroup was shown to be composed of 38 genes with homologies to genes that encode protein kinases, such as *CDKs*, *MAPK*, *MAPKK*, *MAPKKK*, *RLK* and *CDPK*. It is noteworthy that the expression of 21 members of the *CDKs* was down-regulated, whereas the expression of 8 members of the *RLKs* was up-regulated. The third subgroup contained 34 genes, including 15 calcium-binding proteins, 10 calcium-binding EF hand family proteins, 6 calmodulin-binding proteins and 3 calcium:cation antiporters. The final subgroup that was analyzed was related to auxin synthesis (*YUCCA*), transport (*PIN*), signaling (*AUX/IAA* and *ARF*), inactivation (*GH3*) and response (*SAUR*).

Table 1
Summary of read number.

Reads category	Map to genome		Map to annotation gene	
	RL	FL	RL	FL
Total reads	27,088,098	26,386,316	27,088,098	26,386,316
Total base pairs	2,437,928,820	2,374,768,440	2,437,928,820	2,374,768,440
Total mapped reads	19,158,357	18,610,411	16,869,340	16,735,761
Perfect match	13,569,026	13,163,404	11,401,649	11,347,040
≤2 bp mismatch	5,589,331	5,447,007	5,467,691	5,388,721
Unique match	18,398,107	17,826,502	16,143,576	15,953,181
Multi-position match	760,250	783,909	725,764	782,580
Total unmapped reads	7,929,741	7,775,905	10,218,758	9,650,555

2.3. Identification of splice variants and the discovery of novel transcripts

Alternative splicing is a mechanism that brings remarkable diversity to proteins and allows a gene to generate different mRNA transcripts that are translated into distinguishable proteins [13,14]. To assess the genome-wide AS events during leafy head development,

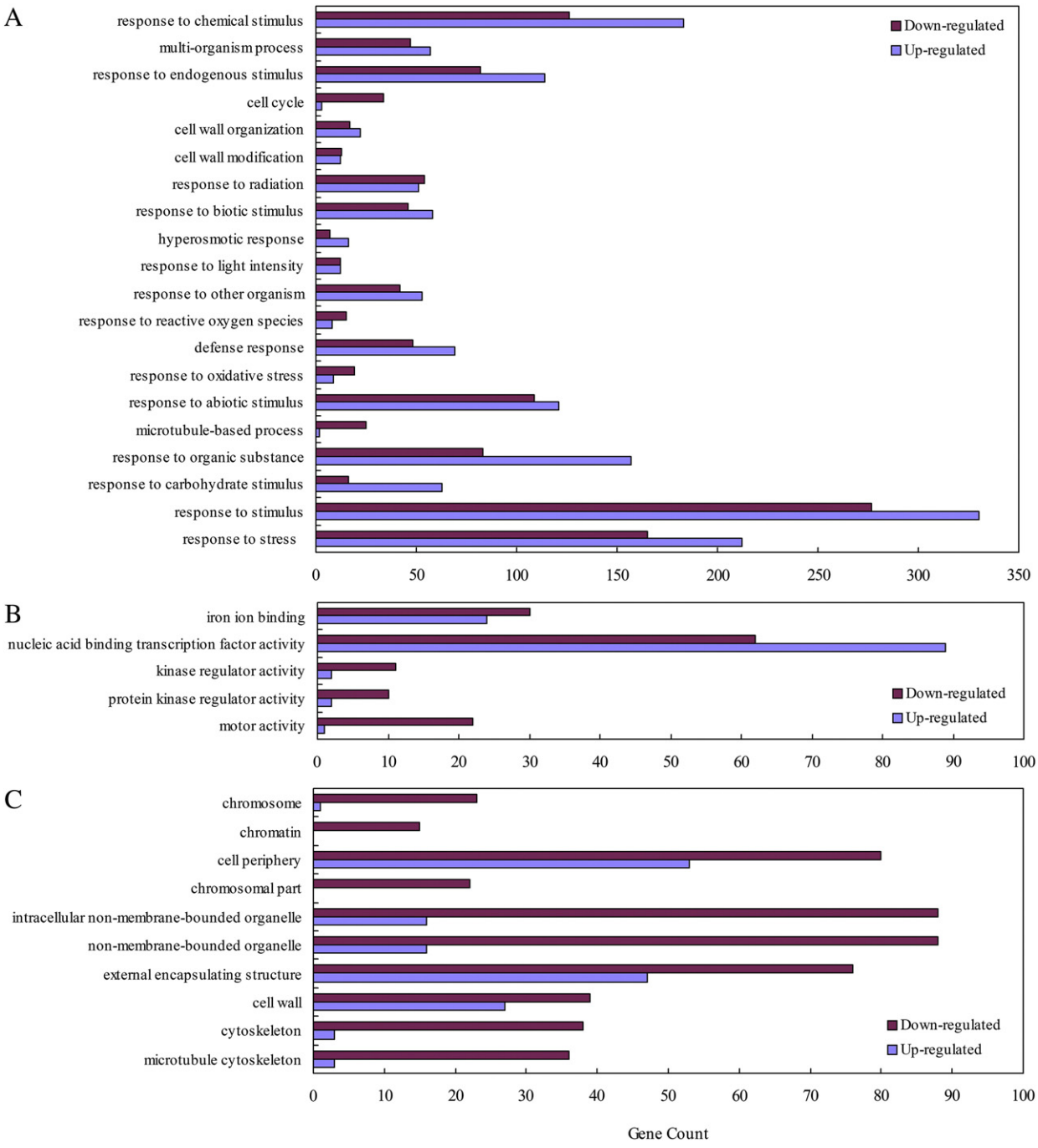

Fig. 1. GO enrichment analysis of differentially expressed genes during leafy heading development. A threshold of corrected p-value ≤ 0.05 was used to judge the significantly enriched GO terms in DEGs. A: Biological process; B: Molecular function; C: Cellular component.

we performed computational analyses to determine all the theoretical splicing junctions and then identified the sequence reads that mapped to these regions. We identified the following seven common types of AS: the alternative 3′ splice site (A3SS), the alternative 5′ splice site (A5SS), exon skipping (ES), the retention intron (RI), the alternative first exon (AFE), the alternative last exon (ALE) and the mutually exclusive exon (MXE). However, the AFE, ALE and MXE splicing mechanisms were not included in our report due to the high number of false positive results that are generated with the program. In this study, we examined the A3SS, A5SS, SE and RI splicing mechanisms and identified 2291, 1131, 147 and 952 events (detailed in Supplementary Table S3), respectively. Of these, 753 (32.87%) A3SS, 467 (41.29%) A5SS, 60 (40.82%) SE and 257 (27.00%) RI events existed in the RL only; 644 (28.11%) A3SS, 352 (31.12%) A5SS, 55 (37.41%) SE and 468 (49.16%) RI events existed in the FL only; and 894 (39.02%) A3SS, 312 (27.59%) A5SS, 32 (21.77%) SE and 227 (23.84%) RI events existed in both types of leaves (Fig. 2). Although the accurate validation of the alternative splicing events is beyond the scope of this investigation, we randomly selected the splicing of the Bra039214 sequence as an example. Because the sequence length differences between the differentially expressed forms of this gene are greater than 100 bp, they are easily distinguishable using agarose gel electrophoresis. According to our RNA-Seq analysis, this gene is constitutively expressed in the A3SS and the RI in three forms. The sequences of the alternative 3′ exon and the retained intron are represented in the public *B. rapa* ESTs database (the NCBI Acc. Nos. are

Table 2
Genes with important functions during plant development, including those encoding transcription factors, protein kinases, calcium signaling related proteins and auxin synthesis, transport and signaling related proteins and their expression patterns in FL.

Gene name	Number of up-regulated	Number of down-regulated
Transcrition factors		
MYB	10	7
LBD	4	1
HD-ZIPIII	0	1
TCP	0	1
MADS	3	0
zf-HD	0	5
Zinc finger protein	16	10
NAM	8	2
WRKY	10	2
bZIP	1	5
bHLH	7	9
ANT	0	1
GRF	0	5
HB	1	3
AP2-EREBP	7	3
ERF	14	3
Protein kinases		
CDK	0	21
MAPK	1	0
MAPKK	3	1
MAPKKK	2	1
RLK	8	0
CDPK	1	0
Calcium signaling related proteins		
CaM	20	5
CaMBP	4	2
CaCA	3	0
Auxin synthesis, transport and signaling related proteins		
AUX/IAA	4	2
SAUR	8	4
GH3	0	3
ARF	0	1
PIN	1	1
YUCCA	1	2

EX036706.1, EX125922.1 and EX024698.1), and the presence of the Bra039214 constitutive and alternative transcript variants was confirmed by reverse transcription (RT)-PCR (Fig. 3).

In a previous study, 41,174 protein-encoding genes were identified in the *B. rapa* genome [15], and these genes were used as the reference sequences for this study. Based on the transcriptome data, other sequence features that are not in the reference sequence annotations were observed. The gene models found in the intergenic regions (200 bp upstream or downstream of the genes) were considered candidate novel TUs. Based on the distribution of the sequence reads and the annotation of the reference genes, we detected 8162 novel TUs in the samples (Supplementary Table S4). Of these novel TUs, approximately 83.08% are shorter than 500 bp, but 3753 of them contain multiple exons.

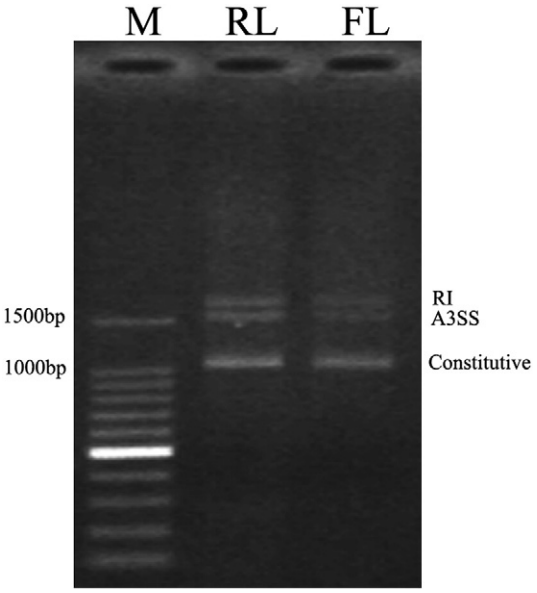

Fig. 3. RT-PCR confirmation of Bra039214 constitutive and alternative splice events in Chinese cabbage resettle leave (RL) and folding leave (FL) developmental stages. M is the DNA marker. The predicted fragment sizes are 1011 bp for the constitutive splice, 1542 bp for the A3SS splice and 1682 bp for the RI splice.

2.4. Gene boundary extension and polymorphism detection

To define the 5′ and 3′ gene boundaries more precisely, we investigated the upstream and downstream regions of the transcripts. In total, 3730 genes were extended at the 5′ end, 3319 genes were extended at the 3′ end, and 16,782 genes were extended at both ends (Supplementary Table S5).

RNA-Seq technology has the potential for comparing the reference genome and the study genome quickly and reliably (Chiifu and

Fig. 2. Frequency of identified splice variants in RL and FL libraries. Venn diagram of quantitative alternative splicing event classification into those specifically identified in one library or those identified in both libraries.

Fushanbaotou) and for identifying the expressed SNPs. For the initial sequence alignment and candidate SNP identification, we used the SOAPsnp software [16]. In the Chiifu reference genome sequences, we identified 22,145 SNPs, approximately 82.77% of which reside in coding regions, 1.51% reside in introns, 0.84% are located in 3′ UTRs, 0.65% are located in 5′ UTRs, and 14.23% reside in unknown regions. (Supplementary Table S6).

2.5. Validation of RNA-Seq-based gene expression by RT-qPCR

To validate the RNA-Seq results, an RT-qPCR analysis was performed using gene-specific primers for the top 5 up-regulated and top 5 down-regulated genes in the FL samples, which were identified by differential screening based on the RNA-Seq analysis (Supplementary Table S1). The results showed that all 10 genes exhibited the same expression profiles as the original RNA-Seq results (Fig. 4). In addition, of the top 5 up-regulated genes, 3 genes have unknown functions, one is homologous to the OB-fold protein, and one is homologous to the AP2-domain transcription factor. Of the top 5 down-regulated genes, one has an unknown function, and the other four genes are homologous to HSP17.6II, ATTI1, phosphatase and WOX1, respectively. The data suggest that the development of the leafy head is a complicated process and that further investigation of these genes, especially the up-regulated genes, may reveal their function in the regulation of leafy head formation.

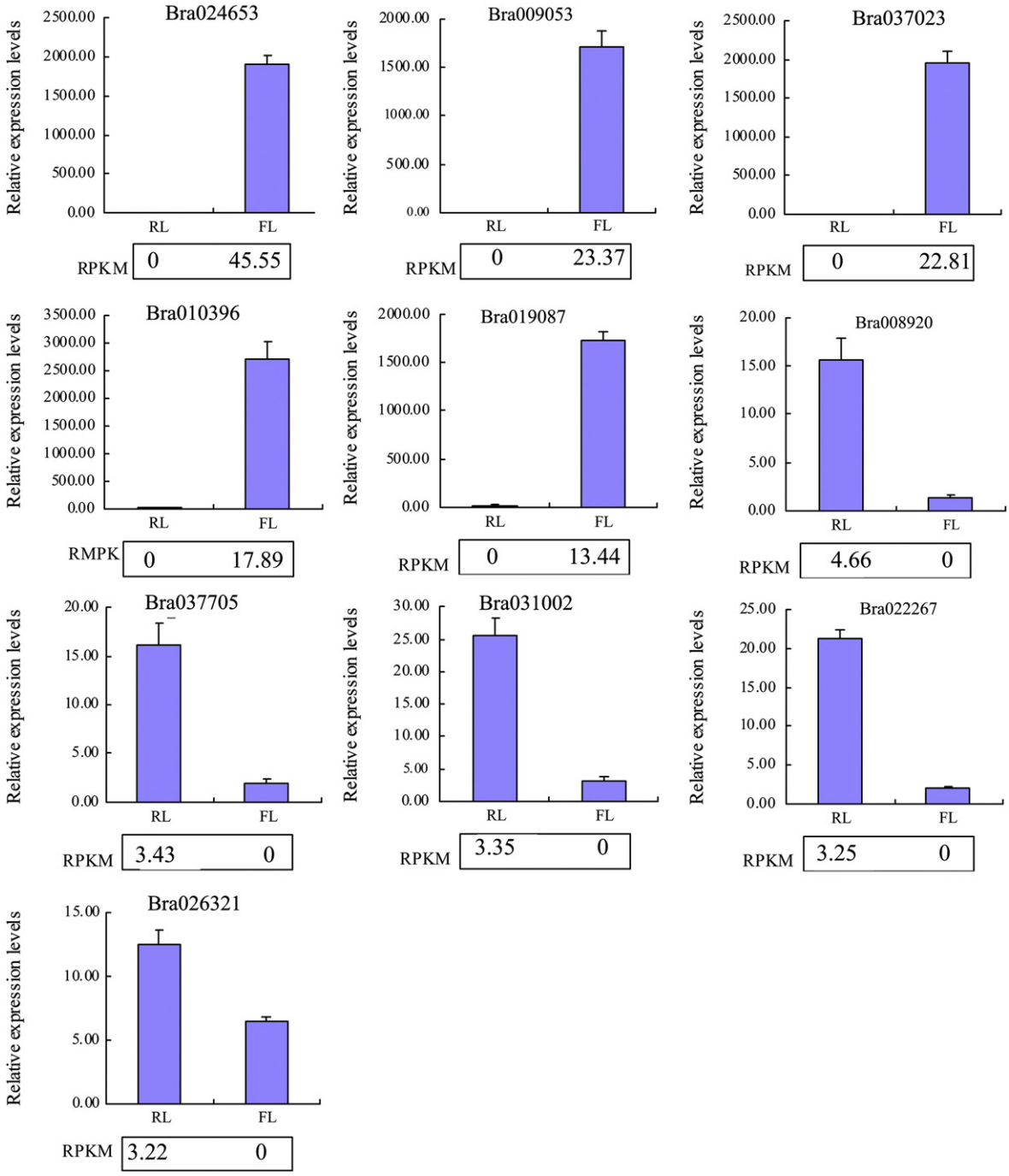

Fig. 4. RT-qPCR validation of the RNA-Seq based gene expression. The values indicate means of three biological replicates ± SD. RPKM, reads per kb per million reads.

3. Discussion

3.1. Involvement of differentially expressed genes in leaf heading development in Chinese cabbage

Analysis of the global gene expression pattern may provide a comprehensive understanding of the regulation mechanism of leafy head development. Previous studies have shown that the formation of the leafy head is affected by many factors, including the uneven distribution of auxin in the leaves, lower temperatures, a greater day–night temperature difference, weak light, a short day length and sufficient carbohydrate nutrition [2]; however, these studies only focused on physiological parameters. In this study, the GO process enrichment analysis revealed that there are hundreds of genes associated with the response to endogenous, carbohydrate, and light intensity stimuli, and other stimuli were significantly enriched (corrected p-value ≤ 0.05). These results suggest that these stimuli are needed for leafy head formation and confirm its [2] findings at a molecular level.

3.1.1. Differentially expressed genes involved in auxin synthesis, transport and signaling

Auxin plays a critical role in numerous plant developmental processes, including tropic responses, apical dominance, lateral root initiation, vascular differentiation, embryo patterning, and shoot elongation [17]. In addition, auxin functions in controlling leafy head formation in Chinese cabbage [2]. In this study, we found significant differences in the expression of genes relating to auxin synthesis (the *YUCCA* gene family), transport (the *PIN* gene family), signaling (the *AUX*/*IAA* and *ARF* gene families), inactivation (the *GH3* gene family) and the auxin response (the *SAUR* gene family). In plants, the AUX/IAA proteins and the ARF proteins are critical regulators of auxin-modulated gene expression [18,19]. The loss-of-function of the *AUX*/*IAA* gene *SHY2*/*IAA3* in *Arabidopsis* resulted in upward curved leaves [20]. The silencing of the *StIAA2* gene in the potato resulted in a number of growth alterations in the aerial parts of the plant, including the curvature of the leaf primordial in the shoot apex, increased plant height and petiole hyponasty [21]. In plants, auxin exists as an active free form and an inactive bound form in which the carboxyl group is conjugated to sugars via ester linkages or to amino acids or peptides via amide linkages [22]. *GH3* genes encode a class of auxin-induced conjugating enzymes, which play a role in modulating the endogenous levels of active auxin through a negative feedback regulation in seedlings [23]. In this study, the RNA-Seq analysis showed that 4, 8, 1 and 1 members belong to the *AUX*/*IAA*, *SAUR*, *PIN* and *YUCCA* gene families, respectively, were significantly up-regulated in the FL, while 2, 4, 3, 1, 1 and 2 members belonging to the *AUX*/*IAA*, *SAUR*, *GH3*, *ARF*, *PIN* and *YUCCA* gene families, respectively, were significantly down-regulated in the FL. These results suggested that auxin might affect the synthesis, transport and signaling processes during leafy head formation. The decrease in the mRNA level of the *GH3* gene indicated that it might promote leafy head formation by increasing the free-auxin concentration in seedlings.

3.1.2. Differentially expressed genes encoding transcription factor proteins

Transcription factors are a group of DNA-binding proteins that interact with other transcriptional regulators to recruit or block the access of RNA polymerases to the DNA template, which plays an important role in controlling plant development and differentiation. Recently, the *AS1* [24,25], *RDL1* [26–28], *ICU4* [29,30], *CIN* [31] and *AS2* [32] genes, which play an important role in controlling leaf curvature or polarity, were confirmed by mutational analysis. These genes are members of the following transcription factor families: *MYB* (*ASI*), *HD-ZIPIII* (*RDL1* and *ICU4*), *TCP* (*CIN*) and *LBD* (*AS2*). In this study, the profiling of the genes encoding the MYB, HD-ZIPIII, TCP and LBD family member proteins in the Chinese cabbage genome using the RNA-Seq method indicated that the expression levels of some of the genes differed between the RL and the FL and were significantly up- or down-regulated. In addition, a number of members of the zinc finger protein family, the bHLH family, the NAM family, the ANT family and the GRF family were differentially expressed. These genes play important roles in regulating plant development. For example, the *Arabidopsis JAGGED* gene encodes a zinc finger protein and may promote leaf tissue development [33]. Additionally, the bHLH transcription factors mediate the brassinosteroid regulation of cell elongation and plant development in rice and *Arabidopsis* [34], and the *NAM* gene from *Petunia* is required for pattern formation in embryos and flowers and is expressed at the meristem and primordia boundaries [35]. The results indicated that these TFs may play critical roles in controlling leafy head development and further investigation of these genes may reveal their function in the regulation of leafy head formation.

3.1.3. Differentially expressed genes encoding protein kinases

In plants, protein phosphorylation has been implicated in responses to many signals, including light, pathogen invasion, hormones, temperature stress, and nutrient deprivation [36]. Protein kinases are a group of enzymes that catalyze protein phosphorylation. In this study, 38 genes including 21 *CDKs*, 1 *MAPK*, 4 *MAPKK*, 3 *MAPKKK*, 8 *RLK* and 1 *CDPK* were significantly up-regulated or down-regulated in the RL and the FL. This result suggests that protein phosphorylation events occur during leafy head development. Plant development requires the stringent control of cell proliferation and cell differentiation. In some cases, the proliferation is positively regulated by *CDKs* [37], *GRF* [38] and *ANT* [39]; however, the expression of the 21 *CDK*, 5 *GRF* and 1 *ANT* genes characterized in this study was down-regulated. Therefore, it is possible that cell proliferation does not play a critical role in the control of leafy head formation, or other genes, such as RLKs and MAPK, are involved this regulation process.

3.1.4. Differentially expressed genes encoding calcium signaling related proteins

Calcium is a ubiquitous secondary messenger in eukaryotic signal transduction cascades. Many external stimuli including light and various other stress factors can alter the cellular Ca^{2+} level, which can affect plant growth and development [40]. Calcium is also a critical nutrient for normal development in the Chinese cabbage, and calcium deficiency leads to leaf tipburn [41]. In plants, three known classes of calcium sensors, including calmodulins (CaM), CDPKs, and calcineurin B-like proteins, recognize specific calcium signatures and transduce them into downstream effects, including altered protein phosphorylation and gene expression patterns. In this study, a series of genes including CaM, calmodulin-binding protein (CaMBP) and calcium:cation antiporter (CaCA), which are related to calcium signal transduction cascades, were significantly up-regulated or down-regulated. Meanwhile, the expression of one *CDPK* gene was up-regulated, suggesting that calcium affects leafy head development and is mediated by this *CDPK* gene.

3.2. Improve the annotation of the existing genes

Our transcriptome data also improves the existing gene annotation in a variety of ways, including providing evidence for numerous AS events, novel TUs, SNP identification and the extension of gene boundaries. Alternative splicing is an efficient way for genomes to encode more transcripts [13,14]. The identification of the AS patterns of Chinese cabbage genes during leafy head development will be helpful for understanding the mechanisms of their transcriptional control. The extension of the 5′ and 3′ end of the genes will be valuable for determining the intact gene boundaries, which is helpful for finding genomic loci containing transcripts. In this study, the dataset of

53.5 million sequence reads (90 bp in length), corresponding to 4.85 Gb of raw sequence data, provides a great resource for the improvement of gene annotations across the Chinese cabbage genome. Our results showed that 37.8 million reads (approximately 70.6%) were successfully mapped to the scaffold sequence of the Chinese cabbage genome, which led to the identification of 8162 novel transcripts, 4521 AS events and 23,831 genes with transcribed sequences extended at the 5′end, the 3′end or at both ends.

In this study, 22,145 SNPs were detected between the reference genome and the studied genome (Chiifu and Fushanbaotou). This result may reflect the real number of SNPs between these two Chinese cabbage subspecies. Other factors that may contribute to the creation of SNPs include RNA editing, sequence errors, mapping errors or reference sequencing errors [9]. RNA editing is an unusual phenomenon in which nucleotides in the mRNA transcript are modified post-transcriptionally and result in a single nucleotide mismatch between the transcript and its genomic template. RNA editing has been reported in a wide spectrum of eukaryotes [42–44]. Adenine (A) to inosine (I) (A-to-I) and cytidine (C) to uridine (U) (C-to-U) are two common forms of RNA editing [45]. Unfortunately, based on the current data, we could not confirm how many of the SNPs were caused by RNA editing. A better method for detecting and analyzing RNA editing may be to use both genomic DNA and RNA from the same genetic background for deep sequencing with a higher base quality [9].

Due to the high cost of the RNA-Seq analysis, our analysis was limited to a single biological sample of pooled tissue from multiple seedlings of Chinese cabbage. Although a global analysis of gene expression and the profound improvement of the existing gene annotation based on a single replication do not allow a solid biological interpretation, this RNA-Seq analysis clearly provided extensive novel information about the Chinese cabbage transcriptome during leafy head development. Further functional analysis of the set of differentially transcribed genes during leafy head development will help us to elucidate the mechanism of leafy head development and may greatly facilitate the breeding practices at a molecular level.

4. Materials and methods

4.1. Plant materials

The inbred line Fushanbaotou, a typical heading Chinese cabbage, was chosen for this study. The seeds from this inbred line were carefully selected and sowed in the field at the *Vegetable Research Institute, Shandong Academy of Agricultural Sciences* farm station at the normal sowing time in Jinan (August 15, 2010). For the RL materials, the developing RL were cut at approximately 5 mm under the shoot tips from plants with 8–10 expanded leaves at the early rosette stage. For the FL materials, plants at the early folding stage with 23–25 expanded leaves and with the young leaves beginning to fold were marked for sampling, and the developing FL were harvested approximately 5 mm under the shoot tips. The developing leaves that were collected from fifteen seedlings for each developmental stage were pooled together and were stored in liquid nitrogen for mRNA extraction.

4.2. Preparation of cDNA library for RNA-sequencing

For Illumina sequencing, the total RNA was extracted from every sample using Trizol (Invitrogen) and was treated with RNase-free DNase I (TaKaRa) for 45 min according to the manufacturer's protocol. The poly(A) mRNA was purified from 20 mg of the total RNA samples using Sera-mag Magnetic Oligo(dT) Beads (Illumina). Because RNA fragmentation produces a more even sequence read distribution than cDNA fragmentation [46], the mRNA was first sheared into short fragments using an RNA fragmentation kit (Ambion) before cDNA synthesis. Using these short mRNA fragments as templates, first strand cDNAs were synthesized using random hexamer primers and reverse transcriptase (Invitrogen). The second-strand cDNA was synthesized using DNA polymerase I, and RNase H was used to remove the RNA. For high-throughput sequencing, the sequence library was constructed following the manufacturer's instructions (Illumina). Fragments of ~300 bp were excised and were enriched by PCR for 18 cycles. The products were loaded onto flow cell channels at a concentration of 2 pM for pair-end 90 bp × 2 sequencing. The Illumina HiSeq™ 2000 platform was used for the sequencing.

4.3. Mapping short reads to the Chinese cabbage genome and the annotated gene

After removing the sequence reads containing sequencing adapters and low quality sequence reads (reads containing more than 5% unknown bases or more than half of their bases with a quality of less than 5), we aligned the reads to the Chiifu Chinese cabbage genome and annotated genes (http://brassicadb.org/brad/) using the SOAP2 software [11], allowing up to two mismatches. The number of reads that were fully located in exons was counted, and the expression level of each gene was determined by calculating the total number of hits in its corresponding exons.

4.4. Gene expression pattern analysis

The unigene expression was calculated according to the method of RPKM (reads per kb per million reads) [47].

To identify the differentially expressed genes (DEGs) in the two samples, a protocol from Audic and Claverie [12] was used. The false discovery rate (FDR) was used to determine the threshold of the p-value in multiple tests, and for the analysis, a threshold of the FDR ≤ 0.001 and an absolute value of $\log_2$Ratio ≥ 1 were used to judge the significance of the gene expression differences [48].

The DEGs were subjected to a gene ontology (GO) analysis. For the GO analysis, the DEGs were first mapped to the GO terms in the database (http://www.geneontology.org/), the gene number was calculated for every term, and the ultra-geometric test was used to find significantly enriched GO terms in the DEGs compared to the genome background. The calculated p-value was determined using the Bonferroni correction, taking the corrected-p-value ≤ 0.05 as a threshold. The GO terms satisfying these conditions are defined as significantly enriched GO terms in the DEGs.

4.5. Alternative splicing analysis

The AS events in the Chinese cabbage were identified using TopHat with all the default parameters to detect the junction sites, which contain information about the boundaries and the combinations of the different exons in a transcript. All the junction sites of the same gene are used to distinguish the type of AS events, including skipped exons (SE), retained introns (RI), alternative 5′-splice sites (A5SS) and alternative 3′-splice sites (A3SS), according to the method of Trapnell [49]. To validate the alternative splicing events, the Bra039214 gene was randomly selected and was subjected to RT-PCR and Sanger sequencing methods. The primers for the RT-PCR amplification of the Bra039214 gene were as follows: forward primer 5′-ATGGCGATGAGGAAGCTTTTGAC-3′ and reverse primer 5′-CTAACTCTGGATAGTTCCAAGAA-3′.

4.6. Identification of novel transcripts

A novel transcriptional active region (TAR) was defined in the intergenic regions (200 bp upstream or downstream of the genes) by contiguous expression with each base supported by at least two uniquely mapped reads and lengths > 35 bp. Supported by the paired-end information, a novel TU was constructed by the

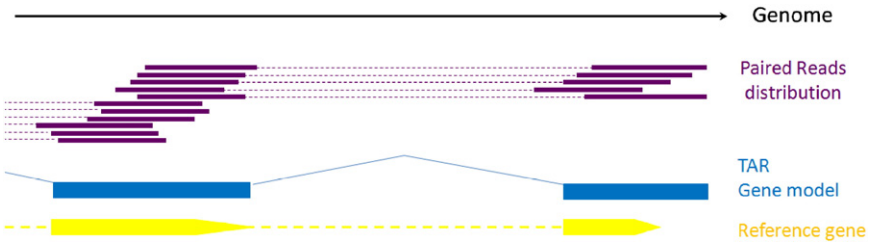

Fig. 5. Gene boundary refining algorithm-sketch.

connectivity between novel TARs, which were joined by at least one paired read. The novel TUs of >150 bp and an average expression of >2 reads per base were counted as candidates for novel transcripts.

4.7. Gene boundary determination

Gene structures were optimized according to the distribution of the reads, the paired-end information and the annotation of the reference gene. First, we obtained the distribution of the reads in the genome by aligning the continuous and overlapping reads to form a TAR. Then, according to the paired-end data, we connected the different TARs to form a potential gene model. Finally, we compared the gene model with the existing annotated gene sequences to extend the 5′ and 3′ ends of the gene. (Fig. 5 indicates the algorithm-sketch of extending the gene ends.)

4.8. SNP calling

The SOAPsnp software [16] was used for SNP discovery between the reference genome and the studied genome (Chiifu and Fushanbaotou). The potential SNPs were identified using the criteria of a minimum base-call quality of 10, an average quality of 20, and minimum best hits of four. The SNP was not allowed to be an ambiguous base (e.g., SNP ≠ "N").

4.9. Real-time quantitative PCR validation of the DEGs

First-strand cDNA synthesis was performed as described above. The transcriptional profiles of the 10 selected genes were analyzed by real-time quantitative PCR (RT-qPCR) using the SYBR Green PCR master mix (Takara) and the IQ5 real-time PCR system (BIO-RAD). The gene-specific primers designed for the 10 genes are listed in Supplementary Table S7. The *actin* gene was used as a constitutive expression control in these experiments. The PCR-cycling conditions comprised an initial polymerase activation step at 95 °C for 1 min, followed by 40 cycles at 95 °C for 10 s and 60 °C for 30 s. After each PCR run, a dissociation curve was designed to confirm the specificity of the product and to avoid the production of primer dimers. The relative amounts of the amplification products were calculated based on the standard curve analysis generated from each standard cDNA. The reactions were performed in triplicate, and each sample was further amplified without reverse transcriptase to confirm the absence of DNA contamination in the samples.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2012.02.005.

Acknowledgments

The reference sequence data were produced by the Chinese *Brassica rapa* genome sequencing project. This work was supported by the Special Prophase Project on the National Basic Research Program of China (2009CB126002); the National Nature Science Foundation of China (31101553); the Promotive Research Fund for Excellent Young and Middle-aged Scientists of Shandong Province, China (BS2010SW027); the China Postdoctoral Science Foundation Funded Project (20100481298) and the Modern Agricultural Industrial Technology System Funding of Shandong Province, China (NYXD2011).

References

[1] X. Yu, J. Peng, X. Feng, S. Yang, Z. Zheng, X. Tang, R. Shen, P. Liu, Y. He, Cloning and structural and expressional characterization of *BcpLH* gene preferentially expressed in folding leaf of Chinese cabbage, Sci. China C 43 (2000) 321–329.
[2] H. Ito, Effect of temperature and photoperiod on head formation of leafy head of Chinese cabbage, J. Hortic. Assoc. Jpn. 26 (1957) 154.
[3] G. Ke, Chinese Cabbage Breeding, China Agricul. Press, Beijing, 2010 (in Chinese).
[4] J. Guo, N. Zhou, R. Ma, M. Cao, mRNA differential display analysis at rosette and folding stage of Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*), J. Agric. Biotechnol. 11 (2003) 456–460 (in Chinese with English abstract).
[5] R.J. Gao, D.P. Dai, R.C. Ma, M.Q. Cao, Y.M. Yan, Y.D. Wang, S.J. Ren, X.Y. Guo, Expressed sequence tags (EST) analysis of the heading leaf of Chinese cabbage (*Brassica rapa* L. *ssp. pekinensis*) at the early heading stage, J. Agric. Biotechnol. 12 (2004) 24–29 (in Chinese with English abstract).
[6] V. Costa, C. Angelini, I.D. Feis, A. Ciccodicola, Uncovering the complexity of transcriptomes with RNA-Seq, J. Biomed. Biotechnol. 2010 (2010) 853916.
[7] O. Morozova, M.A. Marra, Applications of next-generation sequencing technologies in functional genomics, Genomics 92 (2008) 255–264.
[8] R. Lister, R.C. O'Malley, J. Tonti-Filippini, B.D. Gregory, C.C. Berry, A.H. Millar, J.R. Ecker, Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*, Cell 133 (2008) 523–536.
[9] T. Lu, G. Lu, D. Fan, C. Zhu, W. Li, Q. Zhao, Q. Feng, Y. Zhao, Y. Guo, W. Li, et al., Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq, Genome Res. 20 (2010) 1238–1249.
[10] S. Zenoni, A. Ferrarini, E. Giacomelli, L. Xumerle, M. Fasoli, G. Malerba, D. Bellin, M. Pezzotti, M. Delledonne, Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq, Plant Physiol. 152 (2010) 1787–1795.
[11] R. Li, C. Yu, Y. Li, T.W. Lam, S.M. Yiu, K. Kristiansen, J. Wang, SOAP2: an improved ultrafast tool for short read alignment, Bioinformatics 25 (2009) 1966–1967.
[12] S. Audic, J.M. Claverie, The significance of digital gene expression profiles, Genome Res. 7 (1997) 986–995.
[13] D.L. Black, Mechanisms of alternative pre-messenger RNA splicing, Annu. Rev. Biochem. 72 (2003) 291–336.
[14] S. Stamm, S. Ben-Ari, I. Rafalska, Y. Tang, Z. Zhang, D. Toiber, T.A. Thanaraj, H. Soreq, Function of alternative splicing, Gene 344 (2005) 1–20.
[15] X. Wang, H. Wang, J. Wang, R. Sun, J. Wu, S. Liu, Y. Bai, J.H. Mun, I. Bancroft, F. Cheng, et al., The genome of the mesopolyploid crop species *Brassica rapa*, Nat. Genet. 43 (2011) 1035–1039.
[16] R. Li, Y. Li, X. Fang, H. Yang, J. Wang, K. Kristiansen, J. Wang, SNP detection for massively parallel whole-genome resequencing, Genome Res. 19 (2009) 1124–1132.
[17] J. Friml, Auxin transport-shaping the plant, Curr. Opin. Plant Biol. 6 (2003) 7–12.
[18] T. Guilfoyle, T. Ulmasov, G. Hagen, The *ARF* family of transcription factors and their role in plant hormone-responsive transcription, Cell. Mol. Life Sci. 54 (1998) 619–627.
[19] L. Walker, M. Estelle, Molecular mechanisms of auxin action, Curr. Opin. Plant Biol. 1 (1998) 434–439.
[20] Q. Tian, J.W. Reed, Control of auxin-regulated root development by the *Arabidopsis thaliana SHY2/IAA3* gene, Development 126 (1999) 711–721.
[21] B. Kloosterman, R.G. Visser, C.W. Bachem, Isolation and characterization of a novel potato *Auxin/Indole-3-Acetic Acid* family member (*StIAA2*) that is involved in petiole hyponasty and shoot morphogenesis, Plant Physiol. Biochem. 44 (2006) 766–775.
[22] J.D. Cohen, R.S. Bandurski, Chemistry and physiology of the bound auxins, Annu. Rev. Plant Physiol. 33 (1982) 403–430.
[23] P.E. Staswick, B. Serban, M. Rowe, I. Tiryaki, M.T. Maldonado, M.C. Maldonado, W. Suza, Characterization of an Arabidopsis enzyme family that conjugates amino acids to indole-3-acetic acid, Plant Cell 17 (2005) 616–627.
[24] M.E. Byrne, R. Barley, M. Curtis, J.M. Arroyo, M. Dunham, A. Hudson, R.A. Martienssen, Asymmetric leaves1 mediates leaf patterning and stem cell function in *Arabidopsis*, Nature 408 (2000) 967–971.
[25] Y. Sun, Q. Zhou, W. Zhang, Y. Fu, H. Huang, *ASYMMYTRIC LEAVES1*, an *Arabidopsis* gene that is involved in the control of cell differentiation in leaves, Planta 214 (2002) 694–702.

[26] J.O. Hay, B. Moulia, B. Lane, M. Freeling, W.K. Silk, Biomechanical analysis of the *Rolled* (*RLD*) leaf phenotype of maize, Am. J. Bot. 87 (2000) 625–633.
[27] M.T. Juarez, J.S. Kui, J. Thomas, B.A. Heller, M.C. Timmermans, MicroRNA-mediated repression of rolled *leaf1* specifies maize leaf polarity, Nature 428 (2004) 84–88.
[28] J.M. Nelson, B. Lane, M. Freeling, Expression of a mutant maize gene in the ventral leaf epidermis is sufficient to signal a switch of the leaf's dorsoventral axis, Development 129 (2002) 4581–4589.
[29] D. Otsuga, B. DeGuzman, M.J. Prigge, G.N. Drews, S.E. Clark, *REVOLUTA* regulates meristem initiation at lateral positions, Plant J. 25 (2001) 223–236.
[30] J. Serrano-Cartagena, H. Candela, P. Robles, M.R. Ponce, J.M. Perez-Perez, P. Piqueras, J.L. Micol, Genetic analysis of incurvata mutants reveals three independent genetic operations at work in *Arabidopsis* leaf morphogenesis, Genetics 156 (2000) 1363–1377.
[31] U. Nath, B.C.W. Crawford, R. Carpenter, E. Coen, Genetic control of surface curvature, Science 299 (2003) 1404–1407.
[32] L. Xu, Y. Xu, A. Dong, Y. Sun, L. Pi, Y. Xu, H. Huang, Novel *as1* and *as2* defects in leaf adaxial-abaxial polarity reveal the requirement for *ASYMMETRIC LEAVES1* and *2* and *ERECTA* functions in specifying leaf adaxial identity, Development 130 (2003) 4097–4107.
[33] C.K. Ohno, G.V. Reddy, M.G.B. Heisler, E.M. Meyerowitz, The *Arabidopsis JAGGED* gene encodes a zinc finger protein that promotes leaf tissue development, Development 131 (2004) 1111–1122.
[34] L.Y. Zhang, M.Y. Bai, J. Wu, J.Y. Zhu, H. Wang, Z. Zhang, W. Wang, Y. Sun, J. Zhao, X. Sun, et al., Antagonistic HLH/bHLH transcription factors mediate brassinosteroid regulation of cell elongation and plant development in rice and Arabidopsis, Plant Cell 21 (2009) 3767–3780.
[35] E. Souer, A. van Houwelingen, D. Kloos, J. Mol, R. Koes, *The No Apical Meristem* Gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries, Cell 85 (1996) 159–0170.
[36] J.M. Stone, J.C. Walker, Plant protein kinase families and signal transduction, Plant Physiol. 108 (1995) 451–457.
[37] S. Jasinski, C. Riou-Khamlichi, O. Roche, C. Perennes, C. Bergounioux, N. Glab, The CDK inhibitor NtKIS1a is involved in plant development, endoreduplication and restores normal development of cyclin D3;1-overexpressing plants, J. Cell Sci. 115 (2002) 973–982.
[38] G. Horiguchi, G.T. Kim, H. Tsukaya, The transcription factor *AtGRF5* and the transcription coactivator *AN3* regulate cell proliferation in leaf primordial of Arabidopsis thaliana, Plant J. 43 (2005) 68–78.
[39] Y. Mizukami, R.L. Fischer, Plant organ size control: *AINTEGUMENTA* regulates growth and cell numbers during organogenesis, PNAS 97 (2000) 942–947.
[40] P.K. Hepler, Calcium: a central regulator of plant growth and development, Plant Cell 17 (2005) 2142–2155.
[41] C.G. Kuo, J.S. Tsay, C.L. Tsai, R.J. Chen, Tipburn of Chinese cabbage in relation to calcium nutrition and distribution, Sci. Hortic. 14 (1981) 131–138.
[42] J.M. Gott, R.B. Emeson, Functions and mechanisms of RNA editing, Annu. Rev. Genet. 34 (2000) 499–531.
[43] E. Picardi, D.S. Horner, M. Chiara, R. Schiavon, G. Valle, G. Pesole, Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing, Nucleic Acids Res. 38 (2010) 4755–4767.
[44] S. Steinhauser, S. Beckert, I. Capesius, O. Malek, V. Knoop, Plant mitochondrial RNA editing, J. Mol. Evol. 48 (1999) 303–312.
[45] S. Gopal, Computational prediction of RNA editing sites: successes and challenges ahead, Curr. Bioinform. 3 (2008) 162–177.
[46] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (2009) 57–63.
[47] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nat. Methods 5 (2008) 621–628.
[48] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann. Stat. 29 (2001) 1165–1188.
[49] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinformatics 25 (2009) 1105–1111.