



OPEN

# Transcriptome Analysis by RNA–Seq Reveals Genes Related to Plant Height in Two Sets of Parent-hybrid Combinations in Easter lily (*Lilium longiflorum*)

Jewel Howlader<sup>1,2</sup>, Arif Hasan Khan Robin<sup>1,3</sup>, Sathishkumar Natarajan<sup>1</sup>, Manosh Kumar Biswas<sup>1</sup>, Kanij Rukshana Sumi<sup>4,5</sup>, Cheon Young Song<sup>6</sup>, Jong–In Park<sup>1</sup> & Ill–Sup Nou<sup>1</sup>✉

In this study, two different hybrids of Easter lily (*Lilium longiflorum*), obtained from two cross combinations, along with their four parents were sequenced by high–throughput RNA–sequencing (RNA–Seq) to find out differentially expressed gene in parent-hybrid combinations. The leaf mRNA profiles of two hybrids and their four parents were RNA–sequenced with a view to identify the potential candidate genes related to plant height heterosis. In both cross combinations, based to morphological traits mid–parent heterosis (MPH) was higher than high–parent heterosis (HPH) for plant height, leaf length, and number of flowers whereas HPH was higher than MPH for flowering time. A total of 4,327 differentially expressed genes (DEGs) were identified through RNA–Seq between the hybrids and their parents based on fold changes (FC)  $\geq 2$  for up– and  $\leq -2$  for down–regulation. Venn diagram analysis revealed that there were 703 common DEGs in two hybrid combinations, those were either up– or down–regulated. Most of the commonly expressed DEGs exhibited higher non–additive effects especially overdominance (75.9%) rather than additive (19.4%) and dominance (4.76%) effects. Among the 384 functionally annotated DEGs identified through Blast2GO tool, 12 DEGs were up–regulated and 16 of them were down–regulated in a similar fashion in both hybrids as revealed by heat map analysis. These 28 universally expressed DEGs were found to encode different types of proteins and enzymes those might regulate heterosis by modulating growth, development and stress–related functions in lily. In addition, gene ontology (GO) analysis of 260 annotated DEGs revealed that biological process might play dominant role in heterotic expression. In this first report of transcriptome sequencing in Easter lily, the notable universally up–regulated DEGs annotated ABC transporter A family member–like, B3 domain–containing, disease resistance RPP13/1, auxin–responsive SAUR68–like, and vicilin–like antimicrobial peptides 2–2 proteins those were perhaps associated with plant height heterosis. The genes expressed universally due to their overdominance function perhaps influenced MPH for greater plant height—largely by modulating biological processes involved therein. The genes identified in this study might be exploited in heterosis breeding for plant height of *L. longiflorum*.

Heterosis refers to the higher performance in any trait (s) of interest of F<sub>1</sub> hybrid as compared to homozygous parental lines. Charles Darwin<sup>1</sup> described this phenomenon for the first time ever and later Shull<sup>2</sup> and East<sup>3</sup> independently confirmed this phenomenon. Improvements in many traits of different crop and livestock species have

<sup>1</sup>Department of Horticulture, Suncheon National University, 255, Jungang–ro, Suncheon, Jeonnam, 57922, Republic of Korea. <sup>2</sup>Department of Horticulture, Patuakhali Science and Technology University, Dumki, Patuakhali, 8602, Bangladesh. <sup>3</sup>Department of Genetics and Plant Breeding, Bangladesh Agricultural University, Mymensingh, 2202, Bangladesh. <sup>4</sup>Department of Fisheries Science, Chonnam National University, 50, Daehak–ro, Yeosu, Jeonnam, 59626, Republic of Korea. <sup>5</sup>Department of Aquaculture, Patuakhali Science and Technology University, Dumki, Patuakhali, 8602, Bangladesh. <sup>6</sup>Department of Floriculture, Korea National College of Agriculture and Fisheries, 1515, Kongjwipatjiwi–ro, Wansan–gu, Jeonju–si, Jeollabuk–do, 54874, Republic of Korea. ✉e-mail: [nis@suncheon.ac.kr](mailto:nis@suncheon.ac.kr)

been achieved by exploiting heterosis during the last centuries<sup>4–8</sup>. Numerous investigations attempted generalizing the molecular basis of heterosis<sup>9</sup>. Despite, a conclusive molecular mechanism of heterosis remains elusive<sup>10</sup>.

Heterosis could be either positive or negative depending on traits of breeding interest. For example, the yield is a positive heterotic trait and earliness is a negative heterotic trait<sup>11–13</sup>. There are two classic hypotheses namely dominance hypothesis and overdominance hypothesis those are used to differentiate allelic variations in expression in between hybrids and inbred parents<sup>14</sup>. In dominance hypothesis, alleles related to the desired trait of interest from the one parent suppress the alleles linked to less desired trait of interest from another parent. In contrast, overdominance heterosis occurs due to the simultaneous action of two parental alleles resulting different superior traits in a hybrid. These two hypotheses are redefined in relation to additivity and non-additivity to classify differentially expressed genes (DEGs) between parent and hybrids<sup>15,16</sup>. In additive gene effect, the expression level of DEGs in hybrid is similar to the average of both parents (mid-parental expression) while in non-additive effect, the expression level is different from the parental mean. Non-additive gene effect has also been classified into high parent dominance (high parent-like expression), low parent dominance (low parent-like expression), overdominance (above high parent expression value) and underdominance (below low parent expression value)<sup>15</sup>. Experimental evidences show that heterosis in most cases manifested by the incidence of non-additive gene expression<sup>17–22</sup>. In a few cases, however, this phenomenon was also reported to be arisen from additive gene expression levels<sup>20,22,23</sup>.

Next-generation sequencing (NGS) especially high-throughput RNA sequencing (RNA-seq) technology is advantageously used for discovering heterotic genes over other expression profiling technologies in plants and animals<sup>13,24–26</sup>. In plants, using RNA-Seq technology, expression patterns of DEGs between hybrids and their parents can be analyzed that might contribute to understanding the genetic basis of heterosis<sup>13,27,28</sup>. Transcriptome analyses revealed that in nascent hexaploid wheat dominance gene expression was predominant for allopolyploid heterosis<sup>15</sup> but in tobacco, overdominance was key factor for nicotine biosynthesis<sup>29</sup>. Dominance and overdominance effects were displayed by heterotic genes associated with the development of ears earlier in maize inflorescence<sup>30</sup>. In chrysanthemum, two flowering traits—initial blooming time and duration of flowering are regulated by the presence of two pairs of major genes where additivity is predominant<sup>31</sup>. By contrast, the non-additivity might be related to early vegetative development, increased photosynthesis, cell size and number in hybrids that might play the key roles of contribution to the biomass heterosis in *Arabidopsis*<sup>21,32</sup>.

Heterosis in a hybrid as estimated by vegetative growth, flowering time, yield and resistance to biotic and abiotic stresses was shown to be regulated by differential gene expression patterns in plants<sup>4</sup>. In contrast to self-fertilized plants, cross-fertilized one exhibits 10–200% higher heterosis in wheat and maize. Interestingly, when two genetically diverse alleles meet and attain a heterozygous state, the resultant genotypes experience a period of genomic turbulence a so called ‘*Genomic Shock*’ until a stable homeostasis is established<sup>33,34</sup>. This genomic turbulence might trigger the wide range of regulatory genes with differential gene expression patterns<sup>35</sup>. Heterozygosity of the regulating genomic regions that brings about superiority in hybrids in case of most of the quantitative traits could be either monogenic or polygenic<sup>36</sup>. For example, superior hybrid performance in relation to several agronomic traits in tomato<sup>37</sup> and lily<sup>38</sup> might be resulted from a single heterozygous gene linked to heterosis. In contrast, performance of most economically interesting traits of maize<sup>39</sup>, *Arabidopsis*<sup>40</sup> and rapeseed<sup>41</sup> might be regulated by multiple heterotic genes.

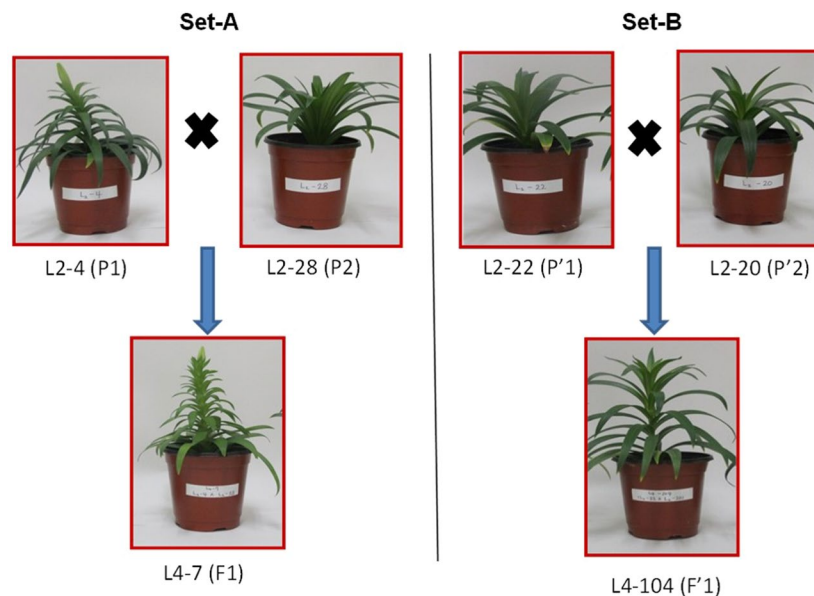
Lilies (*Lilium* spp.,  $2n = 24$ ), the bulbous monocots, are the outcrossing perennial herbs of the Liliaceae family. Lilies have high global market due to their incomparable beauty and commercial value<sup>42,43</sup>. Some Asian countries including China, Korea, Japan and Nepal are considered as one of the important center of diversity of lilies<sup>44</sup>. Lily is a highly popular cut flower in Korea. Annual production of lilies in Korea has an economic value of about US\$ 34 million<sup>45,46</sup>. In lily, modern hybrid breeding exploit hybrid vigor of inter-sectional hybridization of Longiflorum, Oriental and Asiatic hybrids<sup>47</sup>. Breeders selects those hybrids for plant height; number of leaves, flowers and bulbils; leaf area; flower size; days to flowering and petal area which indicated that those traits were often found heterotic<sup>48–50</sup>.

Despite the economic and aesthetic importance of lilies, no attempt was made to date that investigate the heterotic genes related to any phenotypic traits in *Lilium* spp. Besides, RNA-Seq facility may help identify heterotic genes to understand types of proteins or functional domains of the genes are involved in heterosis. In this study, we investigated the heterosis for plant height in Easter lily (*Lilium longiflorum*) by performing RNA-Seq analyses in two hybrids with their respective four parental inbred lines. The objectives of this study were to identify heterosis related DEGs involved in plant height and to interpret underlying genetic reasons that control heterosis in intra-specific hybrids in *L. longiflorum*. We predicted that DEGs in hybrids compared to parents in relation to these traits would exhibit heterosis for plant height in *L. longiflorum*. In this study, we identified some commonly expressed up- and down-regulated candidates DEGs in hybrids in relation to their respective parental lines and analyzed their expression patterns. We then explored the functional features of the candidate genes to understand the molecular mechanism and genetic reason of heterosis for plant height in *L. longiflorum*.

## Results

**Characterization of Hybrids and their parental lines.** We used two different intra-species crosses (L4–7, L2–4 × L2–28 and L4–104, L2–22 × L2–20) of *Lilium longiflorum* (Easter lily) to investigate the phenotypic variations between hybrids and their respective inbred parents (Fig. 1).

Plant height, leaf length, flowering time, number of flowers, and flower diameter showed significant ( $p < 0.05$ ) phenotypic variation between hybrids and their inbred parents in both cross combinations (Table 1). The plant height in hybrid L4–7 (47.1 cm) was 21.0% and 23.67% higher than the high-parent or better-parent (37.2 cm) and mid-parent (35.95 cm), respectively, whereas that was 23.3% and 32.33% higher than the high-parent (39.5 cm) and mid-parent (35.05 cm), respectively in hybrid L4–104 (51.8 cm) (Table 1). The leaf length in hybrid L4–7 (14.5 cm) was 6.89% lower than the high-parent (15.5 cm) and 4.13% higher than the mid-parent (13.9 cm)



**Figure 1.** Comparisons of phenotypic variations in hybrids with their respective parents of *L. longiflorum*. The left panel (Set-A) shows the hybrid (F1, L4-7) and its parents of L2-4 (P1) and L2-28 (P2) while the right panel (Set-B) shows the hybrid (F'1, L4-104) and its parents of L2-22 (P'1) and L2-20 (P'2) at 4 months seedling stage.

whereas that in hybrid L4-104 (18.5 cm) was 29.7% and 31.8% higher than the high-parent (13.0 cm) and mid-parent (12.6 cm), respectively (Table 1). The days to flowering in hybrid L4-7 (184 days) was 2.1% and 3.53% lower than the high-parent (188 days) and mid-parent (190.5 days), respectively, whereas that in hybrid L4-104 (183 days) was 3.27% and 4.64% lower than the high-parent (189 days) and mid-parent (191.5 days), respectively (Table 1). The number of flowers in hybrid L4-7 (10.3) was 44.6% and 48.5% higher than the high-parent (5.7) and mid-parent (5.3), respectively, whereas that in hybrid L4-104 (7.5) was 49.3% and 50.0% higher than the high-parent (3.8) and mid-parent (3.75), respectively (Table 1). The flower diameter in hybrid L4-7 (13.3) was 8.2% and 6.7% lower than the high-parent (14.4) and mid-parent (14.2), respectively, whereas that in hybrid L4-104 (13.8) was 3.6% and 1.4% lower than the high-parent (14.3) and mid-parent (14.0), respectively (Table 1).

The quantification of mid-parent heterosis (MPH) and high-parent heterosis (HPH) in relation to tested traits were varied significantly ( $p < 0.05$ ) between hybrids and their respective parents (Table 1). For plant height, MPH (31.01%) was higher than the HPH (26.61%) in L4-7 hybrid, whereas in hybrid L4-104, MPH (47.79%) was higher than the high-parent heterosis (HPH) (31.34%) (Table 1). For leaf length, MPH (4.32%) was higher than the HPH (-6.45%) in L4-7 hybrid, whereas MPH (46.83%) was higher than the HPH (42.31%) in hybrid L4-104. For days to flowering, MPH (-3.41%) was higher than the HPH (-4.66%) in L4-7 hybrid whereas MPH (-4.44%) was higher than the HPH (-5.67%) in hybrid L4-104. For number of flowers, MPH (94.34%) was higher than the HPH (80.70%) in L4-7 hybrid, whereas MPH (100%) was higher than the HPH (97.37%) in hybrid L4-104. For flower diameter, MPH (-6.34%) was higher than the HPH (-8.27%) in L4-7 hybrid whereas MPH (-1.43%) was higher than the HPH (-3.49%) in hybrid L4-104 (Table 1).

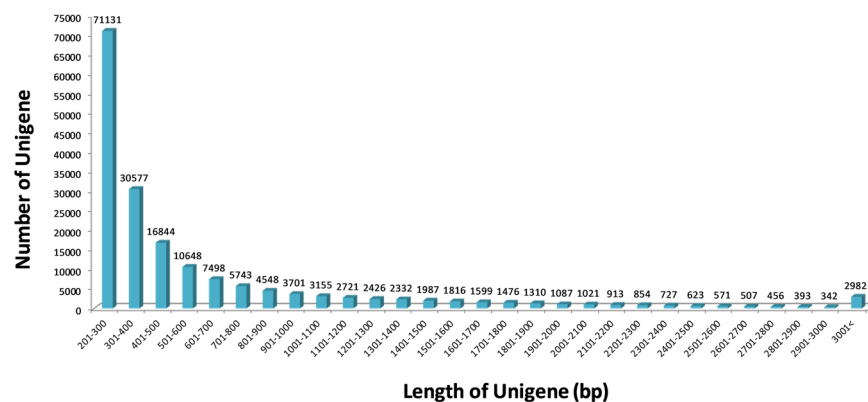
**Illumina paired-end sequencing.** We used RNA-Seq to discover differentially expressed heterotic genes and to investigate the function of those DEGs in hybrids in compared to parents. To identify heterotic transcripts of interest, six cDNA libraries were constructed from leaf RNA samples of six genotypes including two hybrids (L4-7 and L4-104) and four inbred parent lines (L2-4, L2-28, L2-22, and L2-20) of *L. longiflorum* and subjected to RNA-Seq analysis on the Illumina HiSeq. 2000 platform. A total of 743,964,980 short reads, each of approximately 100 bp, were generated from the six genotypes (Supplementary Table S2). After stringent quality assessment and data filtering such as removing adaptor sequences and discarding low quality reads, overall 711,948,744 high-quality 100-bp clean reads with a base quality score of 20% were selected from the six genotypes for further analysis. The GC contents of the six libraries were more than 51% (Supplementary Table S2). The parents and their hybrids showed strong correlation with respect to gene expression levels according to Pearson's correlation with  $-1 \leq r \leq 1$  (Supplementary Fig. S1). The transcripts were then clustered into unigenes. A total of 179,988 unigenes were assembled with a total length of 113,117,791 bp, and the lengths of the transcripts were ranged from 201 to 16,536 with an average of 628 bp (Fig. 2 and Supplementary Table S3).

About 83.72% of the transcripts were in the range of 201-1000 bp (~1 kb), 14.62% transcripts were 1001-3000 bp (1.1 to 3.0 kb) and 1.66% was longer than 3001 bp (>3 kb) (Supplementary Table S3). From the blast results of all 179,988 unigenes in TAIR genome database, finally, 53,209 annotated unigenes were detected with at least one sequence read (Data not shown).

Phenotypic traits	Cross combination A					Cross combination B				
	L2-4 (P1) (♀)	L4-7 (F1)	L2-28 (P2) (♂)	MPH (%)	HPH (%)	L2-22 (P'1) (♀)	L4-104 (F'1)	L2-20 (P'2) (♂)	MPH (%)	HPH (%)
Plant height (cm)	37.2 <sup>b</sup> ± 2.4	47.1 <sup>a</sup> ± 3.4	34.7 <sup>c</sup> ± 4.2	31.01 <sup>**</sup>	26.61 <sup>**</sup>	39.5 <sup>b</sup> ± 3.7	51.8 <sup>a</sup> ± 5.9	30.6 <sup>c</sup> ± 4.3	47.79 <sup>**</sup>	31.14 <sup>**</sup>
Leaf length (cm)	12.3 <sup>c</sup> ± 1.8	14.5 <sup>b</sup> ± 1.2	15.5 <sup>a</sup> ± 2.4	4.32 <sup>**</sup>	-6.45 <sup>**</sup>	13.0 <sup>b</sup> ± 1.7	18.5 <sup>a</sup> ± 2.1	12.2 <sup>c</sup> ± 1.8	46.83 <sup>**</sup>	42.31 <sup>**</sup>
Days to flowering	193 <sup>a</sup>	184 <sup>c</sup>	188 <sup>b</sup>	-3.41 <sup>**</sup>	-4.66 <sup>**</sup>	194 <sup>a</sup>	183 <sup>c</sup>	189 <sup>b</sup>	-4.44 <sup>**</sup>	-5.67 <sup>**</sup>
Number of flowers	4.9 <sup>c</sup> ± 1.1	10.3 <sup>a</sup> ± 2.1	5.7 <sup>b</sup> ± 1.3	94.34 <sup>**</sup>	80.70 <sup>**</sup>	3.7 <sup>bc</sup> ± 1.7	7.5 <sup>a</sup> ± 2.1	3.8 <sup>bc</sup> ± 1.2	100 <sup>**</sup>	97.37 <sup>**</sup>
Flower diameter (cm)	14.4 <sup>a</sup> ± 0.5	13.3 <sup>c</sup> ± 1.2	14.0 <sup>b</sup> ± 1.7	-6.34 <sup>**</sup>	-8.27 <sup>**</sup>	13.7 <sup>bc</sup> ± 1.1	13.8 <sup>bc</sup> ± 1.7	14.3 <sup>a</sup> ± 0.7	-1.43 <sup>ns</sup>	-3.49 <sup>*</sup>
Flower color	white	white	white	—	—	white	white	white	—	—
Stem color	green	green	green	—	—	green	green	green	—	—

**Table 1.** Mid-parent heterosis (MPH) and high-parent heterosis (HPH) of different phenotypic traits in six genotypes of *L. longiflorum*. MPH and HPH were calculated using the following formulas, MPH = (F1 - MP)/MP in % and HPH = (F1 - HP)/HP in %, where F1 denotes the average performance of the hybrid, MP denotes the average performance of the two parents, and HP denotes the average performance of the better-parent between two parents. Each data represent average of three plants. \*\*Significant difference with  $p < 0.01$ , \*Significant difference with  $p < 0.05$ , nsNon-significant difference with  $p < 0.05$ . Different letters within particular traits in each cross are statistically significant and same letters are statistically non-significant.

**The length distribution of unigenes**

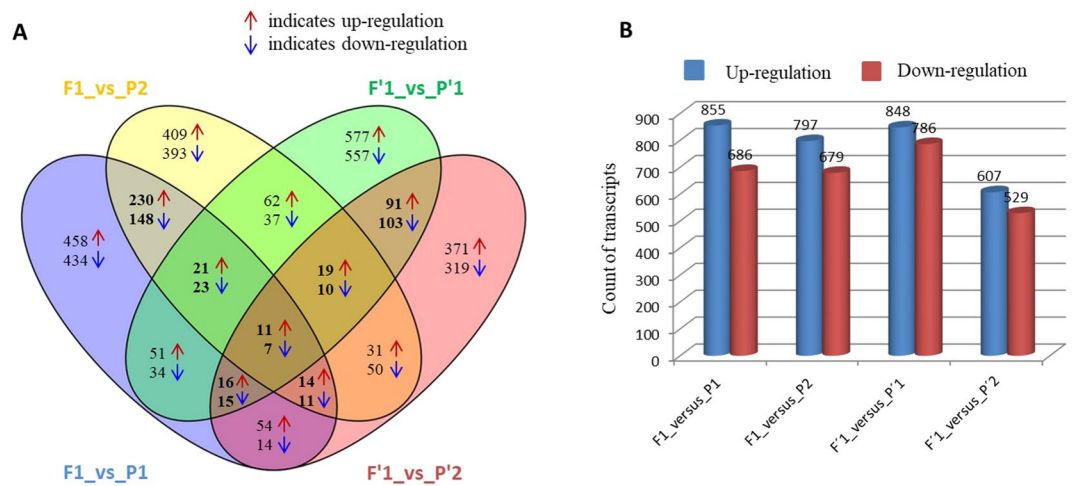


**Figure 2.** The length distribution of unigenes identified in the transcriptomes of lily plants.

**Identification and analysis of DEGs by RNA-Seq.** Putative heterotic DEGs in hybrids over their respective parents were identified using the following criteria, fold change (FC)  $\geq 2$  for up-regulated and  $\leq -2$  for down-regulated DEGs. Based on these criteria, a total of 4,327 up- and/or down-regulated transcripts as reliable DEGs were identified from six genotypes involving two crosses of L2-4  $\times$  L2-28 (F1, L4-7) and L2-22  $\times$  L2-20 (F'1, L4-104) in *L. longiflorum* (Data not shown). The numbers of up- and/or down-regulated DEGs in the hybrid with each parent combination varied considerably (Fig. 3A,B).

Some DEGs were over-expressed either in hybrids or parents. Among 4,327 DEGs, a total of 1,541 DEGs varied from hybrid L4-7 to parent L2-4 (855 up- and 686 down-regulated), 1,476 DEGs varied from hybrid L4-7 to parent L2-28 (797 up- and 679 down-regulated), 1,634 DEGs varied from hybrid L4-104 to parent L2-22 (848 up- and 786 down-regulated), and 1,136 DEGs varied from hybrid L4-104 to parent L2-20 (607 up- and 529 down-regulated) (Fig. 3B). Some DEGs had genotype specific up- and/or down-regulated expression and those were recognized as genotype specific unigenes (Fig. 3A). Venn diagram analysis revealed that overall 737 transcripts showed up and/or down-regulation (highlighted as black bold font) between hybrids and parents in both sets of cross combinations (Fig. 3A). Out of 737 transcripts, a total of 276 up- and 189 down-regulated DEGs (sum of bold font numerals) were commonly observed between L4-7\_vs\_L2-4 and L4-7\_vs\_L2-28 whereas 137 up- and 135 down-regulated DEGs (sum of bold font numerals) were commonly observed between L4-104\_vs\_L2-22 and L4-104\_vs\_L2-20 (Fig. 3A and Supplementary Table S4A). In hybrids from both cross combinations, 11 up- and 7 down-regulated DEGs were found commonly expressed. Again, 14 DEGs were up-regulated in L4-7 and down-regulated in L4-107 hybrids whereas 2 DEGs were up-regulated in L4-107 and down-regulated in L4-7 hybrids (Fig. 3A and Supplementary Table S4A). Thus, a total of 703 up-regulated and down-regulated DEGs were selected for subsequent analysis (Supplementary Table S4A-C).

To explore and categorize the expression alterations, we classify the commonly expressed 703 DEGs into 12 possible expression patterns for hybrids and parents (Table 2). We observed the notable mode of gene action differences especially in additive gene expression effect (classes 1 and 12) on commonly expressed 703 DEGs between two crosses of hybrids L4-104 and L4-7 (Table 2). In both hybrids, total additive gene effect was 19.35% (Table 2).



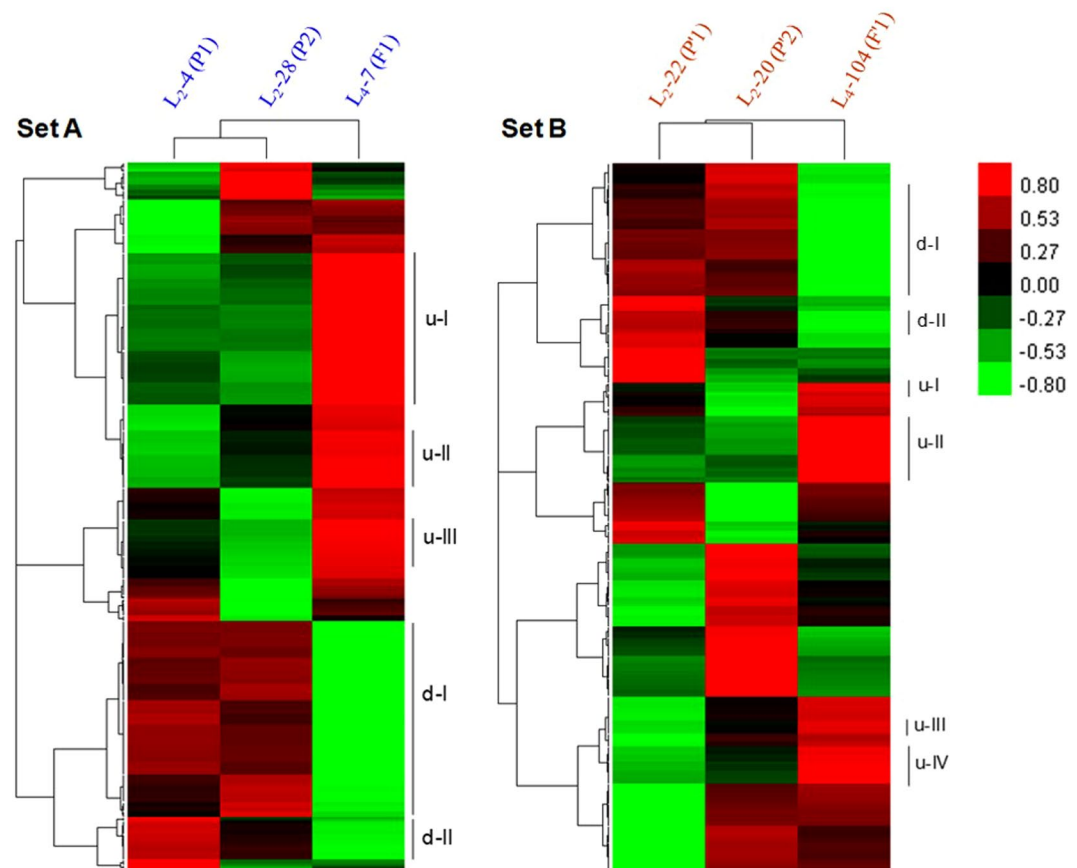
**Figure 3.** Identification of differentially expressed genes (DEGs) between hybrids and parents. Total numbers of DEG in L4-7\_vs\_L2-4, L4-7\_vs\_L2-28, L4-104\_vs\_L2-22, and L4-104\_vs\_L2-20 by venn diagram software version 2.1 (3A). Statistics of up- or down-regulated genes between hybrids and parents was shown by colored arrow heads (3A) and next to colored columns (3B). Bold numerals indicate commonly expressed up- and/or down-regulated DEGs between hybrid (s) and parents (3A).

Sl. No.	Categories	Expression patterns of DEGs											Total DEGs	
		Non-additivity												
		Additivity		Dominance				Overdominance						
		1	12	ELD <sub>♂</sub>		ELD <sub>♀</sub>		Transgressive up-regulation			Transgressive down-regulation			
Classes	1	12	2	11	4	9	5	6	8	3	7	10		
	Relative expression													
1	Hybrid L4-7 = n, (L2-4 × L2-28)	32	51	2	7	3	7	148	156	56	108	9	124	703
	Sum	83		9		10		360			241			
2	Hybrid L4-104 = n, (L2-22 × L2-20)	121	68	14	13	3	18	132	62	32	128	10	102	
	Sum	189		27		21		226			240			
	Average in both (%)	136		18		15.50		293			240.50			
	Per cent (%) in both	19.35		2.56		2.20		41.68			34.21		100	
				4.76		75.89								
				80.65										

**Table 2.** Classification of expression patterns of commonly expressed 703 DEGs in hybrids and their respective parents in *L. formolongi*. DEGs were classified according to the expression levels exhibited by parental and hybrid lines. Additive expression of genes: classes 1 and 12 (blue); dominance expression genes: classes 2, 11, 4 and 9 (green); overdominance expression genes: classes 5, 6, 8, 3, 7 and 10 (red). Classes 5, 6 and 8 represents transgressive upregulation and classes 3, 7 and 10 represent transgressive downregulation<sup>31</sup>. Diagrams of each class represents the relative expression levels observed in the maternal parent (left point), F1 (middle point), and paternal parent (right point). DEGs, differentially expressed genes, ♀, maternal parent, ♂, paternal parent, F1, hybrid, ELD, expression-level dominance, Sl, serial, n, the total number of differentially expressed genes in each class, additivity:  $F1 \approx 1/2 (\varphi + \sigma)$ , non-additivity:  $F1 > 1/2 (\varphi + \sigma)$  or  $F1 < 1/2 (\varphi + \sigma)$ ,  $ELD_{\varphi}$ :  $F1 \approx \varphi > \sigma$  or  $F1 \approx \varphi < \sigma$ ,  $ELD_{\sigma}$ :  $F1 \approx \sigma > \varphi$  or  $F1 \approx \sigma < \varphi$ , transgressive up-regulation:  $F1 > \varphi$  and  $F1 > \sigma$ , transgressive down-regulation:  $F1 < \varphi$  and  $F1 < \sigma$ .

Non-additivity (80.65%) was mainly classified into dominance and overdominance gene expression effect (Table 2). The dominance effect in both crosses (4.76%) is again subdivided as expression level dominance in paternal effect ( $ELD_{\sigma}$ ) (classes 2 and 11) comprised of 2.56% variation and in maternal effect ( $ELD_{\varphi}$ ) (classes 4 and 9) comprised of 2.20% variation (Table 2). Conversely, noteworthy variation in overdominance effect (75.89%) was observed in L4-7 and L4-104 (Table 2). Hybrids L4-7 and L4-104 showed higher transgressive up-regulation (classes 5, 6, 8) comprised of 41.7% while they showed transgressive down-regulation (classes 3, 7, 10) comprised of 34.2% (Table 2).





**Figure 4.** Heat map analysis of functionally characterized DEGs (384) for the four parental inbred lines viz., L2-4 (P1), L2-28 (P2), L2-22 (P'1), and L2-20 (P'2) and two hybrids viz., L4-7 (F1), and L4-104 (F'1). L2-4 (P1), L2-28 (P2) and F1 (L2-4 (P1)  $\times$  L2-28 (P2)) (A); L2-22 (P'1), L2-20 (P'2) and F'1 (L2-22 (P'1)  $\times$  L2-20 (P'2)) (B). The red color denotes the highly expressed up-regulated DEGs, and the green color denotes down-regulated DEGs with lower expression levels. The gradation from red to green represents the transition from large to small values of a FPKM normalized  $\log_2$  transformed counts. The up- (uI, uII, uIII, uIV) and down- (dI, dII) regulated DEGs are enlisted in Table S5B.

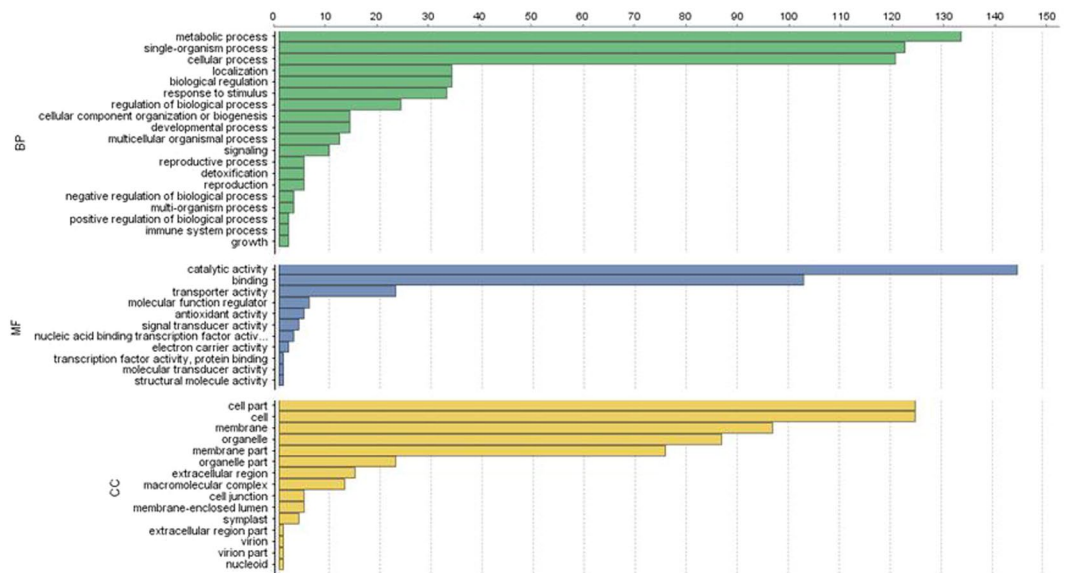
**Transcriptome profiles of hybrids and parents.** Blast2GO search showed that out of 703 DEGs 384 were functionally characterized, and rests 319 were without any blast hits (Supplementary Fig. S3). Among the 384 DEGs, those were functionally characterized, 260 were functionally annotated, 66 had hits with GO mapped, and 58 had only blast hits (Supplementary Fig. S3). Correlations of 384 functionally characterized up- and down-regulated DEGs between hybrids and their respective parents in two cross combinations were investigated using heat map analysis (Fig. 4 and Supplementary Table S5A).

A total of 139 and 127 DEGs were up- and down-regulated, respectively, in L4-7 hybrid compared to the parental lines L2-4 and L2-28 (Fig. 4A and Supplementary Table S5B). In contrast, 68 and 73 DEGs were up- and down-regulated, respectively, in L4-104 hybrid compared to the parental lines L2-22 and L2-20 (Fig. 4B and Supplementary Table S5B). Based on contrasting expression pattern between hybrids and their respective parents, 12 up- and 16 down-regulated DEGs were identified in both crosses and considered them as universal DEGs in *L. longiflorum* (Table 3 and Supplementary Table S5B). Importantly, the up- and down-regulated universal DEGs showed similarity with the different types of proteins and enzymes (Table 3).

**Functional classification by gene ontology.** Gene Ontology (GO) categories revealed that functionally annotated 260 DEGs (36.98%) with coding regions bear at least one functional group (Supplementary Table S6). The 260 proteins were assigned into 45 GO functional groups within three main categories, molecular function (sub-categories, 11), biological process (sub-categories, 19), and cellular components (sub-categories, 15) (Fig. 5). In the cellular component category, cell and cell parts followed by membrane, organelle and membrane part were the predominant subcategories. In the molecular function category, catalytic activity and binding followed by transporter activity were the predominant subcategories. In the biological process category, metabolic process, single organism process, and the cellular process followed by localization, biological regulation, and response to stimulus were the predominant subcategories (Fig. 5).

Genetic hypotheses	DEGs		L4-7_vs_	L4-7_vs_	L4-104_vs_	L4-104_vs_	Functional characterization	
			L2-4	L2-28	L2-22	L2-20		
		Expression level by fold change (FC) value						
Overdominance (1 + 2)	Up-regulated	c47423g1i1	<b>2.75</b>	<b>2.69</b>	<b>2.21</b>	<b>2.14</b>	ABC transporter A family member 7-like	
Dominance (1) Overdominance (2)		c54944g1i1	<b>2.15</b>	1.60	<b>2.69</b>	<b>2.22</b>	ABC transporter C family member 10-like	
Additive (1) Overdominance (2)		c60389g1i1	1.23	1.35	<b>2.75</b>	<b>3.69</b>	B3 domain-containing Os04g0386900-like	
Additive (1) Overdominance (2)		c60389g1i2	1.18	1.14	<b>2.35</b>	<b>3.02</b>	B3 domain-containing Os04g0386900-like	
Overdominance (1 + 2)		c49702g1i1	<b>2.96</b>	<b>2.77</b>	<b>4.85</b>	<b>3.18</b>	monothiol glutaredoxin-S11	
Dominance (1) Overdominance (2)		c52443g1i1	<b>2.27</b>	1.75	<b>7.99</b>	<b>7.26</b>	trans-resveratrol di-O-methyltransferase-like	
Overdominance (1) Additive (2)		c57602g1i1	<b>2.76</b>	<b>3.22</b>	1.16	1.24	auxin efflux carrier component 8	
Overdominance (1 + 2)		c59275g1i2	<b>2.20</b>	<b>2.16</b>	<b>5.80</b>	<b>4.39</b>	probable mannitol dehydrogenase	
Overdominance (1) Additive (2)		c62533g1i1	<b>2.46</b>	<b>2.13</b>	1.25	1.27	tRNA-dihydrouridine(16/17) synthase [NAD(P)(+)]-like	
Additive (1) Overdominance (2)		c63435g1i2	1.29	1.19	<b>3.97</b>	<b>3.22</b>	very-long-chain enoyl-reductase-like	
Additive (1) Overdominance (2)		c64671g4i2	1.63	1.39	<b>2.25</b>	<b>2.38</b>	disease resistance RPP13/1	
Dominance (1) Overdominance (2)		c60887g1i2	1.87	<b>2.21</b>	<b>2.47</b>	<b>2.79</b>	conserved hypothetical protein	
Dominance (1) Overdominance (2)		Down-regulated	c52384g1i2	-1.97	<b>-2.96</b>	<b>-2.38</b>	<b>-2.18</b>	vicilin-like antimicrobial peptides 2-2
Overdominance (1 + 2)			c58513g1i1	<b>-2.63</b>	<b>-4.58</b>	<b>-2.84</b>	<b>-3.66</b>	vicilin-like antimicrobial peptides 2-2
Overdominance (1) Dominance (2)	c47447g1i1		<b>-4.33</b>	<b>-3.18</b>	-1.92	<b>-2.04</b>	auxin-responsive SAUR68-like	
Overdominance (1 + 2)	c63663g3i2		<b>-2.84</b>	<b>-2.49</b>	<b>-2.06</b>	<b>-2.11</b>	auxin-responsive SAUR68-like	
Overdominance (1 + 2)	c47368g1i1		<b>-3.74</b>	<b>-2.04</b>	<b>-2.32</b>	<b>-2.88</b>	probable xyloglucan-glycosyltransferase 12	
Overdominance (1 + 2)	c59731g1i1		<b>-2.76</b>	<b>-4.20</b>	<b>-3.94</b>	<b>-2.22</b>	probable WRKY transcription factor 70	
Additive (1) Overdominance (2)	c47856g3i1		-1.78	-1.66	<b>-2.37</b>	<b>-2.39</b>	regulation of nuclear pre-mRNA domain-containing 1A-like	
Overdominance (1 + 2)	c37944g1i1		<b>-2.41</b>	<b>-5.84</b>	<b>-3.36</b>	<b>-2.88</b>	heavy metal-associated domain containing	
Additive (1) Overdominance (2)	c55548g2i1		-1.38	-1.45	<b>-2.47</b>	<b>-2.08</b>	oxidative stress isoform 1	
Additive (1) Overdominance (2)	c56718g1i1		-1.74	-1.59	<b>-2.60</b>	<b>-2.71</b>	carboxypeptidase 2	
Additive (1) Overdominance (2)	c57368g1i1		-1.34	-1.19	<b>-2.26</b>	<b>-2.03</b>	pentatricopeptide repeat-containing At1g62350-like	
Overdominance (1) Additive (2)	c60709g3i1		<b>-3.59</b>	<b>-4.70</b>	-1.19	-1.13	chloroplastic-like	
Overdominance (1) Additive (2)	c80430g1i1		<b>-2.91</b>	<b>-4.30</b>	-1.27	-1.27	lysine histidine transporter-like 8	
Overdominance (1) Dominance (2)	c96948g1i1		<b>-2.95</b>	<b>-6.14</b>	<b>-2.62</b>	-1.99	naringenin,2-oxoglutarate 3-dioxygenase-like	
Overdominance (1 + 2)	c13258g1i1		<b>-2.33</b>	<b>-2.74</b>	<b>-3.25</b>	<b>-2.23</b>	hypothetical protein PHAVU_005G042200g	
Overdominance (1) Additive (2)	c22243g1i1		<b>-2.83</b>	<b>-2.58</b>	-1.59	-1.18	hypothetical protein MIMGU_mgv1a014176mg	

**Table 3.** Expression level of functionally characterized universal DEGs in *L. longiflorum*. Hybrid 1, L4-7 (F1); Hybrid 2, L4-104 (F1). Bold numerals indicate significant expression level of up- (+) and down- (-) regulated differentially expressed genes (DEG).



**Figure 5.** Comparative Gene Ontology (GO) classifications of commonly expressed functionally annotated DEGs from two cross combinations in the lily transcriptome. The genes corresponded to three main categories, cellular component (CC), molecular function (MF), and biological process (BP).

**Kyoto encyclopedia of genes and genomes (KEGG) pathway mapping.** KEGG pathway-based analysis was performed using Blast2GO software to interrogate the KEGG database for further deepen our understanding of the biological functions and interactions of genes. All 260 functionally annotated transcripts sequences were selected and assigned to the reference canonical pathways in KEGG. A total of 45 DEGs were distributed to 59 different metabolic pathways in the KEGG database (Table S7). We observed that each of 15 KEGG pathways comprised more than one DEG while each of the remaining 44 pathways contained only one DEG (Supplementary Table S7). The most enriched KEGG Orthology terms (ko-term) with the highest levels of gene representation were biosynthesis of antibiotics (ko01055, 8 transcripts) followed by phenylpropanoid biosynthesis (ko00940, 6 transcripts), carbon fixation in photosynthetic organisms (ko00710, 5 transcripts), methane metabolism (ko00680, 5 transcripts), pyrimidine metabolism (ko00240, 5 transcripts), purine metabolism (ko00230, 5 transcripts), fructose and mannose metabolism (ko00051, 4 transcripts), pentose phosphate pathway (ko00030, 4 transcripts glycolysis/gluconeogenesis (ko00010, 4 transcripts), thiamine metabolism (ko00730, 3 transcripts), sulfur metabolism (ko00920, 3 transcripts), starch and sucrose metabolism (ko00500, 2 transcripts), drug metabolism cytochrome P450 (ko00982, 2 transcripts), pentose and glucuronate interconversions (ko00040, 2 transcripts), and other glycan degradation (ko00511, 2 transcripts) (Supplementary Table S7).

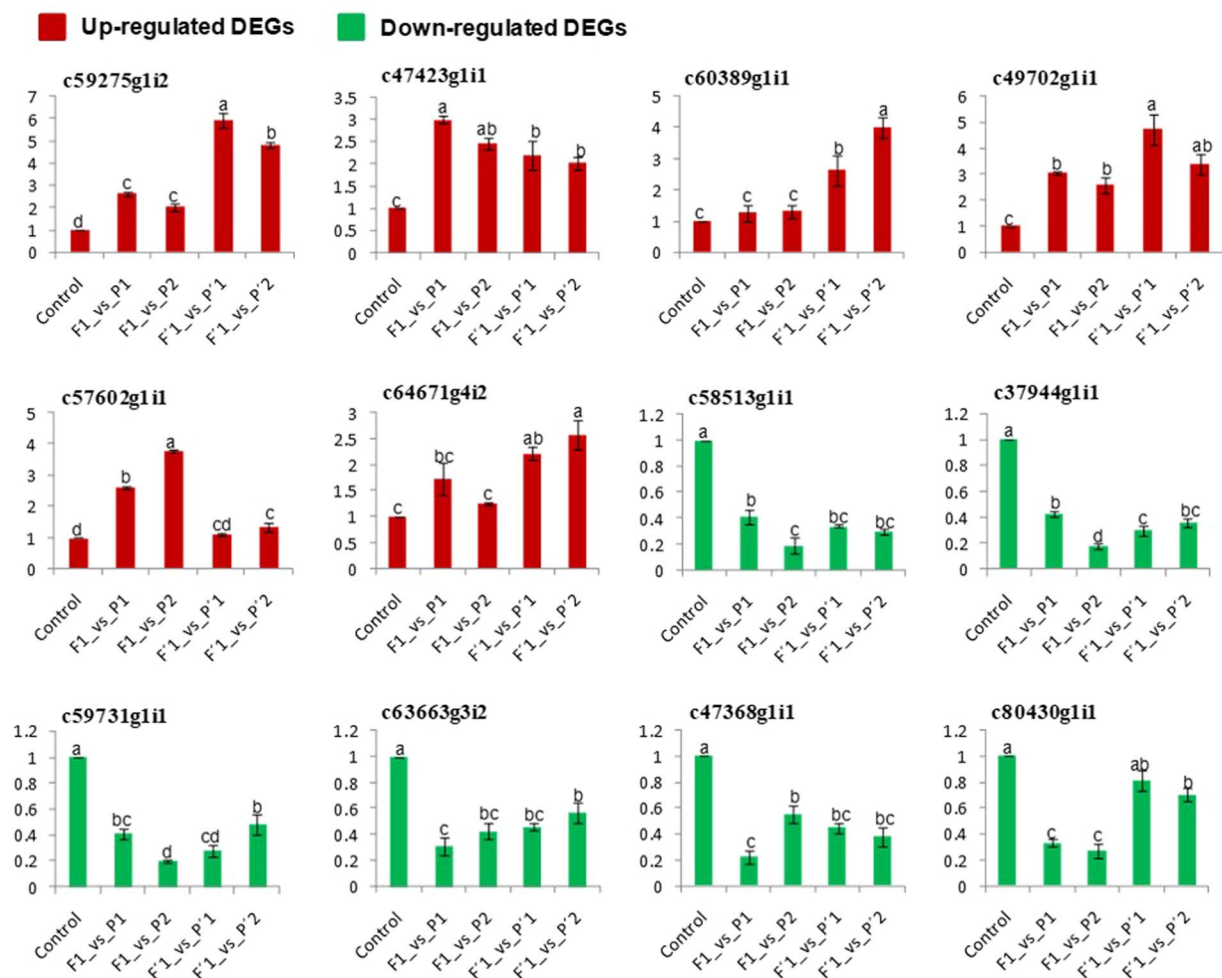
**Validation of DEGs by quantitative real-time PCR.** A subset of the 12 DEGs (6 up- and 6 down-regulation types, yellow highlighted in Table 3) were selected for the verification of the expression patterns using real-time quantitative PCR (qPCR) validation (Supplementary Table S1 and Fig. 6). We compared the transcript profiles obtained from qPCR with those generated from RNA-Seq analysis from the six genotypes of *L. longiflorum*. The data in Fig. 6 confirmed that almost all 12 DEGs displayed the expression patterns consistent well with expression levels obtained from RNA-Seq analysis.

## Discussion

Heterosis is a widely exploited phenomenon in plant breeding in cross-pollinated crops. In general, the higher the heterosis and heritability, the simpler the selection process and greater the response to selection. In this study, the recorded higher plant height, leaf length, and number of flowers but lower days to flowering and flower diameter in both hybrids in contrast to their parents could be attributed by the genetic make-up of the genotypes. Transcriptome analysis identified some universally expressed DEGs in both hybrids. The functional proteins associated with those universally expressed DEGs might be largely regulating heterosis in both hybrid combinations.

This study investigated seven phenotypic traits including plant height, leaf length, flowering time, number of flowers, and flower diameter that showed hybrid vigor, either mid-parent heterosis or high-parent heterosis. However, only leaf sample was taken from the elongating leaves before any flower appearance for RNA-seq, and we did not look at the transcriptome from other organs like flower. As elongation of leaf was associated with plant height we therefore believe that the DEGs expressed in leaves could be associated with increasing plant height in hybrids<sup>51</sup>. Nevertheless, collection of leaf samples at different growth stages, e.g., two or three growth phases, certainly generate additional sets of RNAseq data. Thus, it makes possible to compare the variable DEGs from the different growth stages. The results obtained from such a comprehensive experiment would provide a confirmatory picture to discerning the genes responsible for heterosis. However, in this study, samples were collected from the 2<sup>nd</sup> and 3<sup>rd</sup> youngest leaves those were still elongating. There were more than 20 live leaves per plant at the sampling time and the samples were taken from those plants that did not produce any flower yet





**Figure 6.** Validation of differential expression of 12 universal genes using real-time quantitative PCR (qPCR). The relative transcript levels were normalized with *Lf-actin* as the standard. Red and green colored bars denote up- and down-regulated DEGs, respectively. The error bars indicate the standard error of the mean of three independent replicates. Different lower case letter (a, b, c, d) indicates the significant difference in F1\_vs\_P1 (L4-7\_vs\_L2-4), F1\_vs\_P2 (L4-7\_vs\_L2-28), F'1\_vs\_P'1 (L4-104\_vs\_L2-22), and F'1\_vs\_P'2 (L4-104\_vs\_L2-20) compared to control at  $p < 0.05$ .

(Fig. 1). Leaf elongation and increasing plant height in lily is a dynamic process. Leaf that initiate at the axis of the whorl dynamically undergoes cell division, elongation, expansion, maturation and senescence. Every leaf at any developmental stage of a lily plant pass through those processes. We therefore believe that an elongating leaf at any particular vegetative growth phase of lily plant may produce similar transcriptomic profiling with some variations related to plant development. In our study, both sampled-leaves and sampling plants were still elongating. We therefore believe that the differentially expressed genes between parents and hybrids should be responsible for controlling the plant height and leaf elongation. Thus the transcriptomic profiles obtained in this study are expected to represent plant height. However, the expression of genes related to flowering time, number of flowers, and flower diameter is a subject of further investigation.

Transcriptome analysis as a powerful tool facilitates the identification of DEGs with their expression level and regulatory mechanisms. Previous investigations identified various heterotic DEGs related to a variety of traits between hybrids and parents in plants and animals<sup>13,31,52,53</sup>. Heterotic genes are robust due to multiple allelic combinations between genomes of two parents in both cross and reciprocal cross combinations suggesting that major genes genetically govern this phenomenon<sup>31,54</sup>. Divergent patterns of DEGs between hybrid and its inbred parents play a significant role in hybrid vigor or heterosis<sup>13,18,20,31,55-59</sup>. RNA-Seq generated approximately 711 million high-quality 100-bp paired-end reads from the leaves of six genotypes, which were almost double compared to rice root transcriptome<sup>13</sup>. In addition, 53,209 annotated transcripts were identified that was 1.37 times higher compared to rice root transcriptome (38,872 annotated transcripts)<sup>13</sup>. A total of 4,327 DEGs were selected based on fold change value in hybrid compared to inbred parents indicating that a large number of genes could be functionally associated with hybrid vigor in lily. Out of 4,327, 703 commonly up- or down-regulated DEGs were identified between hybrids and parents in both crosses (Fig. 3). Among them, 384 functionally annotated DEGs were identified using Blast2GO analysis (Supplementary Table S5A).

One of the main interest of this study is to compare the levels of additive and non-additive gene expression in two hybrids with varying levels of heterosis for leaf length which is likely associated with plant height<sup>51</sup>. Interestingly, less than 20% DEGs showed additivity on an average indicating that non-additive gene interactions have major control in heterosis (Table 2)<sup>21,36</sup>. Moreover, between dominance and overdominance non-additive interactions, the overdominance effect was vital as it controls more than 75% gene effect (Table 2). Both transgressive upregulation of DEGs and down-regulation of DEGs have major effect on heterosis (Table 2).

Importantly, heat map analysis revealed that 12 up- and 16 down-regulated universal DEGs in both sets of hybrids were remarkably down- and up-regulated, respectively, in parental lines indicating those universal genes might have a more certain functional role in plant height in comparison with remaining DEGs pool (Fig. 4). Moreover, among the 28 universally up- and down-regulated genes, nine genes exhibited overdominance effect in both hybrid combinations and remaining 19 genes showed overdominance effect either in one of two hybrid combinations (Table 3). This predicted genetic analysis indicated that majority of the 28 universal DEGs have overdominance effect to bring about heterosis for plant height in Easter lily, in this study. In *Nicotiana tabacum*, transcriptome analysis revealed that the heterosis for nicotine biosynthesis is critically controlled by overdominance effect of up-regulated genes related to nicotine anabolism and transport<sup>29</sup>. Functional annotation of these up- and down-regulated DEGs revealed that they were very close to the various classes of proteins and enzymes (Table 3). These common protein classes include ABC transporter family member-like protein and B3 domain-containing proteins among others.

Among 12 up-regulated transcripts, two transcripts showed similarity in functional annotation with the ABC transporter family member-like protein (Table 3) that is actively involved in DNA repair, RNA translocation, and active transport of a wide variety of substrates in cells as well as leading to coping with biotic and abiotic stresses in adverse conditions<sup>60</sup>. Both of these two ABC transporter family members exhibited overdominance in F1 (L4-104) hybrid but one gene showed dominance and another gene showed overdominance in another hybrid, F1 (L4-7), indicating that non-additive gene interaction in involved in this case (Table 3). Another two transcripts showed similarity with the abscisic acids and auxins B3 domain-containing proteins that take part in plant growth and seed maturation process<sup>61,62</sup>. A single transcript showed similarity with the monothiolglutaredoxin-S11 (Table 3) which has anti-oxidative protective roles in the cellular response against reactive oxygen species in plants<sup>63</sup>. Another single transcript showed similarity with the synthase-like and trans-resveratrol di-O-methyltransferase-like enzyme that protect plants against pathogens by catalyzing the biosynthesis of plant phytoalexins<sup>64,65</sup>. Both of these two B3 domain-containing proteins exhibited additive effect in F1 (L4-7) hybrid but overdominance effect in F1 (L4-104) hybrid (Table 3). Other individual transcripts showed similarity with the different families such as auxin efflux carrier family component 8, probable mannitol dehydrogenase, tRNA-dihydrouridine synthase, very-long-chain enoyl-ACP (acyl carrier protein) reductase enzymes (EC 1.3.1.9), disease resistance *RPP13/1* gene (Table 3). These genes play an important role in auxin-related plant growth and development<sup>66-68</sup>, pathogen-induced stress resistance<sup>69-72</sup> and fatty acid biosynthesis<sup>73</sup>. Thus, the twelve commonly expressed DEGs and their associated proteins likely function to tackle plant height under various biotic and abiotic stress environments and excessive transcript abundance of those genes in hybrids eventually might help in maintaining hybrid vigor largely through overdominance gene effect.

Among 16 down-regulated transcripts, two transcripts showed similarity with the auxin-responsive SAUR68-like (SMALL AUXIN UP RNAs) genes (Table 3) that enhance auxin-induced growth in hypocotyls, inflorescence, stems, petals and stamen filaments<sup>74</sup>. One transcript showed similarity with the heavy metal associated domains containing protein that involved in heavy metal transport and heavy metal homeostasis in plants<sup>75,76</sup>. Another single transcript showed similarity with the probable xyloglucan-glycosyltransferase 12 enzyme (Table 3) that functions to swell the cell wall for plant growth process<sup>77</sup>. Thus any DEGs associated with biosynthesis of xyloglucan-glycosyltransferase 12 enzyme might be associated with cell division of plants that is closely related to increasing both leaf length and plant height and the gene effect could be heterotic in hybrids largely due to overdominance (Table 3). Another transcript showed similarity with the regulation of nuclear pre-mRNA domain-containing 1A-like (RPRD1A-like) protein that is involved in a cell-cycle and dimerization process<sup>78</sup>. A single transcript showed similarity with the vicilin-like antimicrobial peptides 2-2 which functions in defense related responses against a wide range of pathogens<sup>79</sup>. Another transcript showed similarity with the oxidative stress isoform 1 (Table 3) which is involved in generating ROS particularly H<sub>2</sub>O<sub>2</sub> for plant defense mechanism<sup>80,81</sup>. Other individual transcripts were from different families such as pentatricopeptide repeat (PPR), WRKY transcription factor 70, chloroplastic-like gene, lysine histidine transporter-like 8, naringenin, 2-oxoglutarate 3-dioxygenase-like enzyme etc (Table 3). These genes have vital role in post-transcriptional processes<sup>82,83</sup>, biotic and abiotic stress tolerance<sup>84</sup>, growth and developmental process<sup>85</sup> and biosynthesis of proteins and secondary metabolites<sup>86,87</sup>. In 16 universally down-regulated genes, at least one hybrid showed overdominance although four of them exhibited additive effect in F1 (L4-7) and three of them exhibited additive gene in F1 (L4-104) (Table 3). Thus, similar to up-regulated DEGs, the sixteen commonly expressed DEGs and their associated proteins might also be involved to tackle plant height in hybrids largely through overdominance effect of genes under various biotic and abiotic stress environments, however, excessive transcript shortage of those genes in hybrids eventually might reduce the hybrid vigor.

Finally, the expression levels of 12 universal up- and down-regulated transcript generated from RNA-Seq analysis were verified by qPCR and the expression patterns of all the DEGs tested were consistent with the RNA-Seq results indicating reliability of the RNA-Seq data from the six genotypes of *L. longiflorum* (Fig. 6). Taken together, the genes identified herein might be useful for creating genetic diversity with increased hybrid vigor in *L. longiflorum*. However, further experiments using genes with specific trait of interest are needed to confirm the heterotic association between the gene expression patterns and the target agronomic traits in lily.

In this study, the majority of the commonly expressed DEGs revealed non-additive effect especially overdominance (transgressive up- and down-regulation) mode of gene expression patterns in both cross combinations (Table 2). Thus, overdominance mode of gene expression patterns might contribute more than the additive as well as dominance effects to intraspecific heterosis between hybrids and parents in *L. longiflorum*<sup>52</sup>.

Functional analysis placed the 260 DEGs into 45 subcategories (Fig. 5). GO terms representing biological processes, accounted for almost half of the 45 categories, indicating a diverse range of biological processes especially metabolic process, single-organism process, and cellular process might play major role in increasing plant height of lily hybrids (Fig. 5). Previous reports showed that biological processes including metabolic process and cellular process are related to root development and yield of rice as extensive metabolic activities were taking place in roots of hybrid plants in both vegetative and reproductive stages<sup>13</sup>. KEGG analysis revealed that 45 functional DEGs contributed to 59 different metabolic pathways possibly regulated hybrid vigor in lily (Fig. 5 and Supplementary Table S7). Reports showed that differentially expressed heterotic genes in rice control hybrid vigor by controlling metabolic pathways especially carbohydrate metabolism in root tissues for root elongation<sup>13</sup>. However, some DEGs were associated with more than one metabolic pathway indicating their overlapping functions in different metabolic activities and might be candidate heterotic genes for the growth, development, and stress-related mechanism in *L. longiflorum* (Supplementary Table S7).

The results presented in this study and the discussion has been made mostly focused a common set of DEGs in the two hybrids which could be of interest in regulating heterosis of leaf elongation and plant height. The confirmation of the specificity of genes responsible for heterosis of leaf elongation and plant height is subject of further investigation. As majority of DEGs showed non-additivity and specifically overdominance effect, we therefore comment on that overdominance gene interaction might be responsible of heterosis related to leaf elongation and plant height in Easter lily.

## Conclusions

In summary, we observed higher MPH for positive heterotic traits such as plant height, leaf length, and number of flowers than HPH whereas HPH was higher for the negative heterotic trait flowering time in *L. longiflorum*. Blast2GO analysis of 703 DEGs revealed that 384 DEGs were functionally characterized. The expressed DEGs might be associated with overdominance effects in hybrids to show heterosis in lily. Heat map analysis of 384 DEGs revealed that 12 and 16 DEGs were up- and down-regulated, respectively, in both sets of hybrids and showed similarity with different types of proteins and enzymes. The major protein families possibly involved in heterosis are ABC transporter A family member-like, B3 domain-containing, disease resistance RPP13/1, auxin-responsive SAUR68-like, and vicilin-like antimicrobial peptides 2-2 proteins etc. In addition, GO analysis indicated that biological process might contribute more in heterotic related genes expression compared to other processes. Furthermore, KEGG pathway-based analysis revealed that some DEGs were involved in more than one metabolic pathways indicating that some heterotic related genes might be functionally redundant in *L. longiflorum*. Validation of up- and down regulated universal DEGs showed that the expression trends were consistent with the RNA-Seq data suggested the reliability of the transcriptome data in *L. longiflorum*. This study identified some important protein families associated with universally up- or down-regulated DEGs. The results of this study advance the molecular underpinning of heterosis for plant height in lily plants.

## Materials and Methods

**Plant materials and growing conditions.** Seeds of two hybrids (L4-7, L2-4 × L2-28 and L4-104, L2-22 × L2-20) with their four inbred parents (L2-4, L2-28, L2-22 and L2-20) of *Lilium longiflorum* (Easter lily) were obtained from the Department of Floriculture, Korea National Agricultural College, Hwaseong, Korea. Seeds of hybrids and inbred lines were sown in the plastic pots (BN-150, 13.5 cm × 10.5–15 cm) containing sterilized soil mixture for germination. Plants were grown under greenhouse environmental condition with 18 ± 2 °C temperature and 80–85% relative humidity. Leaves (2<sup>nd</sup> and 3<sup>rd</sup>) of four months-aged plants, that did not produce any flower yet, were sampled with three replicates. Collected samples were immediately transferred to liquid nitrogen and kept at –80 °C for subsequent analysis.

**RNA isolation, cDNA library construction, and RNA-sequencing.** RNeasy mini kit (Qiagen, Hilden, Germany) was used for total RNA isolation from leaf samples (2<sup>nd</sup> and 3<sup>rd</sup>) of three replicates from each of six genotypes following the manufacturer's guidelines. The quality of isolated RNA was checked using NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). DNA was removed from the samples using RNase-free DNase (Promega, Madison, WI, USA), also according to the manufacturer's instructions. The integrity of RNA was checked by agarose gel electrophoresis<sup>88</sup>. cDNA synthesis was performed using equal quantities of high-quality RNA from each sample following IlluminaTruSeq™ RNA sample preparation kit. Sequencing libraries were generated according to manufacturer's guidelines (Illumina, San Diego, CA) for cluster analysis. Two consecutive purification steps were performed using poly-T oligo-attached magnetic beads for the isolation of poly-A-containing mRNA from the total RNA and then they were fragmented using an RNA fragmentation kit. Synthesis of first strand cDNA was done using reverse transcriptase and random primers whereas second strand cDNA was synthesized by DNA polymerase I and RNase H. After second strand cDNA synthesis and adaptor ligation, cDNA fragments of 200-bp were separated by gel electrophoresis. The cDNA fragments were subsequently amplified with PCR primer cocktail with minimum cycles to avoid skewing the representation of the library. The quantity and quality of amplified cDNA library were controlled by MacroGen (Seoul, South Korea). Following loading of amplified products into the Illumina HiSeq. 2000 instrument, they were exposed to automated cycles of paired-end-sequencing (2 × 100 bp). The processing of fluorescent images into sequences, base-calling and quality value calculations were carried out using the Illumina data processing pipeline version 1.8 (Solexa's Sequencing-by-Synthesis).

**De novo assembly and assessment of differential gene expression.** High-quality reads were filtered by eliminating low-quality reads with Q < 20 from the raw reads. To perform quality trimming and removal of adapter, fast and multi-threaded command line tool 'trimmomatic' was used (<http://www.usadellab>).

[org/cms/?page=trimmomatic](#))<sup>89</sup>. Further, high-quality reads (transcriptome) were assembled by Illumina/Solexa with *de novo* using trinity program (<http://trinityrnaseq.sourceforge.net>)<sup>90</sup> and were quantified of transcript abundances using RSEM (RNA-Seq by Expectation-Maximization) (v1.2.15) (<http://deweylab.biostat.wisc.edu/rsem/>)<sup>91</sup>. As lily has no reference genome, therefore, RSEM was used for quantification. Transcript expression level was quantified in terms of FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values ranging from  $> 0$  to over  $10^4$ , and estimated fold change (FC) (Macrogen, Seoul, Korea). Expression of the transcript that satisfied with  $FC \geq 2$  and  $FC \leq -2$  was considered as significantly up- and down-regulated DEGs, respectively at least one of total comparisons (Macrogen, Seoul, Korea). The transcriptome sequences of *L. longiflorum* (Easter lily) was deposited in our internal lily database<sup>92</sup>.

**Identification of DEG and cluster analysis.** Commonly expressed up- and down-regulated DEGs between hybrids and parents was selected using the Venny diagram (v2.1.0)<sup>93</sup>. Differential up- and down-regulation of annotated transcripts between hybrids and their parents were analyzed using cluster analysis. The cluster of FPKM-normalized expression values for each transcript was made with the cluster software version 3.0 followed by Java TreeView to visualize the results<sup>94</sup>.

**Functional annotation.** Blast (blastx) search was conducted to identify functionally annotated DEGs protein against the GenBank non-redundant (NR) database (<http://www.ncbi.nlm.nih.gov/>), SWISSPROT database (<http://www.ebi.ac.uk/swissprot/>), Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg/>) and Clusters of Orthologous Groups (COG) of protein databases with an E-value of  $1e^{-5}$ . Unique transcripts were functionally classified according to gene ontology (GO) using TAIRGO slim of Blast2GO software version 3.3 (<https://www.blast2go.com/>) and plotted based on their distribution<sup>95</sup>. Association between unique transcripts and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was performed using blastp against the KEGG database<sup>96</sup>.

**Phenotypic traits and DEGs expression pattern measurements.** Plant height (cm), leaf length (cm), the number of flowers, flowering duration (days), and flower diameter (cm) of the six genotypes were recorded. Calculation of mid-parent heterosis (MPH) and high-parent heterosis (HPH) were performed using the following formulas,  $MPH = (F_1 - MP)/MP$  in % and  $HPH = (F_1 - HP)/HP$  in %, where  $F_1$  denotes the average performance of the hybrid, MP denotes the average performance of the two parents, and HP denotes the average performance of the better-parent between two parents. The mode of gene expression patterns including additivity and non-additivity were performed based on Rapp *et al.*<sup>97</sup>, Zhang *et al.*<sup>52</sup>, and Wu *et al.*<sup>31</sup>. The relative expression level of DEGs were categorized into additive and non-additive (dominant and overdominant) classes based on Wu *et al.*<sup>31</sup>. The relative expression levels of DEGs in between two parents were classified as additive. The relative expression levels of DEGs similar to the dominant parent was classified as dominance. The relative expression levels of DEGs either higher (transgressive upregulation) or lower (transgressive downregulation) than both parents were classified as overdominance.

**Validation of RNA-Seq by qPCR.** The quantitative real-time PCR (qPCR) analysis was carried out for the further verification of RNA-Seq data using gene-specific primer sets (Supplementary Table S1). The details of plant culture, sample collection, and RNA extraction have been described in sections 2.1 and 2.2. The cDNA synthesis was carried out for three independent biological replicates against each parent and hybrid genotype. The cDNA was synthesized by Superscript III First-Strand synthesis kit (Invitrogen, California, USA) following the manufacturer's manual. The qPCR reaction mixture consisted of 1  $\mu$ L cDNA in a 20  $\mu$ L reaction volume employing 2 $\times$  qPCR BIO SyGreen Mix Lo-Rox SYBR Green Super-mix with ROX (PCR Biosystems Ltd., London, UK). There were three technical replicates against each biological replicate for qPCR analysis. The primers used for qPCR were designed by primer premier 3.0 (Premier, Canada) (<http://bioinfo.ut.ee/primer3-0.4.0/>) and those were synthesized by Macrogen Company (Seoul, South Korea). The thermal cycling conditions for qPCR comprised of pre-incubation at 95 °C for 10 min, followed by 3 step amplifications at 95 °C for 20 s, 58 °C for 20 s, and 72 °C for 25 s for 40 cycles. The melting temperature comprised of 95 °C for 10 s, 65 °C for 60 s, and 97 °C for 1 s as a default setting. For quantification (Cq), the fluorescence was measured at the end of each cycle, and three technical replicates ( $N = 3$ ) were pooled. Using LightCycler96 (Roche, Germany), amplification, detection, and data analysis were conducted. The *Lf-Actin* from *L. longiflorum* (DQ019459), as an internal control, was used to normalize the gene expression level<sup>98</sup>. The  $2^{-\Delta\Delta Ct}$  method<sup>99</sup> was used to calculate the relative gene expression level.

**Statistical analysis.** The phenotypic data were analyzed following Zhai *et al.*<sup>13</sup> and Ding *et al.*<sup>34</sup>. Analysis of variance (ANOVA) was conducted following a generalized linear model using Statistical Analysis System (SAS) software version 9.1 (SAS Inst. Inc., Cary, NC)<sup>100</sup> to determine the significance of variation for phenotypic traits and qPCR expression level between hybrids and parents.

Received: 9 December 2019; Accepted: 12 May 2020;

Published online: 03 June 2020

## References

1. Darwin, C. The effects of cross and self fertilisation in the vegetable kingdom. *J. Murray & Co.*, 169 pp. (1876).
2. Shull, G. H. The composition of a field of maize. *Am Breeders Assoc Rep* **4**, 296–30 (1908).
3. East, E. M. Inbreeding in corn. *Conn Agric Exp Sta Rep* **1907**, 419–428 (1908).
4. Lippman, Z. B. & Zamir, D. Heterosis, revisiting the magic. *Trends Genet.* **23**, 60–66 (2007).
5. Stupar, R. M. *et al.* Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. *BMC Plant Biol* **8**, 33 (2008).



6. Groszmann, M. *et al.* Changes in 24–ntsiRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proc. Natl. Acad. Sci. USA* **108**, 2617–2622 (2011).
7. Gebreyohannes, G., Koonawootrittriron, S., Elzo, M. A. & Suwanasopee, T. Variance components and genetic parameters for milk production and lactation pattern in an Ethiopian multibreed dairy cattle population. *Asian–Australas J Anim Sci* **26**, 1237 (2013).
8. Lopes, M. S., Bastiaansen, J. W., Harlizius, B., Knol, E. F. & Bovenhuis, H. A genome–wide association study reveals dominance effects on number of teats in pigs. *PLoS one* **9**, e105867 (2014).
9. Kaeppeler, S. Heterosis, one boat at a time, or a rising tide? *New Phytol* **189**, 900–902 (2011).
10. Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D. & Veitia, R. A. Heterosis. *The Plant Cell Online* **22**, 2105–2112 (2010).
11. Nuruzzaman, M. *et al.* Studies on parental variability and heterosis in rice. *Pakistan J Biol Sci* **5**, 1006–1009 (2002).
12. Rahimi, M., Rabiei, B., Samizadeh, H. & Kafighasemi, A. Combining ability and heterosis in rice (*Oryza sativa* L.) cultivars. *J AgrSci Tech-Iran* **12**, 223–231 (2010).
13. Zhai, R. *et al.* Transcriptome analysis of rice root heterosis by RNA–Seq. *BMC Genom* **14**, 19 (2013).
14. Birchler, J. A., Auger, D. L. & Riddle, N. C. In search of the molecular basis of heterosis. *The Plant Cell* **15**, 2236–2239 (2003).
15. Swanson–Wagner, R. A. *et al.* All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proc. Natl. Acad. Sci. USA* **103**, 6805–6810 (2006).
16. Li, A. *et al.* mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *The Plant Cell* **26**, 1878–1900 (2014).
17. Auger, D. L. *et al.* Nonadditive gene expression in diploid and triploid hybrids of maize. *Genetics* **169**, 389–397 (2005).
18. Użarowska, A. *et al.* Comparative expression profiling in meristems of inbred–hybrid triplets of maize based on morphological investigations of heterosis for plant height. *Plant Mol Biol* **63**, 21–34 (2007).
19. Hoecker, N. *et al.* Comparison of maize (*Zea mays* L.) F1–hybrid and parental inbred line primary root transcriptomes suggests organ–specific patterns of nonadditive gene expression and conserved expression trends. *Genetics* **179**, 1275–1283 (2008).
20. Zhang, H. Y. *et al.* A genome–wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. *Mol Plant* **1**, 720–731 (2008).
21. Fujimoto, R., Taylor, J. M., Shirasawa, S., Peacock, W. J. & Dennis, E. S. Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proc. Natl. Acad. Sci. USA* **109**, 7109–7114 (2012).
22. Meyer, S., Pospisil, H. & Scholten, S. Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. *Plant Mol Biol* **63**, 381–391 (2007).
23. Thiemann, A. *et al.* Genome–wide meta–analysis of maize heterosis reveals the potential role of additive gene expression at pericentromeric loci. *BMC Plant Biol* **14**, 88 (2014).
24. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA–Seq. *Nat Methods* **5**, 621–628 (2008).
25. Wang, Z., Gerstein, M. & Snyder, M. RNA–Seq, a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).
26. Wu, X., Li, R., Li, Q., Bao, H. & Wu, C. Comparative transcriptome analysis among parental inbred and crosses reveals the role of dominance gene expression in heterosis in *Drosophila melanogaster*. *Sci Rep* **6**, 21124 (2016).
27. Hansey, C. N. *et al.* Maize (*Zea mays* L.) genome diversity as revealed by RNA–sequencing. *PLoS one* **7**, e33071 (2012).
28. Saxena, R. K., Edwards, D. & Varshney, R. K. Structural variations in plant genomes. *Brief Funct Genomics* **13**, 296–307 (2014).
29. Tian, M. *et al.* Transcriptomic analysis reveals overdominance playing a critical role in nicotine heterosis in *Nicotiana tabacum* L. *BMC Plant Bio* **18**, 48 (2018).
30. Ding, H. *et al.* Heterosis in early maize ear inflorescence development, a genome–wide transcription analysis for two maize inbred lines and their hybrid. *Int J Mol Sci* **15**, 13892–13915 (2014).
31. Zhang, F. *et al.* Genetic analysis and associated SRAP markers for flowering traits of chrysanthemum (*Chrysanthemum morifolium*). *Euphytica* **177**, 15–24 (2011).
32. Meyer, R. C. *et al.* Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. *Plant J* **71**, 669–683 (2012).
33. Gallais, A. Heterosis, its genetic basis and its utilisation in plant breeding. *Euphytica* **39**, 95–104 (1988).
34. McClintock, B. The significance of responses of the genome to challenge 180–199 (World Scientific Pub. Co, Singapore, 1993).
35. Ha, M. *et al.* Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. *Proc. Natl. Acad. Sci. USA* **106**, 17835–17840 (2009).
36. Semel, Y. *et al.* Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. Sci. USA* **103**, 12981–12986 (2006).
37. Krieger, U., Lippman, Z. B. & Zamir, D. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat Genet* **42**, 459–463 (2010).
38. Shahin, A. *et al.* Genetic mapping in *Lilium*, mapping of major genes and quantitative trait loci for several ornamental traits and disease resistances. *Plant Breed* **130**, 372–382 (2011).
39. Frascaroli, E. *et al.* Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. *Genetics* **176**, 625–644 (2007).
40. Meyer, R. C. *et al.* QTL analysis of early stage heterosis for biomass in Arabidopsis. *Theo Appl Genet* **120**, 227–237 (2010).
41. Radoev, M., Becker, H. C. & Ecker, W. Genetic analysis of heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by quantitative trait locus mapping. *Genetics* **179**, 1547–1558 (2008).
42. Shahin, A. *et al.* Generation and analysis of expressed sequence tags in the extreme large genomes. *Lilium Tulipa. BMC Genom* **13**, 640 (2012).
43. Robinson, B. S. Green seraglios: tulips, Turbans, and the global market. *Journal Early Modern Cultural Studies*. 93–122 (2009).
44. Lim, K. B. & Van Tuyl, J. M. Lily. In *Flower Breeding and Genetics* (pp. 517–537). Springer, Dordrecht (2007).
45. Howlader, J., Park, J.-I., Robin, A. H. K., Sumi, K. R. & Nou, I.-S. Identification, Characterization and Expression Profiling of Stress-Related Genes in Easter Lily (*Lilium formolongi*). *Genes* **8**, 172 (2017).
46. Korea International Trade Association (KITA). Available online, <http://www.kita.net>.
47. Marasek-Ciolakowska, A., Nishikawa, T., Shea, D. J. & Okazaki, K. Breeding of lilies and tulips—Interspecific hybridization and genetic background—. *Breed Sci* **68**(1), 35–52 (2018).
48. Ramzan, F. *et al.* Evaluation of F1 and BC1 hybrids of *Lilium lancifolium* × Asiatic hybrid ‘Chianti’ by morphological analysis and fluorescence *in situ* hybridization. *J Hort Sci Biotech* **93**(6), 614–624 (2018).
49. Roh, M. S., Griesbach, R. J. & Lawson, R. H. Evaluation of interspecific hybrids of *Lilium longiflorum* and *L. elegans*. In: *International Symposium on the Genus Lilium 414* (pp. 101–110) (1994, August).
50. Roh, M. S., Griesbach, R. J., Gross, K. C. & Line, M. Identification and evaluation of the interspecific hybrid between *Lilium longiflorum* and *L. callosum*. In: *International Symposium on the Genus Lilium 414* (pp. 111–124) (1994, August).
51. Dhiman, M. R., Parkash, C., Kumar, R., Guleria, M. S. & Dhiman, M. Studies on genetic variability and heritability in Asiatic hybrid lily (*Lilium x elegans* L.). *Mol Plant Breed* **6** (2015).
52. Zhang, J. *et al.* Transcriptome analysis of interspecific hybrid between *Brassica napus* and *B. rapa* reveals heterosis for oil rape improvement. *Int J Genomics* **230985**, 1–11 (2015).
53. Yang, M., Zhu, L., Xu, L., Pan, C. & Liu, Y. Comparative transcriptomic analysis of the regulation of flowering in temperate and tropical lotus (*Nelumbo nucifera*) by RNA–Seq. *Ann Appl Biol* **165**, 73–95 (2014).



54. Li, X., Wei, Y., Nettleton, D. & Brummer, E. C. Comparative gene expression profiles between heterotic and non-heterotic hybrids of tetraploid *Medicago sativa*. *BMC Plant Biol* **9**, 107 (2009).
55. Guo, M. *et al.* Genome-wide transcript analysis of maize hybrids, allelic additive gene expression and yield heterosis. *Theo Appl Genet* **113**, 831–845 (2006).
56. Ge, X. *et al.* Transcriptomic profiling of mature embryo from an elite super-hybrid rice LYP9 and its parental lines. *BMC Plant Biol* **8**, 114 (2008).
57. Wei, G. *et al.* A transcriptomic analysis of super hybrid rice LYP9 and its parents. *Proc. Natl. Acad. Sci. USA* **106**, 7695–7701 (2009).
58. He, G. *et al.* Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* **22**, 17–33 (2010).
59. Song, G. S. *et al.* Comparative transcriptional profiling and preliminary study on heterosis mechanism of super-hybrid rice. *Mol Plant* **3**, 1012–1025 (2010).
60. Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev* **72**, 317–364 (2008).
61. McCarty, D. R., Carson, C. B., Stinard, P. S. & Robertson, D. S. Molecular analysis of viviparous-1, an abscisic acid-insensitive mutant of maize. *Plant Cell* **1**, 523–532 (1989).
62. Ulmasov, T., Hagen, G. & Guilfoyle, T. J. ARF1, a transcription factor that binds to auxin response elements. *Science* **276**, 1865–1868 (1997).
63. Laporte, D. *et al.* Glutaredoxin GRXS13 plays a key role in protection against photooxidative stress in Arabidopsis. *J Exp Bot* **63**, 503–515 (2011).
64. Nozaki, H. *et al.* Momilactone A and B as allelochemicals from moss *Hypnum plumaeforme*, first occurrence in bryophytes. *Biosci Biotechnol Biochem* **71**, 3127–3130 (2007).
65. Schmidlin, L. *et al.* A stress-inducible resveratrol O-methyltransferase involved in the biosynthesis of pterostilbene in grapevine. *Plant Physiol* **148**, 1630–1639 (2008).
66. Grieneisen, V. A., Xu, J., Marée, A. F., Hogeweg, P. & Scheres, B. Auxin transport is sufficient to generate a maximum and gradient guiding root growth. *Nature* **449**, 1008 (2007).
67. Benjamins, R. & Scheres, B. Auxin, the looping star in plant development. *Annu Rev Plant Biol* **59**, 443–465 (2008).
68. Kleine-Vehn, J. *et al.* PIN auxin efflux carrier polarity is regulated by PINOID kinase-mediated recruitment into GNOM-independent trafficking in Arabidopsis. *Plant Cell* **21**, 3839–3849 (2009).
69. Williamson, J. D., Stoop, J. M., Massel, M. O., Conkling, M. A. & Pharr, D. M. Sequence analysis of a mannitol dehydrogenase cDNA from plants reveals a function for the pathogenesis-related protein ELI3. *Proc Natl Acad Sci USA* **92**, 7148–7152 (1995).
70. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **26**, 148–153 (1998).
71. Bittner-Eddy, P. D., Crute, I. R., Holub, E. B. & Beynon, J. L. RPP13 is a simple locus in Arabidopsis thaliana for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *Plant J* **21**, 177–188 (2000).
72. Rose, L. E. *et al.* The maintenance of extreme amino acid diversity at the disease resistance gene, RPP13, in *Arabidopsis thaliana*. *Genetics* **166**, 1517–1527 (2004).
73. Massengo-Tiassé, R. P. & Cronan, J. E. Diversity in enoyl-acyl carrier protein reductases. *Cell Mol Life Sci* **66**, 1507–1517 (2009).
74. Chae, K. *et al.* Arabidopsis SMALL AUXIN UP RNA63 promotes hypocotyl and stamen filament elongation. *Plant J* **71**, 684–697 (2012).
75. Lin, S. J., Pufahl, R. A., Dancis, A., O'Halloran, T. V. & Culotta, V. C. A role for the *Saccharomyces cerevisiae* ATX1 gene in copper trafficking and iron transport. *J Biol Chem* **272**, 9215–9220 (1997).
76. Chu, C. C. *et al.* A copper chaperone for superoxide dismutase that confers three types of copper/zinc superoxide dismutase activity in Arabidopsis. *Plant Physiol* **139**, 425–436 (2005).
77. Rose, J. K. & Bennett, A. B. Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls, parallels between cell expansion and fruit ripening. *Trends Plant Sci* **4**, 176–183 (1999).
78. Ni, Z. *et al.* RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nat Struct Mol Biol* **21**, 686–695 (2014).
79. Marcus, J. P., Green, J. L., Goulter, K. C. & Manners, J. M. A family of antimicrobial peptides is produced by processing of a 7S globulin protein in *Macadamia integrifolia* kernels. *Plant J* **19**, 699–710 (1999).
80. Schröder, E. & Eaton, P. Hydrogen peroxide as an endogenous mediator and exogenous tool in cardiovascular research, issues and considerations. *Curr Opin Pharmacol* **8**, 153–159 (2008).
81. Kim, Y. W. & Byzova, T. V. Oxidative stress in angiogenesis and vascular disease. *Blood* **123**, 625–631 (2014).
82. Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N. & Schuster, G. HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell* **15**, 1480–1495 (2003).
83. Nakamura, T., Schuster, G., Sugiura, M. & Sugita, M. Chloroplast RNA-binding and pentatricopeptide repeat proteins. *Biochem Soc Trans* **32**, 571–574 (2004).
84. Kayum, M. A. *et al.* Identification and expression analysis of WRKY family genes under biotic and abiotic stresses in *Brassica rapa*. *Mol Genet Genomics* **290**, 79–95 (2015).
85. Chen, L. & Bush, D. R. LHT1, a lysine- and histidine-specific amino acid transporter in Arabidopsis. *Plant Physiol* **115**, 1127–1134 (1997).
86. Forkmann, G., Heller, W. & Grisebach, H. Anthocyanin biosynthesis in flowers of *Matthiola incana* flavanone 3- and flavonoid 3'-hydroxylases. *ZNaturforsch* **35**, 691–695 (1980).
87. Wellmann, F., Matern, U. & Lukačín, R. Significance of C-terminal sequence elements for Petunia flavanone 3β-hydroxylase activity. *FEBS Lett* **561**, 149–154 (2004).
88. Laila, R. *et al.* Developmental and genotypic variation in leaf wax content and composition, and in expression of wax biosynthetic genes in *Brassica oleracea* var. Capitata. *Front Plant Sci* **7**, 1972 (2017).
89. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic, a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
90. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-Seq, reference generation and analysis with Trinity. *Nat Protoc* **8**, 1494–1512 (2013).
91. Li, B. & Dewey, C. N. RSEM, accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
92. Lily database. Available online, [http://210.110.86.160/Lidb/Lilidb\\_Home.html](http://210.110.86.160/Lidb/Lilidb_Home.html).
93. Oliveros, J. C. Venny. An interactive tool for comparing lists with Venn's diagrams. 2007–2015, (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>).
94. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**, 14863–14868 (1998).
95. Conesa, A. *et al.* Blast2GO, a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).

96. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**(suppl\_1), D277–D280 (2004).
97. Rapp, R. A., Udall, J. A. & Wendel, J. F. Genomic expression dominance in allopolyploids. *BMC Biol.* **7**, 18 (2009).
98. Xin, H. *et al.* Cloning and characterization of HsfA2 from Lily (*Lilium longiflorum*). *Plant Cell Rep* **29**, 875–885 (2010).
99. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* **25**, 402–408 (2001).
100. SAS. SAS/STAT, statistical analysis systems for windows. SAS Institute Inc., Cary, N.C., USA. Release 9.1, p. 423 (2004).

## Acknowledgements

A plant material was provided by Department of Floriculture, Korea National Agricultural College, Hwaseong 445–893, Republic of Korea. This work was supported by Golden Seed Project (Center for Horticultural Seed Development) (Grant No. 213007–05–4–CG100), Ministry of Agriculture, Food and Rural Affairs, Korea.

## Author contributions

The work presented here was conducted in collaboration among all authors. Ill–Sup Nou and Jong–In Park conceived and planned the experiments. Jewel Howlader designed and carried out all experiments, analyzed the data and drafted the manuscript. Sathishkumar Natarajan and Manosh Kumar Biswas assisted in bioinformatics analysis. Kanij Rukshana Sumi participated in bioinformatics analysis, samples collection and qPCR validation. Cheon Young Song provided plant materials and improved the technical sites of the project. Arif Hasan Khan Robin participated in analyzing the data, critically revised the manuscript and gave suggestions for improving it. All authors have read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-65909-x>.

**Correspondence** and requests for materials should be addressed to I.-S.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020