

An efficient experiment design helps to identify differential expressed genes (DEGs) in RNA-seq data in studies of plant qualitative traits

Xingfeng Li^{1,2}, Zhenqiao Song^{2*}, Yinguang Bao², Honggang Wang^{1,2}**

¹State Key Laboratory of Crop Biology, Shandong Agricultural University, Taian, Shandong, China

²Agronomy College, Shandong Agricultural University, Taian, Shandong, China

*Xingfeng Li and Zhenqiao Song contributed equally to the work.

** Corresponding author: Honggang Wang.

Xingfeng Li Phone: 86-13515481834, Fax: 86-538-8242226, E-mail: lixf@sdau.edu.cn

Zhenqiao Song Phone: 86-538-8246821, Fax: 86-538-8242226, E-mail: szq@sdau.edu.cn

Yinguang Bao Phone: 86-538-8246821, Fax: 86-538-8242226, E-mail: 42008229@qq.com

Honggang Wang Phone: 86-538-8246821, Fax: 86-538-8242226, E-mail: b00007@sdau.edu.cn

** Corresponding author: Xingfeng Li and Honggang Wang

Running Title: Designs help identifying DEGs in RNA-seq of qualitative traits

Abstract:

In this study, we conducted comparative transcriptome analysis between homozygous dominant parent and heterozygous F₁ hybrid with homozygous recessive parent in qualitative trait study of common wheat (*Triticum aestivum* L.). Two sets of near-isogenic lines (NILs) were used: one set of NILs carrying powdery mildew resistance and susceptible *Pm2* alleles, the other set of NILs carrying different awn inhibition gene *BI* alleles. The results demonstrated that 2,932 DEGs were identified between L031 (*Pm2Pm2*) and Chancellor (*pm2pm2*), while 1,494 DEGs presented between F₁ hybrid (*Pm2pm2*) and Chancellor, the co-regulated DEGs were 1,028. For the wheat awn inhibition gene *BI* test, 720 DEGs were identified between SN051-2 (*BIBI*) and SN051-1 (*b1b1*), and 231 DEGs were identified between F₁ hybrid (*B1b1*) and SN051-1, the co-regulated DEGs were 180. Hierarchical clustering analysis of co-regulated DEGs showed that dominant parent and F₁ hybrid were clustered as the nearest neighbors, while recessive parent showed an apparent departure. The results showed that the overlapping DEGs between dominant parent and F₁ hybrid with recessive parent reduced the number of interested DEGs to only one-quarter (or one-third) of that between dominant and recessive parent, these overlapping loci could provide insights into molecular mechanisms that are affected by causal mutations.

Keywords

Comparative transcriptomics, qualitative traits, F₁ hybrid, Dominant and recessive parent, Near-isogenic lines, *Triticum aestivum*

Introduction

RNA sequencing (RNA-seq) is an increasingly popular and effective approach for transcriptome analysis, with both high-throughput and high resolution capabilities for quantitative measurement of gene expression, detection of splice variants and single-nucleotide variations, and discovery of novel coding and noncoding transcripts (Shendure, 2008; Wang et al., 2009; Grabherr et al., 2011; Wang et al., 2014).

The identification of differentially expressed genes (DEGs) between two or more genotypes, tissues, time-points or conditions is a fundamental task in the analysis of RNA-seq data (Robles et al., 2012; Soneso and Delorenzi, 2013). The results obtained from the analysis of DEGs may facilitate our understanding of the molecular basis of phenotypic variation, disease resistance, developmental changes, and a diverse range of biological problems. Recently, some research groups have elucidated the performance characteristics of RNA-seq, different sources of measurement noise, and effects of different analysis approaches and algorithms (Kratz and Caminci, 2014; Li et al., 2014; Risso et al., 2014; Consortium 2014). Such studies have helped to improve the detection power of DEGs in RNA-seq analysis.

According to the theory of qualitative Mendelian traits, a cross between a homozygous dominant and a homozygous recessive line produces a heterozygous F_1 hybrid that exhibits the same phenotype as the dominant parent. Therefore, the gene expression profiles of the causal genes and the affected downstream genes in the F_1 hybrid would be similar to those of the dominant line. However, the F_1 hybrid contains two alleles, one contributed by each parent; therefore, the differentially expressed genes between the F_1 hybrid and the recessive parent could be very useful and more accurate for identifying important DEGs.

However, few studies have investigated experimental designs in this manner. Transcriptome studies of F₁ hybrids and its parents have mostly been used to study phenomena associated with regulatory divergence (*cis/trans*), additive or transgressive gene expression, and allele-specific gene expression, thus providing an opportunity to speculate on mechanisms of heterosis (Paschold et al., 2012; Song et al., 2013; Bell et al., 2013; Song et al., 2014).

In this study, we conducted a comparative transcriptome analysis of the F₁ hybrid with its parents for two qualitative traits of common wheat. These two RNA-seq experiments both revealed that DEGs between the F₁ hybrid and recessive parent were clearly lower than that between the dominant and recessive parents, and the overlapping DEGs between them could provide more accurate information regarding the molecular mechanisms that are affected by the causal mutation.

Materials and Methods

Materials

Two sets of near-isogenic lines (NILs) of wheat were used: one set of NILs carrying powdery mildew resistance and the susceptible *Pm2* alleles, and the other set of NILs carrying differences in the wheat awn inhibition gene *BI*.

The wheat cultivar Chancellor (susceptible, *pm2pm2*) and its near-isogenic line L031 (resistant), which carries the powdery mildew resistance gene *Pm2*, were used in this study. They were obtained by crossing Ulka (donor of resistance gene) with Chancellor, and then backcrossed with Chancellor for 7 generations, and finally self-crossed and selected under *Blumeria graminis* f. sp. *tritici* inoculation (Wang et al. 2011). The plants of Chancellor, L031 and its F₁ hybrid (resistant) were raised in a growth chamber with a 16-h photoperiod at 18°C. The *Bgt.* isolate E09 was used for

infection at the trifoliate stage. Sixteen hours after inoculation, leaf samples were collected, pooled from three plants, immediately immersed in liquid nitrogen, and stored at -80°C until processing.

A pair of wheat lines with different awn lengths, SN051-1 (long awn, *b1b1*) and SN051-2 (short awn, *B1B1*), were used in the current study. They were developed from the F_8 progeny of Octoploid *Triticale* Jinsong49 and Octoploid *Trititrigia* Xiaoyan7430, and then were self-crossed for 5 generations (Du et al. 2010). The morphological, genetic and molecular marker analyses showed that they differed in the wheat awn inhibition gene *B1*: line SN051-2 contained the dominant allele *B1*, and SN051-1 possessed the recessive allele *b1*. Genotypic data via the use of DNA markers showed that SN051-1 and SN051-2 were truly isogenic (additional file 1). The hybrid was obtained from the cross of SN051-1 \times SN051-2. The SN051-1, SN051-2, and F_1 hybrid were planted in the same field at Shandong Agricultural University agronomy farm, Tai'an, Shandong, China. Pools of young spikes from ten plants were hand-dissected when they developed into the anther connective tissue formation phase, while the awn primordia at the distal end of the lemma began to elongate in the awned genotypes. The young spike samples of SN051-1, SN051-2, and the hybrid were collected, immediately frozen in liquid nitrogen, and then stored at -80°C until use.

RNA isolation and library preparation for sequencing

Total RNA was extracted using the RNAsimple total RNA kit according to the manufacturer's instructions (Tiangen, Beijing, China). The quantity and quality of the extracted RNA were verified by gel electrophoresis and spectrophotometry. Oligo(dT) - coated magnetic beads were used to enrich the mRNA, which was then broken into fragments using fragmentation buffer. Using these cleaved mRNA fragments as templates, first- and second-strand cDNAs were synthesized. The double-stranded cDNA was further purified using the QiaQuick PCR extraction kit (Qiagen, Hilden,

Germany), resolved for final reparation and poly (A) addition, and then connected using different sequencing adapters. The expressed sequence tag (EST) libraries were constructed by PCR amplification after checking the quality by agarose gel electrophoresis and sequencing using an Illumina HiSeq™ 2000 platform by Biomarker Technology Co., Ltd (Beijing, China).

***De novo* transcriptome assembly and annotation**

To obtain clean reads, raw reads such as reads with adaptors, reads in which unknown bases represented more than 5% of the total bases, and low-quality reads (percentage of low-quality bases with a quality value ≤ 5 was more than 50% of a read) were removed. The clean reads were then assembled *de novo* into longer contigs based on their overlap regions using the Trinity platform. Different contigs from another transcript and their distance were further recognized by mapping clean reads back to the corresponding contigs based on their paired-end information, and thus, the sequences of the transcripts were determined. These were defined as unigenes. All of the unique sequences were annotated using BLASTX against the NCBI non-nucleotide (Nt) sequence database, NCBI non-redundant (Nr) protein database, Swiss-Prot, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Clusters of Orthologous Groups (COG) to predict and classify the functions.

Annotation of expression

To evaluate the depth of coverage, all of the usable reads were realigned to each unitranscripts using Short Oligonucleotide Analysis Package (SOAP) aligner (<http://soap.genomics.org.cn/soapaligner.html>) and then normalized into RPKM values (reads per kb per million reads; Mortazavi et al., 2008). The DEG analysis was then performed using the Bioconductor package DESeq based on the ratio of the RPKM values. The false discovery rate (FDR)

control method was used to identify the threshold of the P-value in multiple tests to compute the significance of differences in transcript abundance. In the present analysis, only unitranscripts with an absolute \log_2 ratio ≥ 1 and an FDR significance score < 0.01 were used for subsequent analyses. Hierarchical clustering and heat map generation were performed in R. The gene expression data were \log_2 -transformed and then quantile-normalized prior to generating the heat map for direct comparison of the data.

Validation of DEG expression by quantitative real time PCR (qRT-PCR)

Gene-specific primers were designed using Primer software (version 5.0) and the *Actin* gene was used as the internal control gene. The qRT-PCR reaction for target gene transcript amplification was carried out in a final volume of 25 μ L containing PCR buffer, 1 mM $MgCl_2$, 0.2 mM dNTP, 1 μ L of SYBR green I, 1U of Taq, 0.4 μ M of each forward and reverse primers, and 2 μ L diluted cDNA. The PCR reaction conditions were: denaturation at 95 °C for 5 min followed by 40 cycles of 95 °C for 10 s, annealing at the appropriate temperature (from 57 to 61 °C) for 30 s, extension at 72°C for 30 s, with a final extension at 72 °C for 10 min. All reactions were done in triplicate. The amplification data were analyzed using iQ5 software version 1.0 (Bio-RAD, USA). The threshold cycle (Ct) values of the triplicate PCRs were averaged and relative quantification of the transcript levels was undertaken using the comparative Ct method. The ΔCt value of the calibrator (the sample with the highest ΔCt value) was subtracted from every other sample to produce the $\Delta\Delta Ct$ value, and $2^{-\Delta\Delta Ct}$ was taken as the relative expression level for each sample.

Results

RNA-seq results of wheat NILs with contrasting disease reactions

For the powdery mildew resistance gene *Pm2* test, leaf samples of the wheat cultivar Chancellor (susceptible, *pm2pm2*), the near-isogenic line L031 (resistance, *Pm2Pm2*) and its F₁ hybrid were inoculated with *Bgt*. E09 and then used for RNA-seq analysis. After cleaning and checking the read quality, 4.16 Gb, 3.62 Gb and 4.06 Gb of clean data were generated, respectively. Among the clean reads, more than 90% had a read quality of Q30 (sequencing error rate, 0.1%) or higher. These reads were assembled *de novo* using Trinity platform software, resulting in 316,787 transcripts and 112,033 unigenes with an N50 length of 1,742 bp and 1,158 bp, respectively. Among these, 63,210 unigenes were annotated after BLAST searches using the Nr, Swiss-Prot, KEGG, COG and GO databases. The percentage of aligned reads mapping to unigenes in the library was generally approximately 80%, which indicated an acceptable quality of the aligned reads.

The gene expression levels were measured as reads per kilobase per million reads (RPKM), and putative DEGs were identified using a false discovery rate (FDR) of less than 0.01 and a fold change greater than 2. The box plots of the relative log RPKM values for each RNA-seq library showed little distributional differences among the three libraries (Fig. 1A). As shown in Figure 1B, comparison of the L031 and Chancellor libraries (Pm vs. Ch) showed that 2,932 genes (account for 4.64% of annotated unigenes) were differentially expressed, 1,438 and 1,494 of these DEGs were up- and down-regulated after inoculation, respectively. Concomitantly, 1,419 DEGs (account for 2.24% of annotated unigenes) were identified from 'F₁ vs. Ch' comparison, 692 DEGs were up-regulated, and 727 DEGs were down-regulated, respectively. Venn diagram of the differentially co-regulated genes revealed a total of 1,028 DEGs (account for 1.63% of annotated unigenes) presented in comparison

between resistant (L031 and the F₁) and susceptible genotypes (Chancellor). Four hundred and seventy-seven DEGs showed relatively higher transcript levels in resistant seedlings than susceptible seedlings, while 551 DEGs showed the opposite expression pattern (Fig. 1B).

GO enrichment analysis were conducted in order to determine the functions of the DEGs, these DEGs were categorized into 49 functional groups in the three main categories, “cellular components (CC)”, “molecular functions (MF)” and “biological processes (BP)” (Figure 1C). The term “cell”, “organelle” and “cell part” were significantly enriched in the CC category, term “catalytic activity” and “binding” were significantly enriched in the MF category, term “metabolic process” and “cellular process” were significantly enriched in the BP category. GO analysis also showed DEG numbers enriched in most terms predicted between F₁ and Chancellor were reduced to only 1/2 or 2/3 of that between L031 and Chancellor, while term “macromolecular complex” in CC and term “structural molecule activity” did not (Fig. 1C).

Hierarchical clustering analysis of 1028 co-regulated DEGs in Chancellor, L031 and the F₁ hybrid was performed (Fig. 1D). L031 and F₁ hybrid were clustered as the nearest neighbors as anticipated, while Chancellor showed an apparent departure. Cluster analysis of the RPKM data showed that the differentially co-expressed DEGs could be grouped into seven clusters; clusters D and F were enriched with transcripts that showed large fold changes in expression between resistant and recessive genotypes.

Eight transcripts, which encoding the LRR family of receptor serine/threonine-protein kinases and up-regulated by *Bgt.* infection in the resistant genotypes (L031 and F₁), were selected for quantitative real-time PCR (qRT-PCR) analyses (Fig. 2). The RPKM values of these transcripts were much lower (some were 0) in Chancellor, hence these genes may be responsible for perception of fungal invasion

and activation of down-stream signaling cascades to induce defense responses. qRT-PCR analyses showed that the expression patterns of seven genes were highly consistent with the RNA-Seq data, suggesting that our transcriptome analysis was accurate and reliable.

RNA-seq results of wheat NILs that differ in wheat awn inhibition gene *BI*

For the wheat awn inhibition gene *BI* test, global transcriptional changes of the young spike samples at the anther connective tissue formation phase of wheat lines SN051-1 (awned, *b1b1*) and SN051-2 (awnless, *B1B1*), and its hybrid were compared. Approximately 9.34, 9.18 and 10.27 Gbs of clean data were generated using the Illumina HiSeq™ 2500 sequencing device, respectively. The Q30 values were higher than 85.0%. *De novo* assemblies using the Trinity program generated 265,244 transcripts and 96,360 unigenes with an N50 length of 1,628 and 1,231 bp, respectively. A total of 49,289 unigenes were annotated.

Transcriptome profiles of young spike samples of SN051-2 (Al), SN051-1 (An) and its F₁ hybrid were compared in order to obtain a better understanding of awn development at the transcriptome level. The box plots of the relative log FPKM values based on all of the genes showed profiles similar to the profiles obtained in the *Pm2* test (Fig. 3A). A total of 720 and 231 DEGs were identified by pairwise comparisons of the RNA-seq data between ‘SN051-2 vs. SN051-1’ (Al vs. An) and ‘F₁ hybrid vs. SN051-1’ (F₁ vs. An), with 549 and 168 up-regulated and 171 and 63 down-regulated genes, respectively (Fig. 3B). The overlap analysis between ‘Al vs. An’ and ‘F₁ vs. An’ identified 180 DEGs (account for 0.36% of annotated unigenes) that showed constant differential expression (134 up- and 46 down-regulated) (Fig. 3B), representing only one-quarter of the DEGs between dominant and recessive parent.

GO enrichment analysis revealed that interested DEGs between the awnless and awned genotypes

were categorized into 36 functional groups in CC, MF and BP categories. The term “cell”, “organelle” and “cell part” in the CC category, term “catalytic activity” and “binding” in the MF category, and term “metabolic process” and “cellular process” in the BP category were also significantly enriched (Figure 3C). DEG numbers between F₁ and SN051-1 enriched in most terms were significantly down-regulated than that between SN051-2 and SN051-1.

Hierarchical cluster analysis showed that dominant parent SN051-2 and F₁ hybrid tended to cluster together as expected, which revealed similar overall gene expression profiles between them (Fig. 3D). Clustering analysis arranged 180 significantly DEGs into five groups. Cluster B was enriched with 44 transcripts that were uniquely expressed in the awnless genotype, and although most of the results were hypothetical proteins, they might be considered genes involved in regulation of awn development. However, this hypothesis requires further investigation.

Among 46 down-regulated DEGs (overlapped between ‘Al vs. An’ and ‘F₁ vs. An’), four transcripts were selected for quantitative real-time PCR (qRT-PCR) analyses (Fig. 4). Results showed gene expression trend of these DEGs were also highly consistent with to RNA-Seq data.

Discussion

Although differences in expression between wild-type and mutant isolates can help to answer biological questions, DEG analysis does not generally facilitate the identification of an initial list of potential candidate genes. Even in RNA-seq and DEG analyses of qualitative traits, thousands of DEGs have been identified between near-isogenic lines (Zhou et al., 2010; Zhang et al., 2014). Some of the DEGs are likely from the result of measurement noise or biological complexity, whereas a wide range of the DEGs likely demonstrate the downstream effects of the causal genes (Cai et al.,

2014; Lin et al., 2014). An increase in the sequencing depth and numbers of samples would decrease these biases to some extent (McIntyre et al., 2011; Rapaport et al., 2013).

In the qualitative Mendelian traits study, F_1 hybrid exhibits the same phenotype as the dominant parent, the gene expression profiles of the causal genes and the affected downstream genes between the F_1 hybrid and its dominant parent would be similar. In that means, the DEGs presented between the F_1 hybrid and the recessive parent could be very useful and more accurate for identifying important DEGs. In this study, according to all of the DEGs statistics for the 4 pairwise comparisons in the two treatments, we found that the DEG numbers between the F_1 and recessive parent was clearly less than that between the dominant and recessive parent. Using the DEGs that overlapped between the homozygous dominant parent and heterozygous F_1 hybrid with the homozygous recessive parent condensed the interested DEGs to 1/3 or 1/4 of DEGs between the dominant and recessive parent.

This finding indicated that we could compare the F_1 generation with the susceptible parent to reduce the number of false positive genes and improve the veracity of the experiment. In other words, although this approach does not provide a direct identification of the ‘Trigger’ gene of interest, it brings us closer to the answer.

Previous studies proved that differential gene expression in F_1 hybrid and its parents may be very complex. For example, F_1 hybrid could accumulate levels of transcript equal to the mid-parent (additivity), the high or low parent (high or low parent dominance), above the high parent (over-dominance), or below the low parent (under-dominance) (Swanson-Wagner et al., 2006).

Recent studies have demonstrated that mono-allele specific expression play important roles in development and stress induced responses, thus contributed to heterosis in plants (Song et al., 2013;

Song et al., 2014; Guo et al., 2015). The object of the present study were to study the molecular mechanism of the qualitative traits, but we should not ignored the fact that differential expressed genes between F₁ hybrid and parents could be the results of induction and suppression of the genes caused by genetic and epigenetic alteration. These genes also possess the biological information, some of these DEGs should be studied in detail while should not be considered as the false positives in hasty.

Conclusions

Two different RNA-seq experiment of qualitative trait in wheat showed that the differentially expressed genes between F₁ and recessive parent were lower than those observed between dominant and recessive parents, overlapping loci between the homozygous dominant parent and heterozygous F₁ hybrid with the homozygous recessive parent could reduce the number of interest DEGs to only one-quarter (or one-third) of the total number of DEGs between dominant and recessive parent, which would provide more accurate information regarding the molecular mechanisms that are affected by the causal mutation.

Additional file

Additional Table S1. Percentage of the polymorphic markers between L031 and Chancellor.

Additional Table S2. Percentage of the polymorphic markers between SN051-1 and SN051-2.

Additional Table S3. The 1,028 overlap DEGs in the comparison between resistant (L031 and the F1) and susceptible genotypes (Chancellor) in *Pm2* test.

Additional Table S4: The 180 overlap DEGs between ‘Al vs. An’ and ‘F₁ vs. An’.

Additional Figure S1. Expression patterns of the 8 R protein-like transcripts which were up-regulated in L031 and F1 by RNA-seq data

Additional figure S2: Expression patterns of the 4 DEGs which were down-regulated in SN051-1 by RNA-seq data.

Competing Financial Interests

The authors have declared that no competing interests exist.

Acknowledgments

This work was supported by The National Key Research and Development Program of China (2016YFD0102000) and National Natural Science Foundation of China (31671675), Natural Science Foundation of Shandong Province (ZR2015CM034).

References

1. Bell, G. D., Kane, N. C., Rieseberg, L. H. & Adams, K. L. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol Evol* **5**, 1309-1323 (2013).
2. Cai, W. et al. Genome-wide analysis of regulation of gene expression and H3K9me2 distribution by JIL-1 kinase mediated histone H3S10 phosphorylation in *Drosophila*. *Nucleic Acids Res* **42**, 5456-5467 (2014).
3. Consortium, S. M.-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and

information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* **32**, 903-914 (2014).

4. Du B, Cui F, Wang H & Li X. Characterization and Genetic Analysis of Near-isogenic Lines of Common Wheat for Awn-inhibitor Gene B1. *Molecular Plant Breeding* (in Chinese) **8**, 259-264 (2010)

5. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652 (2011).

6. Guo, Z. et al. Global epigenomic analysis indicates that epialleles contribute to Allele-specific expression via Allele-specific histone modifications in hybrid rice. *BMC Genomics* **16**, 232 (2015).

7. Kratz, A. & Carninci, P. The devil in the details of RNA-seq. *Nat Biotechnol* **32**, 882-884 (2014).

8. Li, S. et al. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat Biotechnol* **32**, 915-925 (2014).

9. Lin, F. et al. Molecular response to the pathogen *Phytophthora sojae* among ten soybean near isogenic lines revealed by comparative transcriptomics. *BMC Genomics* **15**, 18 (2014).

10. McIntyre, L. M. et al. RNA-seq: technical variability and sampling. *BMC Genomics* **12**, 293 (2011).

11. Paschold, A. et al. Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. *Genome Res* **22**, 2445-2454 (2012).

12. Rapaport, F. et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* **14**, R95 (2013).

13. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor

analysis of control genes or samples. *Nat Biotechnol* **32**, 896-902 (2014).

14. Robles, J. A. et al. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics* **13**, 484 (2012).

15. Shendure, J. The beginning of the end for microarrays? *Nat Methods* **5**, 585-587 (2008).

16. Sonesson, C. & Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14**, 91 (2013).

17. Song, G. et al. Global RNA sequencing reveals that genotype-dependent allele-specific expression contributes to differential expression in rice F1 hybrids. *BMC Plant Biol* **13**, 221 (2013).

18. Song, G. et al. The phenotypic predisposition of the parent in F1 hybrid is correlated with transcriptome preference of the positive general combining ability parent. *BMC Genomics* **15**, 297 (2014).

19. 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 U S A* **103**, 6805-6810 (2006).

20. Wang L, Zhu Y, Li X & Wang H. Screening for SSR markers linked to wheat powdery mildew resistance gene Pm2. *Journal of Plant Protection* (in Chinese) **38**, 216-220 (2011).

21. Wang, C. et al. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol* **32**, 926-932 (2014).

22. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63 (2009).

23. Zhang, Y. et al. The Aux/IAA gene *rum1* involved in seminal and lateral root formation controls

- 1 vascular patterning in maize (*Zea mays* L.) primary roots. *J Exp Bot* **65**, 4919-4930 (2014).
- 2 24. Zhou, Y. L. et al. Genome-wide gene responses in a transgenic rice line carrying the maize
- 3 resistance gene *Rxo1* to the rice bacterial streak pathogen, *Xanthomonas oryzae* pv. *oryzicola*.
- 4 *BMC Genomics* **11**, 78 (2010).

Figure Legends

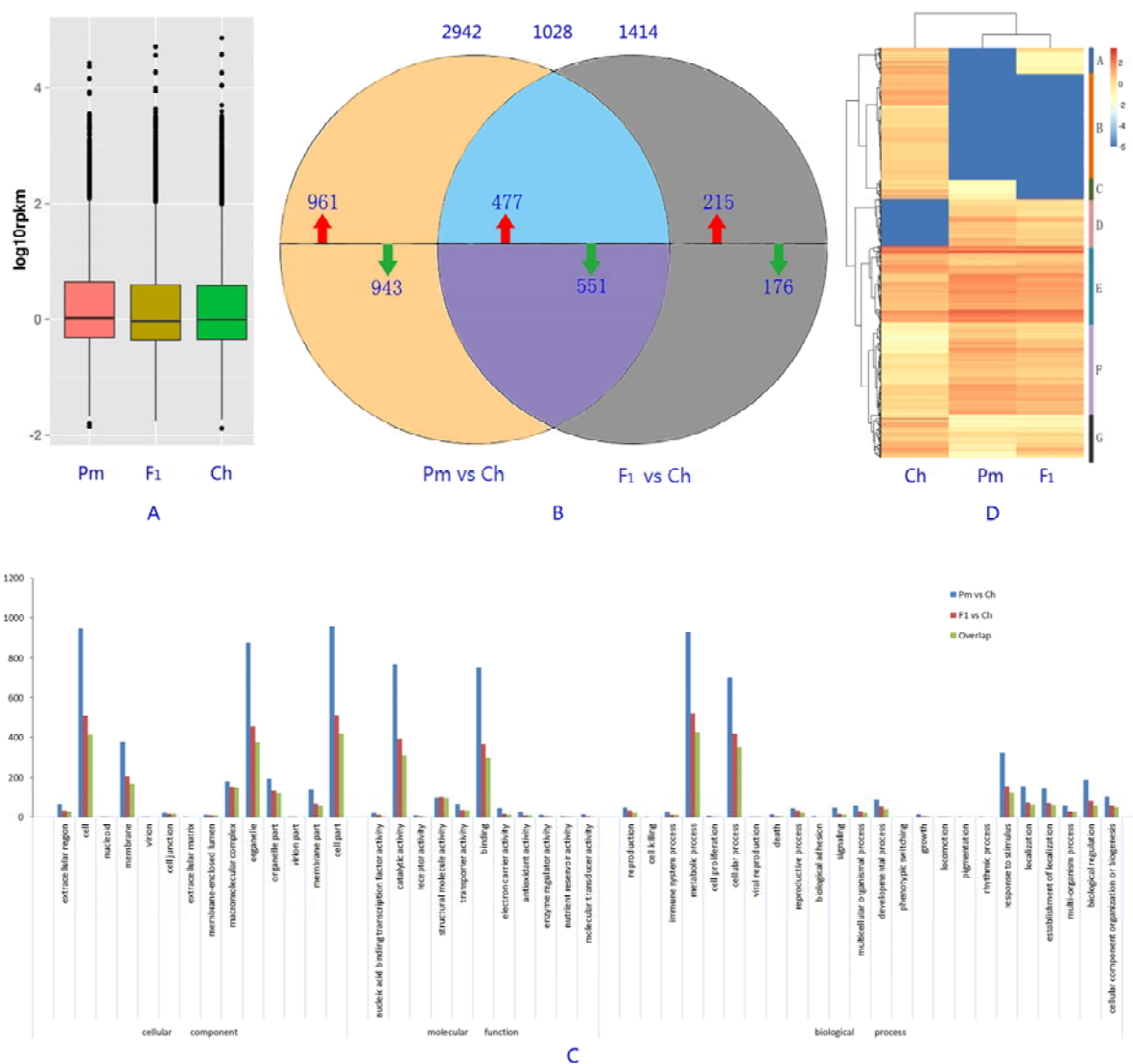


Figure 1 RNA-seq analysis of L031, Chancellor and its F₁ hybrid after inoculation with *Bgt*.

A. Box plots of the relative log FPKM values based on all of the genes for each RNA-seq library.

B. Venn diagram showing the number of differentially co-expressed genes among Chancellor (susceptible, *pm2*), L031 (resistant, *Pm2*) and its F₁ hybrid. (Red and blue arrows indicate up- and down-regulated DEGs, respectively).

C. GO classifications of DEGs. The results are summarized into three main categories: 'biological

processes, cellular components, and molecular functions’.

D. Heat map showing all of the DEGs among the three samples. The gene expression data were \log_2 -transformed and then quantile-normalized prior to generating the heat map for direct comparison of the data. DEGs (red and blue indicate up- and down-regulated DEGs, respectively) for each sample were mapped by lane. Hierarchical clustering and heat map generation were performed in R.

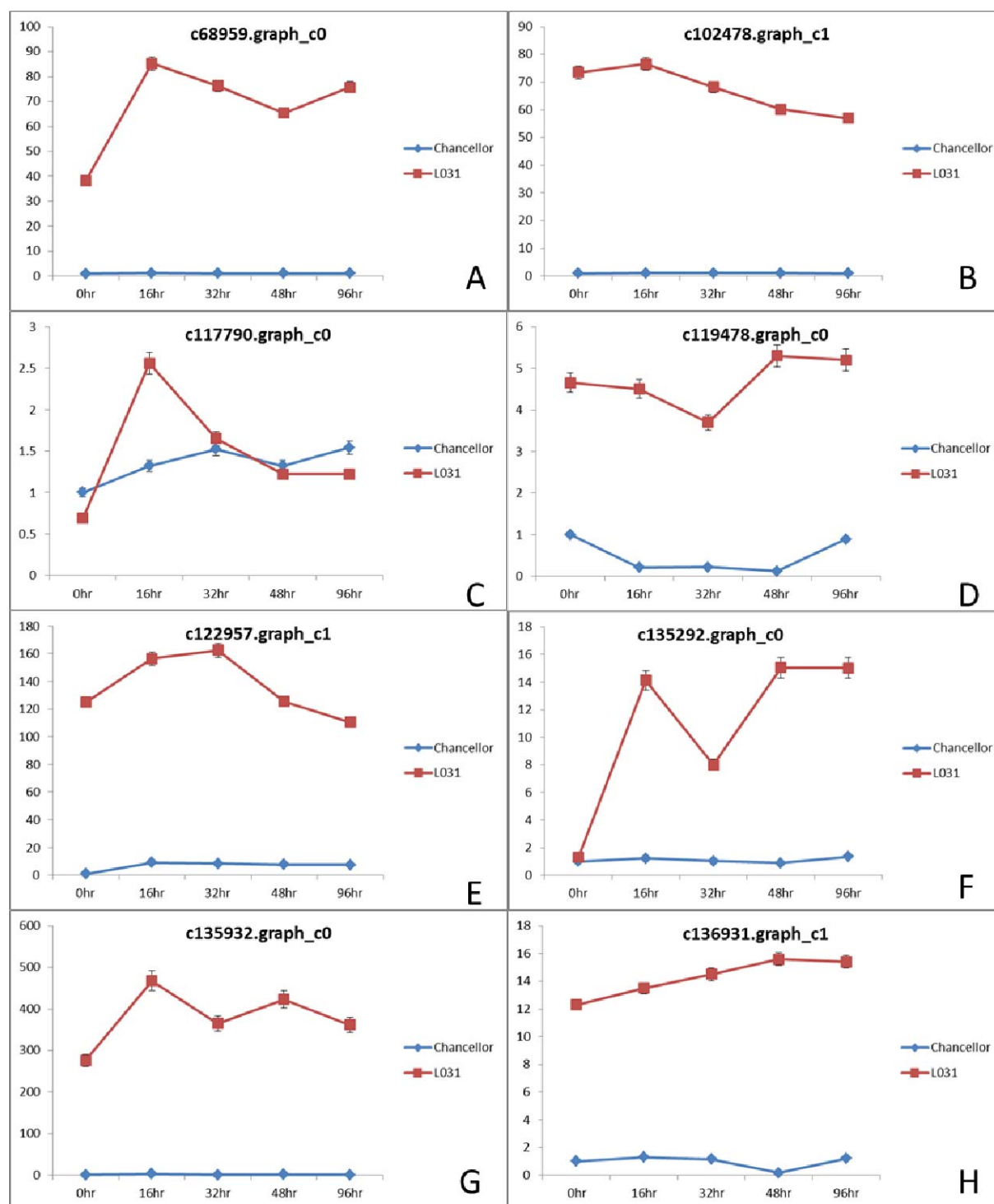


Figure 2. Validation of 8 DEGs in L031 and Chancellor using quantitative real time RT-PCR (qRT-PCR). The x-axis shows the samples at different hours after inoculation with *Bgt.* E09.

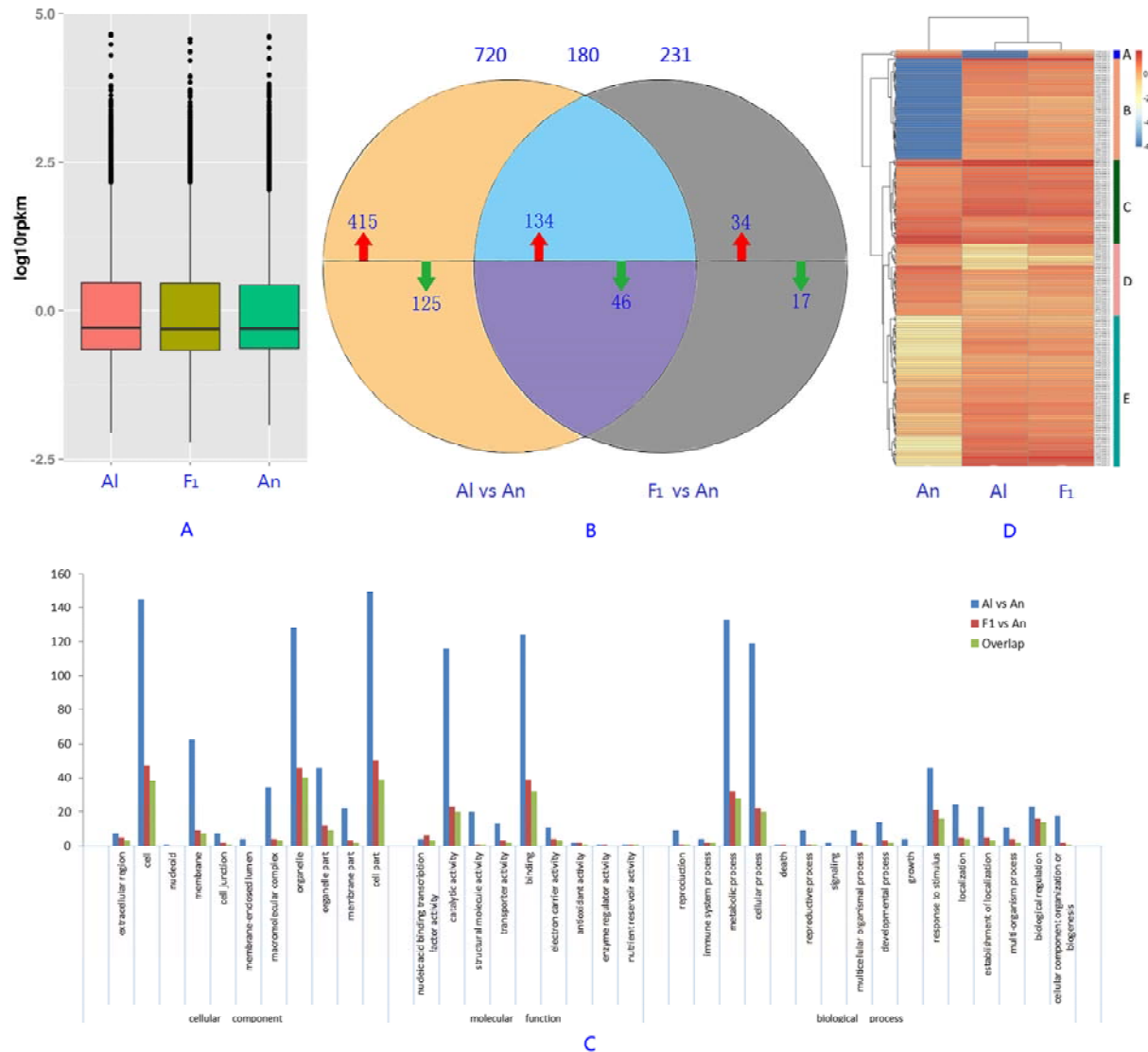


Fig. 3 RNA-seq analysis of young spikes of SN051-2 (AI), SN051-1 (An) and its F₁ hybrid.

A. Box plots of the relative log FPKM values based on all of the genes in each RNA-seq library.

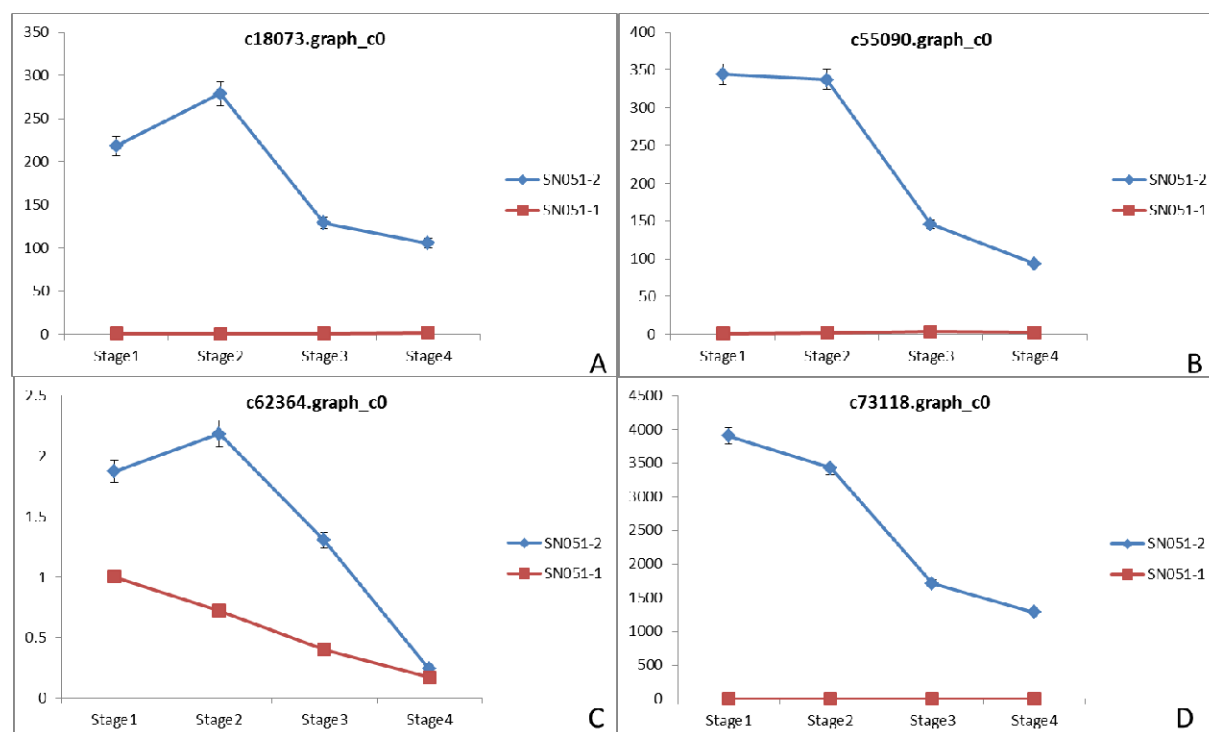
B. Venn diagram showing the number of differentially co-expressed genes among Chancellor (susceptible, *pm2*), L031 (resistant, *Pm2*) and its F₁ hybrid. (Red and blue arrows indicate up- and down-regulated DEGs, respectively).

C. GO classifications of DEGs. The results are also summarized into three main categories: 'biological processes, cellular components, and molecular functions'.

D. Heat map showing all of the DEGs among three samples. The gene expression data were

1 \log_2 -transformed and then quantile-normalized prior to generating the heat map for direct
2 comparison of the data. DEGs (red and blue indicate up- and down-regulated DEGs, respectively) for
3 each sample were mapped by lane. Hierarchical clustering and heat map generation were performed
4 in R.

5



6

7 **Figure 4. Validation of 4 DEGs in SN051-1 and SN051-2 using quantitative real time RT-PCR**
8 **(qRT-PCR).** The x-axis shows the samples at different develop stage.

9

10