Integrated Single-Molecule Long-Read Sequencing and Illumina sequencing reveals the resistance mechanism in response to barley yellow dwarf virus-GAV in Psathyrostachys huashanica

Chuan Shen
Northwest Agriculture and Forestry University
Caiyan Wei
Northwest Agriculture and Forestry University
Jingyuan Li
Northwest Agriculture and Forestry University
Xudong Zhang
Northwest Agriculture and Forestry University
Yunfeng Wu (wuyunfeng@nwafu.edu.cn)
Northwest Agriculture and Forestry University https://orcid.org/0000-0002-8059-0842

Research article

Keywords: P. huashanica, full-length transcriptome, Illumina sequencing, PacBio, BYDV-GAV

Posted Date: February 20th, 2020

DOI: https://doi.org/10.21203/rs.2.24039/v1

Version of Record: A version of this preprint was published at Phytopathology Research on July 30th, 2020. See the published version at https://doi.org/10.1186/s42483-020-00057-8.

Abstract

Background: Even though *Psathyrostachys huashanica* has great potential and promising prospects for resistance gene mining and molecular genetic breeding and a significant germplasm resource value, there is not yet a reference genome available. To date, most of the studies of *P. huashanica* have focused on the creation of translocation lines and addition lines as well as the development of molecular markers. Therefore, there is a great lack of research at the transcriptional level.

Results: The full-length transcriptome of *P. huashanica* was sequenced using PacBio isoform sequencing (Iso-Seq) of a pooled RNA sample derived from roots, stems, buds and leaves in equal amounts to explore potential full-length transcript isoforms. A total of 112,596 unique transcript isoforms were obtained, with a total length of 114,957,868 base pairs (bp). Illumina sequencing reads were used to correct and trim PacBio isoforms. In total, 103,875 unigenes were annotated with at least one functional database. In addition, a mass of DEGs involved in BYDV-GAV defence were identified; most of these resistance related genes showed gene ontology (GO), and KEGG enrichment analysis showed that these DEGs were mostly involved in the response to plant-pathogen interaction, plant hormone signal transduction and the MAPK signalling pathway. Then, twenty resistance-related upregulated genes, including MAPKs, CRPKs, CDPKs, PR proteins, WRKYs and disease resistance proteins were selected for RNA-seq data validation using quantitative real-time PCR, which showed consistencies between these two sets of data. Besides, the results also indicated that a series of defense-related genes were induced during BYDV-GAV infection at different stages.

Conclusions: The full-length transcriptome dataset should contribute to improved *P. huashanica* stress resistance gene utilisation, and will serve as a reference database for the analysis of transcript expression in *P. huashanica*. In addition, we identified genes in our transcriptome dataset involved in the regulation of the BYDV-GAV infection response in *P. huashanica.*

Background

Psathyrostachys huashanica Keng (2n = 2x = 14, NsNs) is a perennial cross-pollinating plant (subfamily Pooideae, tribe Triticeae, family Poaceae) [1, 2], which is a rare and endangered plant at the national level and a wild relative of crops in need of protection (Chinese National Forest Bureau and Agriculture Ministry 1999). It is distributed throughout central Asia, from Eastern Turkey to central China and Mongolia, and grows only on the residual soil of the mountainous braes and rocky slopes of Huashan Pass in the Qinling Mountains of Shaanxi Province, central China [3]. Psathyrostachys huashanica is already regarded as a potential omnipotent germplasm resource because of its excellent performance at early maturity, multiple florets and kernels and a high resistance to both abiotic (drought, barren soil, and salinity) and biotic stress (fungal disease: stripe rust, leaf rust, powdery mildew, take-all wheat fungus and virus disease and wheat yellow dwarf) [4–11].

The development of alien addition lines is important for transferring useful genes from exotic species into common wheat [12]. In 1991, scientists began to create hybrid materials between common wheat and P. huashanica, which is regarded as an effective and economic method for introducing agronomically desirable characters into available wheat cultivars [6]. Chromosome engineering techniques for the targeted introgression of resistance genes from wild wheat relatives to hexaploid wheat have been widely used for many years [13, 14]. By using embryo culture, backcrossing, and chromosome doubling, scientists crossed common wheat cv. 7182 and P. huashanica to generate heptaploid hybrid H8911 (2n = 49, AABBDDNs) [6]. In order to study its resistance to stripe rust, Giemsa C-banding, genomic in situ hybridization (GISH) and EST-SSR techniques were used to characterise a Triticum aestivum-P. huashanica Keng monosomic addition line PW11 with superior resistance to stripe rust. This subsequently proved that a small terminal segment of the 3Ns genome carried new stripe rust resistance genes that showed high levels of disease resistance [15]. In the same way, wheat x P. huashanica Keng 2Ns, 4Ns, 6Ns and 7Ns disomic addition lines were generated and showed enhanced tiller numbers and disease resistance [16–19]. In addition, a recent study showed that P. huashanica exhibited high resistance to barley yellow dwarf virus-GAV (BYDV-GAV) and disrupted the movement of the viral particles from inoculated leaves to new leaves [11].

Barley yellow dwarf viruses (BYDVs) are a class of single-stranded RNA viruses belonging to the genus Luteovirus, which are phloem-limited and obligately transmitted by aphids in a persistent or circulative manner [20]. Based on the specificity of their aphid vectors and serological properties of the virus particles, BYDVs have been classified into four distinct isolates, including BYDV-PAV, -GPV, -PAV, and -RMV, among of which, BYDV-GAV is the species most often found in China in recent years [21–23]. In addition, BYDVs can infect some cereal crops and grass species belonging to the family Gramineae (Poaceae), such as wheat (Triticum aestivum), barley (Hordeum vulgare), and oat (Avena sativa) [24, 25]. In the field, infection of BYDVs results in a number of wheat cultivars showing symptoms of leaf yellowing and plant dwarfism, but severe cases can cause yield loss [26].

Transcriptomics is the study of gene structure, expression and regulation. Short-read second-generation sequencing (SGS) has become a powerful tool for quantitative gene expression levels and exploring gene transcriptional regulation [27]. However, SGS has its limitations, for example short read lengths, which makes it poorly suited for the assembly of complex genomes and transcriptomes as well as full-length isoforms and methylation detection [28, 29]. Single-molecule real-time (SMRT) sequencing, a form of third-generation sequencing, was developed by Pacific BioSciences (PacBio) and offers an alternative approach to overcome these limitations; this enhances our understanding of the complexity of the transcriptome [28, 29].

Full-length cDNA sequencing is fundamental to studies of structural and functional genomics, and is beneficial for genome annotation, the identification of novel genes and isoforms as well as the characterization of long non-coding RNAs (lncRNAs), especially those without a reference genome [30, 31]. Besides, SMRT sequencing has been applied to the assembly of plant genomes based on its long reads, such as Triticum aestivum

[32], loblolly pine [33], Zea mays [34] and rubber tree [35]. On the other hand, full-length cDNA sequencing has been used to characterise the complexity of transcriptomes combined with Illumina-based RNA-seq datasets in Sorghum bicolor [36], Zea mays [37], cotton [38], sugarcane [39], coffee bean [40] and sweet potato [41]. Moreover, SMRT sequencing has also been widely used to construct plant transcriptomes without a reference genome, for example Camellia sinensis [42], Astragalus membranaceus [43] and Halogeton glomeratus [44], but it had not previously been used in P. huashanica.

In this study, we constructed full-length cDNA libraries from P. huashanica and performed SMRT sequencing to generate large-scale full-length transcripts. This is the first study to perform SMRT sequencing of the full-length transcriptome of P. huashanica. Then, we used full-length transcripts as a reference to profile the differential gene expression in response to BYDV-GAV infection, and found plenty of defence-related genes that had been induced by the viral invasion. The results of our study provide novel insights into the resistance response to BYDV-GAV infection in P. huashanica and could potentially contribute to the utilization of resistance resources of P. huashanica in the future.

Results

Construction of a long-read reference transcriptome for P. huashanica

In order to obtain desirable transcript isoforms, high-quality RNA was extracted from four tissues of P. huashanica to generate four RNA samples. These RNA samples were pooled with the same amount of each sample. Two size-fractionated, full-length cDNA libraries (1–5 kb and 4.5–10 kb) were constructed and subsequently sequenced using the Pacific Bioscience Sequel platform with three cells (2 and 1 cell, respectively). A total of 24.15 Gb raw polymerase reads with an average of 8.05 Gb per cell and 20.77 Gb subreads with an average of 6.9 Gb per cell were collected from SMRT sequencing (Additional file : Table S1 and Figure S1). After quality control, 3,677,921,244 bp CCS (reads of insert, 1,688,359 reads) were obtained, including 1,140,528 in the 1–5 kb library and 547,831 in the 5–10 kb library (Table. 1). The ROI length distributions of each library are shown in the Additional File: Figure S2). Of these reads of insert, 507,104 (~ 30.04%) were classified as full-length and non-chimeric, and 902,746 (~ 53.47%) were classified as non-full-length reads based on whether 5’ and 3’ ends and poly A tails were detected. Only reads with two ends and a poly A tail were classified as full-length (Additional file: Figure S3). Only full-length non-chimeric reads and non-full-length reads were used in further analysis, while short reads with a length of < 300 bp and chimeric reads were discarded from the subsequent analysis.

Functional annotation of coding genes

Full-length non-chimeric reads were clustered into consensus groups. For each cluster, if there was sufficient FL and non-FL coverage, then Quiver was run to polish the consensus. Only high QV consensus sequences were used for downstream analysis. High quality consensus isoforms of each library were merged into the final results, and redundancy was removed. After clustering and polishing these three libraries, we obtained 112,596 isoforms from 177,882 isoforms (Table 2), and the length distribution of final consensus isoforms is shown in (Fig. 1).To improve PacBio transcript quality, RNA-Seq reads were used to trim all final consensus isoforms, and mapping identity, substitution, deletion, insertion and mapping coverage were obtained (Additional file: Table S2). Then, we used BLAST, BLAST2GO and InterProScan5 to perform functional annotation terms from the non-redundant (NR), NCBI nucleotide sequence (NT), Gene Ontology (GO), Cluster of Eukaryotic Orthologous Groups (KOG), Kyoto Encyclopaedia of Genes and Genomes (KEGG), SwissProt, and Interpro protein databases with transcripts (Additional file: Table S3). As a result, 103,875 (92.25%) unigenes were annotated with at least one functional database (Table 3). Also, with KOG, GO and KEGG annotation, functional distribution was statistically assessed and the results are shown in (Fig. 2). In addition, we obtained 36,283 annotated transcripts among NR, KEGG, Swissprot and Interpro (Fig. 3).

Table 1
Summary of reads of insert classification, the size fraction of F01 and G01 libraries were 1–5 kb, and the size fraction of H01 library was 4.5–10 kb

Sample	Library	reads of insert	five prime reads	three prime reads	poly-A reads	full-length non-chimeric reads	full-length non-chimeric read length(bp)
Psathyrostachys_huashanica	F01	565467	301,093(53.25%)	316,921(56.05%)	292,646(51.75%)	194,333(34.37%)	451.5
Psathyrostachys_huashanica	G01	575061	345,129(60.02%)	360,354(62.66%)	333,960(58.07%)	232,059(40.35%)	457.5
Psathyrostachys_huashanica	H01	547831	267,902(48.9%)	269,092(49.12%)	139,934(25.54%)	80,712(14.73%)	2723

Table 2
Summary of the number of isoforms, high-quality sequences from sLibrary were clusted and corrected to create Final isoforms

Library	sLibrary_isoforms	Final_isoforms
Psathyrostachys_huashanica_F01	70137	49273
Psathyrostachys_huashanica_H01	21969	3158
Psathyrostachys_huashanica_G01	85776	60165
TOTAL	177882	112596

Table 3
Summary of functional annotation result, final isoforms were used to perform functional annotation with seven databases (Nr, Nt, Swissprot, KEGG, KOG, Interpro, GO)

Values	Total	Nr	Nt	Swissprot	KEGG	KOG	Interpro	GO	Intersection	Overall
Number	112,596	92,515	95,790	65,275	67,659	67,051	54,331	45,466	20,919	103,875
Percentage	100%	82.17%	85.07%	57.97%	60.09%	59.55%	48.25%	40.38%	18.58%	92.25%
1. Polymerase reads length distribution										

Illumina sequencing and assessment of sequencing quality

To determine the appropriate time points for RNA-seq analysis in P. huashanica, time course virus titers were quantified using qRT-PCR at different time points after BYDV-GAV inoculation. As shown in (Fig. 4), BYDV-GAV titer increased dramatically at 3 dpi and then peaked at 10 dpi and then declined gradually until 14 dpi when the virus was barely detectable. Then, mock-infected and BYDV-GAV-infected leaves (3 dpi, 7 dpi and 14 dpi) were selected to conduct RNA-seq analysis.

Since the genome sequence of P. huashanica was not available, we used our PacBio sequencing dataset as a reference to perform RNA-SEq. High quality RNA was extracted from the four samples; each had three repetitions to generate 12 libraries. The libraries were sequenced on an Illumina HiSeq2000 platform. More than 16 GB raw reads and 115 million clean reads were generated from each library using Illumina paired-end sequencing (Additional file: Table S4). Thereafter, clean reads were aligned to the PacBio dataset and a total of 76,313 transcripts were detected, including 16,827 lncRNA transcripts and 47,196 mRNA transcripts which were used for FPKM analysis. The assessment of transcript coding quality was determined by three types of prediction software, and the results are shown in (Fig. 5).

Identification of differentially expressed genes (DEGs)

Differentially expressed genes (DEGs) were obtained by comparison of BYDV-GAV-inoculated samples with mock-inoculated samples at each time point. DEGs were identified with the criteria of $|\text{fold change}| \geq 2$ and q-value ≤ 0.001 at particular time points (Fig. 6). In this way, 11,282 genes were identified to be DEGs at 3 dpi, of which 5991 genes were upregulated and 5291 were downregulated. Interestingly, there were 4689 genes downregulated, and there were more than 3864 upregulated genes at 7 dpi. At the later time points, the upregulated genes increased to 4729, which again exceeded the 4528 downregulated genes. In addition, we used volcano plots to visualize DEGs in a more intuitive way (Fig. 7). The results of the distribution of upregulated and downregulated genes were consistent with the accumulation of virus in P. huashanica. All statistically significant DEGs were further characterized by functional annotation and enrichment analysis.

Short time series expression miner (STEM) analysis

To explore the major expression trends of DEGs from the sequencing dataset over the four time points, we performed a clustering and visualization analysis using STEM software and found that all DEGs were assigned to 46 clusters that displayed similar expression patterns. We found that 15 clusters were statistically significant (P-value ≤ 0.05) (Fig. 8a). Among them, clusters 45, 46, 47, 48 and 49 exhibited upregulated expression patterns during viral infection, while the rest of the clusters showed downregulated expression patterns. Clusters 31, 33, 34, 35, 36, 43, 44 and 45 showed initial increased expression levels at 3 dpi but were downregulated at a later infection stage. Clusters 29, 30, 38, 39, 40, 41, 42, 47, 48 and 49 exhibited gradually increasing expression levels from 3 dpi to 7dpi, but showed downregulated expression levels in the following period. Those upregulated clusters were induced by the BYDV-GAV infection to resist viral replication and movement.

In addition, we also used the R 'Mfuzz' package to cluster the expression levels of all detective genes, and achieved 12 clusters (Fig. 8b), among which, clusters 9 (7.9%) and 10 (5.2%) exhibited upregulation during all stages of viral infection. On the contrary, clusters 6 (9.1%) and 11 (5.6%) showed downregulation during all stages of viral infection. Moreover, clusters 5 (5.3%), 7 (6.2%), 8 (8.0%) and 12 (11.8%) also played critical roles in the response to the viral invasion.

GO and KEGG enrichment analysis of DEGs

To investigate the biological function of viral-induced genes in P. huashanica, GO and KEGG pathway analysis were performed on DEGs at different time points. The GO enrichment results for DEGs between 3 dpi and mock-infected indicated that the cellular process and metabolic process were the most significantly represented groups in the biological process category, while cells and cell parts were the most common enrichment terms in the cellular component. Within the molecular functional category, catalytic activity and binding were the predominant enrichment groups (Fig. 9a). The above results were similar to what had been observed for other viral infection time points relative to mock (Fig. 9b and Fig. 9c). However, certain differences also occurred between these time points and mock. For example, the number of upregulated genes enriched in metabolic processes and catalytic activity were more downregulated at 3dpi, while upregulated enriched genes decreased at 7 dpi and increased at 14 dpi, which corresponded to the amount of virus.

According to the annotation from the KEGG database, the most enriched pathway was the metabolic pathway, followed by biosynthesis of secondary metabolites at all infection stages. Additionally, we found that DEGs related with plant-pathogen interaction, phenylpropanoid biosynthesis, starch and sucrose metabolism, the MAPK signalling pathway and plant hormone signal transduction were significantly enriched at 3 dpi (Fig. 10a). The pathways of plant-pathogen interaction, mRNA surveillance, phenylpropanoid biosynthesis, starch and sucrose metabolism and glycolysis/gluconeogenesis were predominant at 7 dpi (Fig. 10b). Among main pathway categories, RNA transport, plant-pathogen interaction, plant hormone signal transduction, phenylpropanoid biosynthesis, mRNA surveillance pathway and the MAPK signalling pathway were significantly represented at 14 dpi (Fig. 10c). In summary, most of the DEGs were putatively involved in resistance-related pathways. These genes appeared to participate in the interactions between P. huashanica and BYDV-GAV.

Identification of transcription factors and transcriptional regulators

We identified 605 TFs distributed in 46 families among the 15,499 DEGs. Then, these TFs were classified, and mainly include the following families: AP2-EREBP (apetala2-ethylene-responsive element binding proteins) (25 genes), WRKY (41 genes), bHLH (basic helix-loop-helix) (33 genes), MYB-related (40 genes), MYB (13 genes), NAC (NAM, ATAF1-2, and CUC2) (13 genes), C2H2 (13 genes), NF-YA (13 genes), GARP-G2-like (17 genes) and bZIP (basic region/leucine zipper motif) (22 genes) (Additional file: Table S6). Transcriptional regulators mainly focused on AUX/IAA (35 genes), SET (19 genes), TRAF (Tumour necrosis factor receptor-associated factor) (18 genes), SNF2 (17 genes), GNAT (Gcn5-related N-acetyltransferases) (9 genes), Jumonji (9 genes), HMG (high-mobility-group) (8 genes) and PHD (8 genes) (Additional file: Table S5). The results indicated TFs widely participate in fighting BYDV-GAV infection.

Validation of gene expression profiles using qRT-PCR

The expression trends of 20 representative genes were verified by qRT-PCR in a separate experiment. The genes selected that were related to the defence response were divided into three categories. The first class included protein kinases, such as cysteine-rich receptor-like protein kinase (CRPK), mitogen-activated protein kinase (MAPK), calcium-dependent protein kinase related genes (CDPK) and wall-associated receptor kinase (WARK). The second class included genes involved in defence-related proteins, such as pathogenesis-related protein (PR), heat shock protein 70 (HSP70) and WRKY transcription factor (WRKY). The last class selected included resistance genes, such as those coding for disease resistance proteins. The results of all 20 genes selected using RT-qPCR were consistent with the results obtained from RNA-Seq analysis (Fig. 11).

Discussion

Although many studies related to P. huashanica have been performed, and numerous important germplasm resources and resistant materials have been obtained, there was still no reference genome available for of P. huashanica, with little progresses achieved in functional genomics. Therefore, it was necessary to urgently produce a reference genome or transcripts of P. huashanica for further study.

Iso-Seq sequencing was developed and applied to the characterisation of transcriptomes of different species including plants and animals. It has a great many advantages, for example, creating long reads, the lack of a need for assembly and so forth, making it especially suitable for non-model organisms that lack genomic sequences [30, 31].

The long-standing and widespread use of P. huashanica in current resistance gene mining and molecular genetic breeding, along with its significant value as a germplasm resource, has led to intense interest in uncovering the mechanism of resistance to numerous pathogens [5, 11]. Because of the lack of a reference genome and reference transcripts, much of this interest has focused on the creation of translocation and addition lines. To tackle this problem, we applied long-read SMRT sequencing of four distinct tissues (i.e. the root, stem, leaf and bud), from which we were able to generate a meritorious complete transcriptome. Finally, we generated an initial collection of 112,596 unique P. huashanica transcript isoforms in our experiment. This transcriptome dataset offers a broad range of possibilities to study the multifaceted characteristics of P. huashanica.

Using the PacBio full-length isoform sequencing to obtain transcripts without assembly overcomes the difficulty of short-reads posed by next generation sequencing [45]. Even though PacBio offers longer reads than other current platforms, it has a higher error rate [46]. The short-reads from the Illumina platform have been widely used for RNA-Seq differential gene expression analysis, since it provides sufficient depth and a lower error rate compared to reads generated from PacBio [27]. Therefore, we used clean reads created from short-read next-generation sequencing (NGS) technology to correct the PacBio transcript isoforms [47].

P. huashanica serves as a resistant germplasm resource library contains a plethora of resistance resources; it not only confers resistance to biotic stress, but also confers resistance to abiotic stress. By sequencing, we found that P. huashanica contains a huge transcription factor and resistance

gene library. In previous research, using wild relative species to create high-quality and durable resistant materials has been proved an extraordinary way to recover genetic diversity of wheat germplasm [48–50]. Starting twenty years ago, research groups introgressed BYDV (Barley Yellow Dwarf Virus) resistance genes, which were exploited in more than ten wild relative species belonging to Thinopyrum, Agropyron, Elymus, Leymus, Roegneria and Psathyrostachy genera into wheat [51].

In this study, Illumina sequencing was performed to profile DEGs between BYDV-GAV-infected and mock-infected P. huashanica at different time points. A total of 11,282, 8,553 and 9,257 genes were identified as DEGs at 3 dpi, 7 dpi and 14 dpi, respectively, most of which were enriched in plant-pathogen interaction, plant hormone signal transduction and MAPK signalling pathways. In the long-term interaction between plants and pathogens in nature, plants have evolved a complicated immune system to recognize pathogens and activate defence responses. Pattern recognition receptors (PRRs) localized on the plasma membrane belong to the receptor-like kinase (RLK) or receptor-like protein (RLP) family, which combine conserved molecular features derived from pathogens to trigger activation of PAMP-triggered immunity (PTI) [52, 53]. In addition to PTI, plants also recognize phytopathogen-synthesized effector proteins known as resistance (R) proteins to promote their infection of the host plants [54, 55]. In our experiment, we found a number of MAPKs (mitogen-activated protein kinases), CRPKs (cysteine-rich receptor-like protein kinases), CDPKs (calcium-dependent protein kinases), WAKs (wall-associated kinase-like proteins) and serine/threonine-protein kinases were significantly upregulated during viral infection, which might be associated with the innate immunity of the plant [56–59]. Additionally, plant proteins belonging to the nucleotide-binding site-leucine-rich repeat (NBS-LRR) family, one of the largest gene families known in plants, were clearly correlated with resistance responses and pathogen detection, especially the effector molecules of pathogens responsible for virulence [60–62]. Our sequencing results identified many NLR resistance proteins that were induced to high expression to defend against viral invasion at different time points. The results were suggestive of a defence response involving both ETI and PTI.

Transcription factors (TFs) play a central role in the response to multifaceted biotic stresses, such as insect attack and pathogen infection [63]. In order to respond to such stresses, many TF families have been reported to be differentially expressed in numerous plants as a positive response to bacterial, fungal and viral infections [64]. Several families of TFs, such as MYB, WRKY, ERF, NAC and bZIP, which are known key regulators in the defence response, were highly expressed in P. huashanica after viral infection [65–69].

In summary, the full-length transcript sequencing of P. huashanica, which is one of the wild relatives of wheat, provided an immense opportunity to identify high-quality germplasm resources for crop breeding. In addition, a diversity of resistance genes induced by BYDV-GAV was obtained with Illumina sequencing, which could contribute to the antiviral breeding in wild relative of wheat.

Conclusion

This study generated 1,688,359 reads of insert from three SMRT cells and obtained 112,596 isoforms with an N50 of 1,204 bp. Thereafter, we combined strand-specific Illumina RNA-seq and PacBio full-length cDNA sequences to achieve a better and lower error rate in the reference transcriptome database. According to the functional annotation results of isoforms, 82.17% isoforms had annotations in the NR database. Approximately 40.38% and 60.09% of the total isoforms were annotated against the gene ontology and KEGG pathway databases, respectively. The Venn diagram of protein alignment results between NR, KOG, KEGG, Swissprot and Interpro indicated that there were 36,283 isoforms annotated in all of these databases. We used all transcripts to predict CDS, and achieved a total number of 91,210 sequences. Furthermore, Illumina RNA-seq detected 76,313 transcripts from the PacBio dataset, including 47,196 mRNAs and 16,827 LncRNAs. The identification of DEGs involved in viral resistance will extend our understanding of the complex molecular mechanisms of the interactions between BYDV-GAV and P. huashanica.

In summary, the total set of transcripts in the study could be used to improve P. huashanica stress resistance gene utilization and provide a foundation for future studies on molecular breeding for resistance to wheat disease.

Materials And Methods

Sample preparation and RNA extraction

P. huashanica was provided by Professor Jinxue Jing from North West Agriculture and Forestry University, China. To obtain a good representation of the P. huashanica transcriptome, all four tissues, including roots, stems, buds and leaves, including inoculated and uninoculated leaves, were harvested from different development stages of five independent plants. Total RNA was extracted from each sample using the protocol described in [70], employing a Trizol kit (Invitrogen), followed by a Qiagen RNeasy Plant minikit (Qiagen) and subsequently pooled in equal amounts.

For illumina RNA-Seq, P. huashanica were inoculated with viruliferous aphids carrying BYDV-GAV, and control plants (mock-infected) were inoculated with non-viruliferous aphids, 5 aphids per leaf. The viruliferous and non-viruliferous aphids were raised on susceptible wheat line 7182. For the experimental groups, at the three-to-four leaf stage, the samples of RNA-Seq were collected at day 0 (mock-infected), day 3 (dpi), day 7 (dpi) and day 14 (dpi). Then, all these samples were immediately frozen in liquid nitrogen for more than ten minutes and stored at -80 °C for further use. The concentration of each RNA sample was measured using a QubitFluorometer (Life Technologies, USA), optical density was checked using the NanoDrop (Thermo Fisher Scientific, MA, USA), and RNA integrity was assessed with an Agilent 2100 bioanalyzer (Agilent, USA). Only those with an RIN value of more than eight were retained for further use.

Library preparation and PacBio sequencing

The preparation of cDNA was strictly according to the PacBio Iso-seq protocol. Total RNA was used to prepare the first- and second-strand cDNA using the SMARTer PCR cDNA Synthesis Kit (Clontech) and Phusion High-Fidelity DNA Polymerase (NEB). Then, PCR amplified fragments of dsDNA were

used to construct and sequence SMRTbell libraries. The cDNA fractions were generated with the BluePippin size selection system (Sage Science). We constructed two libraries, 1–5 kb and 4.5–10 kb, of which the former fragment library ran two cells and the latter ran only one. The size selection, PCR amplification and sequencing of the three SMRT cells were conducted on the PacBio Sequel platform by BGI (BGI-Shenzhen, China).

Data processing and read correction

Raw data from libraries produced by the Pacific Biosciences Sequel were processed following the SMRT analysis pipeline. Consistent sequences were obtained after removing redundant subreads, which were used to identify insert fragments; the sequences achieved in this step were called reads of insert (ROI). The ROI sequences were divided into four categories: full-length non-chimeric, chimeric, non-full-length and short reads. Only those sequences that had 5' adapters, 3' adapters and a poly-A tails could be defined as full-length reads [71]. Therefore, full-length non-chimeric sequences and non-full-length sequences were retained for further analysis.

These classified ROI were used to predict de novo consensus isoforms by performing an Iterative Clustering for Error Correction (ICE) algorithm. Then, predicted consensus isoforms were polished using the Quiver algorithm to finally obtain the full-length polished consensus sequences. Depending on the Quiver output (QV), which indicated how confident the consensus calls were, the algorithm classified the polished isoforms into high QV (expected accuracy ≥ 99%) or low QV isoforms; only high QV isoforms were used for further analysis.

Because there was no available reference genome for P. huashanica, we used CD-HIT-EST version 4.6.5 (-c 0.98 -T 6 -G0 -aL 0.90 -AL 100 -aS 0.98 -AS 30) to remove the redundancy of multiple libraries based on sequence similarity to obtain final full-length isoforms [72].

Illumina RNA-Seq and PacBio data correction

The total RNA was digested with DNase I, and Ribo-Zero™ rRNA Removal Kit (Illumina, USA) was used to remove the rRNA. Purified mRNA from the previous steps was fragmented into small pieces. The first strand cDNAs were synthesized with random hexamer primers and reverse transcriptase, followed by second strand cDNAs synthesis using DNA polymerase I (New England BioLabs) and RNase H (Invitrogen). Then, A-Tailing Mix and RNA Index Adapters were added by incubating to carry out end repair. The cDNA fragments with adapters were amplified by PCR, and the products were purified by Ampure XP Beads. The purity and quality of the libraries were measured using an Agilent 2100 Bioanalyzer and Qubit 2.0. The PCR amplification and sequencing of the twelve libraries were conducted on the Illumina HiSeq platform by BGI (BGI-Shenzhen, China). The raw Illumina sequencing reads were cleaned by removing adapter sequences, reads with over 5% unknown reads, and reads with low quality scores ($Q \leq 15$).

Subsequently, the paired-end (PE) reads generated from Illumina Hiseq X-ten platform was used to correct PacBio isoforms. Then, clean reads from next-generation sequencing were delivered to proofread version 2.14.0 with the default parameters to correct single-bases and trim the regions of chimeras and low quality [73].

Identification of coding RNAs

We used three analysis tools: CPC [74], txCDsPredict [75] and CNCI [76], as well as a protein database, pfam [77], to distinguish between coding and non-coding sequences based on their coding potentials. Three analysis tools marked the code capacity of isoforms and set a threshold value to distinguish lncRNAs from mRNAs. Isoforms that could align to the Pfam database were identified as mRNA; others were considered lncRNAs. We confirmed that the transcript was mRNA or lncRNA on the basis of at least three of the four judgment results being consistent, long non-coding RNAs having a nucleotide sequence of at least 300 bp and an ORF of no more than 100 amino acids.

Functional annotation

Final full-length isoforms were searched against the NCBI non-redundant (NR), NCBI nucleotide sequence (NT), SwissProt, InterPro, Cluster of EuKaryotic Orthologous Groups (KOG), GO and KEGG databases with a threshold E-value $\leq 10^{-5}$. The annotation of NT, NR, KOG, KEGG and SwissPro was performed using BLAST version 2.2.23 [78], GO using Blast2Go version 2.5.0 [79] and InterPro using InterProScan5 version 5.11-51.0 [80]. GO terms were enriched and plotted by WEGO [81]. KEGG pathway mapping was performed on the KEGG Automatic Annotation Server (KAAS) v2.0 [82]. Venn diagrams of NR, KEGG, SwissPro and InterPro annotation results were drawn by the online Venn tool [83].

Time series and transcription factors expression analyses

Time series-cluster analyses, based on the Short Time-series Expression Miner (STEM) method, were conducted with Short Time-series Expression Miner (STEM) software [84]. This was used to identify the global trends and similar temporal model patterns of the expression of the total DEGs. Meanwhile, the R package "Mfuzz" (version = 3.8) was used to draw the cluster patterns of all expressed genes [85]. The prediction and classification of plant transcription factors and transcriptional regulators from DEGs used a web-based tool iTAK, Plant Transcription Factor & Protein Kinase Identifier and Classifier (http://itak.feilab.net/cgi-bin/itak/index.cgi) [86].

qRT-PCR validations for differentially expressed genes

To validate the RNA-Seq analysis, qRT-PCR was performed with the same plant materials and sampling time. Total RNAs extracted using a Plant Total RNA Extraction Kit (Bioer Technology, Hangzhou) from three inoculated replications were used to examine the expression patterns of selected genes. Total RNA (1 μg) was used for reverse transcription with M-MLV Reverse Transcriptase (Promega, USA), and the cDNA samples were diluted two-fold. For qRT-PCR, quantitative assays were conducted on each cDNA dilution with UltraSYBR Mixture (Cwbio, China) and a Roche LightCycler480 Real-time PCR Detection System instrument (Roche, Switzerland).

Twenty genes that were found to be differentially expressed were selected based on the RNA-seq results (i.e. >2-fold change, $p < 0.05$) and analysed using real-time quantitative PCR. The gene-specific primer pairs for each gene were designed using Primer3web version 4.0.0 (http://primer3plus.com/primer3web/primer3web_input.htm). P. huashanica housekeeping gene 18S rRNA was used as the control to normalize all data (for a list of these genes primer sequences, see Additional file 3: Table S6). The data of the relative expression level of DEGs were normalized using a $2^{-\triangle\triangle Ct}$ method [87]. All qRT-PCR reactions were repeated in three biological and three technical replications, and the means and corresponding standard errors were calculated.

Abbreviations

GISH:genomic in situ hybridization; SGS:second-generation sequencing; SMRT:Single-molecule real-time; NR:non-redundant; NT:NCBI nucleotide sequence; GO:Gene Ontology; KOG:Cluster of Eukaryotic Orthologous Groups; KEGG:Kyoto Encyclopaedia of Genes and Genomes; DEGs:Differentially expressed genes; CRPK:cysteine-rich receptor-like protein kinase; MAPK:mitogen-activated protein kinase; CDPK:calcium-dependent protein kinase related genes; WARK:wall-associated receptor kinase; PR:pathogenesis-related protein

Declarations

Availability of data and materials

The raw PacBio reads of inserts are available with the NCBI SRA database accession numbers SRR7785349-SRR7785350. The raw reads of Illumina sequencing are available with the NCBI SRA database accession numbers SRR8441810- SRR8441814 and SRR7820544-SRR7820546.

Acknowledgments

Not applicable

Funding

This research was supported by the National key research and development program (2018YFD0200402-2) and the 111 project from the Education Ministry of China (No.B07049).

Authors' contributions

YW and CS conceived and designed the experiments. CW, JL and CS collected the samples. JL, XZ and CW conducted analysis. CS prepared the first draft. CW, JL and YW revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Baden C. A taxonomic revision of *Psathyrostachys* (Poaceae). Nordic Journal of Botany. 2010;11(1):3-26.
2. Kuo PC. Flora Reipublicae Popularis Sinicae. Science Press, Beijing. 1987;Pp.51–104.
3. Lu BR. Diversity and conservation of *Triticum* genetic resource. Chinese Biodiversity. 1995;3: 63–8.
4. Wang L, Liu Y, Du W, Jing F, Wang Z, Wu J, Chen X. Anatomy and Cytogenetic Identification of a Wheat-*Psathyrostachys huashanica* Keng Line with Early Maturation. PLoS One. 2015;10(10):e0131841.
5. Kang HY, Zhang ZJ, Xu LL, Qi WL, Tang Y, Wang H, Zhu W, Li DY, Zeng J, Wang Y et al. Characterization of wheat-*Psathyrostachys huashanica* small segment translocation line with enhanced kernels per spike and stripe rust resistance. Genome. 2016;59(4):221-9.
6. Chen SY, Zhang AJ, Fu J. The Hybridization Between *Triticum aestivum* and *Psathyrostachys huashanica*. Journal of Genetics & Genomics. 1991;18(6):508-12.
7. Du Y N, Chen X H, Zhao J X, et al. Analysis of drought resistance of *Triticum aestivum -Psathyrostachys* derivatives at the germination and seedling stages. Journal of Triticeae Crops. 2010;30(4):670-5.
8. He X, Chen Z, Wang J, Li W, Zhao J, Wu J, Wang Z, Chen X. A sucrose:fructan-6- fructosyltransferase (6-SFT) gene from *Psathyrostachys huashanica* confers abiotic stress tolerance in tobacco. Gene. 2015;570(2):239-47.

9. Wang Y, Xie Q, Yu K, Poysa V, Lin L, Kang H, Fan X, Sha L, Zhang H, Zhou Y. Development and characterization of wheat-*Psathyrostachys huashanica* partial amphiploids for resistance to stripe rust. Biotechnol Lett. 2011;33(6):1233-8.
10. Wang M, Shang H. Evaluation of resistance in Psathyrostachys huashaica to wheat take-all fungus. Journal of Northwest Agricultural University. 2000;28:69-71.
11. Song S, Tao Y, Zhang H, Wu Y. *Psathyrostachys huashanica*, a potential resource for resistance to Barley yellow dwarf virus-GAV. European Journal of Plant Pathology. 2013;137(2):217-21.
12. Ali N, Heslop-Harrison JP, Ahmad H, Graybosch RA, Hein GL, Schwarzacher T. Introgression of chromosome segments from multiple alien species in wheat breeding lines with wheat streak mosaic virus resistance. Heredity (Edinb). 2016;117(2):114-23.
13. Zhang P, Dundas IS, Xu SS, Friebe B, McIntosh RA, Raupp WJ. Chromosome Engineering Techniques for Targeted Introgression of Rust Resistance from Wild Wheat Relatives. In: Wheat Rust Diseases. 2017;163-72.
14. Wang Y, Yu K, Xie Q, et al. Cytogenetic, genomic in situ hybridization (GISH) and agronomic characterization of alien addition lines derived from wheat-*Psathyrostachys huashanica*. African Journal of Biotechnology. 2013;10(12):2201-11.
15. Du W, Wang J, Pang Y, Wang L, Wu J, Zhao J, Yang Q, Chen X, Belzile F. Isolation and characterization of a wheat – *Psathyrostachys huashanica* 'Keng' 3Ns disomic addition line with resistance to stripe rust. Genome. 2014;57(1):37-44.
16. Du W, Wang J, Wang L, Wu J, Zhao J, Liu S, Yang Q, Chen X. Molecular characterization of a wheat– *Psathyrostachys huashanica* Keng 2Ns disomic addition line with resistance to stripe rust. Molecular Genetics and Genomics. 2014;289(5):735-43.
17. Du W, Wang J, Lu M, Sun S, Chen X, Zhao J, Yang Q, Wu J. Characterization of a wheat-*Psathyrostachys huashanica* Keng 4Ns disomic addition line for enhanced tiller numbers and stripe rust resistance. Planta. 2014;239(1):97-105.
18. Du W, Wang J, Pang Y, Li Y, Chen X, Zhao J, Yang Q, Wu J. Isolation and characterization of a *Psathyrostachys huashanica* Keng 6Ns chromosome addition in common wheat. PLoS One. 2013;8(1):e53921.
19. Du W, Wang J, Wang L, Zhang J, Chen X, Zhao J, Yang Q, Wu J. Development and characterization of a *Psathyrostachys huashanica* Keng 7Ns chromosome addition line with leaf rust resistance. PLoS One. 2013;8(8):e70879.
20. Miller, W.A.; Rasochova, L. Barley yellow dwarf virues. Annu. Rev. Phytopathol. 1997, 35, 167-190.
21. King, A.M.Q.; Adams, M.J.; Carstens, E.B.; Lefkowitz, E.J. "Family Luteoviridae" in: Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses; Elsevier Academic Press: London, UK, 2012; pp. 1045-1053.
22. Zhou, G.; Zhang, S. Identification and application of four strains of wheat yellow dwarf virus. Sci. Agric. Sin. 1987, 298, 434-440.
23. Liu, F.; Wang, X.; Liu, Y.; Xie, J.; Gray, S.M.; Zhou, G.; Gao, B. A Chinese isolate of barley yellow dwarf virus-PAV represents a third distinct species within the PAV serotype. Arch. Virol. 2007, 152, 1365-1373.
24. Jin, Z.; Wang, X.; Chang, S.; Zhou, G. The complete nucleotide sequence and its organization of the genome of Barley yellow dwarf virus-GAV. Sci. China Ser. C Life Sci. 2004, 47, 175-182.
25. Zhang, Z.Y.; Lin, Z.S.; Xin, Z.Y. Research progress in BYDV resistance genes derived from wheat and its wild relatives. J. Genet. Genom. 2009, 36, 567-573.
26. Banks, P.M.; Davidson, J.L.; Bariana, H.; Larkin, P.J. Effects of barley yellow dwarf virus on the yield of winter wheat. Aust. J. Agric. Res. 1995, 46, 935-946.
27. Buermans H P, den Dunnen J T. Next generation sequencing technology: Advances and applications. BBA - Molecular Basis of Disease. 2014;1842(10):1932-41.
28. Rhoads A, Au K F. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics. 2015;13(5):278-89.
29. An D, Cao H X, Li C, et al. Isoform Sequencing and State-of-Art Applications for Unravelling Complexity of Plant Transcriptomes. Genes. 2018;9(1):43.
30. Roberts R J, Carneiro M O, Schatz M C. The advantages of SMRT sequencing. Genome Biology. 2013;14(6):405.
31. Liu Xiaoxian,Mei Wenbin,Soltis Pamela S et al. Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. Mol Ecol Resour. 2017;17(6): 1243-56.
32. Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL. The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. Gigascience. 2017;6(11):1-7.
33. Zimin AV, Stevens KA, Crepeau MW, Puiu D, Wegrzyn JL, Yorke JA, Langley CH, Neale DB, Salzberg SL. An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. Gigascience. 2017;6(1):1-4.
34. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS et al. Improved maize reference genome with single-molecule technologies. Nature. 2017;546(7659):524-7.
35. Pootakham W, Sonthirod C, Naktang C, Ruang-Areerate P, Yoocha T, Sangsrakru D, Theerawattanasuk K, Rattanawong R, Lekawipat N, Tangphatsornruang S. De novo hybrid assembly of the rubber tree genome reveals evidence of paleotetraploidy in *Hevea* Sci Rep. 2017;7:41457.
36. Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS. A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun. 2016;7:11706.

37. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nat Commun. 2016;7:11708.
38. Wang M, Wang P, Liang F, Ye Z, Li J, Shen C, Pei L, Wang F, Hu J, Tu L et al. A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. New Phytol. 2018;217(1):163-78.
39. Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJ. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. BMC Genomics. 2017;18(1):395.
40. Cheng B, Furtado A, Henry RJ. Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. Gigascience. 2017;6(11):1-13.
41. Luo Y, Ding N, Shi X, Wu Y, Wang R, Pei L, Xu R, Cheng S, Lian Y, Gao J et al. Generation and comparative analysis of full-length transcriptomes in sweetpotato and its putative wild ancestor I. trifida. 2017;17(1):225
42. Xu Q, Zhu J, Zhao S, Hou Y, Li F, Tai Y, Wan X, Wei C. Transcriptome Profiling Using Single-Molecule Direct RNA Sequencing Approach for In-depth Understanding of Genes in Secondary Metabolism Pathways of Camellia sinensis. Front Plant Sci. 2017;8:1205.
43. Li J, Harata-Lee Y, Denton MD, Feng Q, Rathjen JR, Qu Z, Adelson DL. Long read reference genome-free reconstruction of a full-length transcriptome from Astragalus membranaceus reveals transcript variants involved in bioactive compound biosynthesis. Cell Discov. 2017;3:17031.
44. Wang J, Yao L, Li B, Meng Y, Ma X, Wang H. Single-Molecule Long-Read Transcriptome Dataset of Halophyte Halogeton glomeratus. Frontiers in Genetics. 2017;8:197.
45. Bayega A, Wang YC, Oikonomopoulos S, Djambazian H, Fahiminiya S, Ragoussis J. Transcript Profiling Using Long-Read Sequencing Technologies. Methods Mol Biol. 2018;1783:121-47.
46. Au KF, Underwood JG, Lee L, Wong WH. Improving PacBio long read accuracy by short read alignment. PLoS One. 2012;7(10):e46679.
47. Mahmoud M, Zywicki M, Twardowski T, Karlowski WM. Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. Genomics. 2017; doi:1016/j.ygeno.2017.12.011
48. Warburton ML, Crossa J, Franco J, Kazi M, Trethowan R, Rajaram S, Pfeiffer W, Zhang P, Dreisigacker S, Ginkel Mv. Bringing wild relatives back into the family: recovering genetic diversity in CIMMYT improved wheat germplasm. Euphytica. 2006;149(3):289-301.
49. Nevo E, Chen G. Drought and salt tolerances in wild relatives for wheat and barley improvement. Plant Cell Environ. 2010;33(4):670-85.
50. Wulff BB, Moscou MJ. Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. Front Plant Sci. 2014;5:692.
51. Zhang Z, Lin Z, Xin Z. Research progress in BYDV resistance genes derived from wheat and its wild relatives. Journal of Genetics and Genomics. 2009;36(9):567-73.
52. Yu, X., Feng, B., He, P., and Shan, L. 2017. From Chaos to Harmony: Responses and Signaling Upon Microbial Pattern Recognition. Annual review of phytopathology.
53. Lu, D., Wu, S., Gao, X., Zhang, Y., Shan, L., and He, P. 2010. A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. Proc Natl Acad Sci U S A 107:496-501.
54. Li, X., Kapos, P., and Zhang, Y. 2015b. NLRs in plants. Current opinion in immunology 32:114-121.
55. Gouveia Bianca C., Calil Iara P., Machado João Paulo B., Santos Anésia A., Fontes Elizabeth P B., (2016). Immune Receptors and Co-receptors in Antiviral Innate Immunity in Plants., Front Microbiol, 7, 2139.
56. Liu, N., Hake, K., Wang, W., Zhao, T., Romeis, T., and Tang, D. 2017. CALCIUM-DEPENDENT PROTEIN KINASE5 associates with the truncated NLR protein TIR-NBS2 to contribute to exo70B1-mediated immunity. The Plant cell:tpc. 00822.02016.
57. Meng, X., and Zhang, S. 2013. MAPK cascades in plant disease resistance signaling. Annu Rev Phytopathol 51:245-266.
58. Rayapuram, C., Jensen, M. K., Maiser, F., Shanir, J. V., et al. (2012). Regulation of basal resistance by a powdery mildew-induced cysteine-rich receptor-like protein kinase in barley. Molecular Plant Pathology, 13(2), 135-147.
59. Kohorn BD, Kohorn SL: The cell wall-associated kinases, WAKs, as pectin receptors. Front Plant Sci 2012, 3:1-5.
60. Jacob, F., Vernaldi, S., and Maekawa, T. 2013. Evolution and Conservation of Plant NLR Functions. Frontiers in immunology 4:297.
61. Maekawa, T. , Kufer, T. A. , & Schulze-Lefert, P. . (2011). Nlr functions in plant and animal immune systems: so far and yet so close. NATURE IMMUNOLOGY, 12(9), 817-826.
62. Chiang, Y. H. , & Coaker, G. . (2015). Effector triggered immunity: nlr immune perception and downstream defense responses. The Arabidopsis Book, 13, e0183.
63. Benko-Iseppon, A. M. , Crovella, S. , Guida-Santos, M. , Neto, J. P. B. , Santos, R. D. F. D. , & Amorim, L. L. B. A. . (2017). Transcription factors involved in plant resistance to pathogens. Curr Protein Pept Sci, 18(4), -.
64. Birkenbihl Rainer P., Liu Shouan., Somssich Imre E., (2017). Transcriptional events defining plant immune responses., Curr. Opin. Plant Biol., 38, 1-9.

65. Chen, C. and Chen, Z. (2000) Isolation and characterization of two pathogen- and salicylic acid-induced genes encoding WRKY DNA-binding proteins from tobacco. Plant Mol. Biol. 42, 387-396.
66. Yang, P., Chen, C., Wang, Z., Fan, B. and Chen, Z. (1999) A pathogen- and salicylic acid-induced WRKY DNA-binding activity recognizes the elicitor response element of the tobacco class I chitinase gene promoter. Plant J. 18, 141-149.
67. Yoshii, M. . (2009). Disruption of a novel gene for a nac-domain protein in rice confers resistance to rice dwarf virus. Plant J., 57(4), 615-625.
68. Xiaolin, L. , Shuhong, F. , Wei, H. , Guoyin, L. , Yunxie, W. , & Chaozu, H. , et al. (2017). Two cassava basic leucine zipper (bzip) transcription factors (mebzip3 and mebzip5) confer disease resistance against cassava bacterial blight. Frontiers in Plant Science, 8, 2110-.
69. Fischer, U., & Dröge-Laser, W. (2004). Overexpression of nterf5, a new member of the tobacco ethylene response transcription factor family enhances resistance to tobacco mosaic virus. Molecular Plant-Microbe Interactions, 17(10), 1162-1171.
70. Furtado A. RNA extraction from developing or mature wheat seeds. Methods Mol Biol. 2014;1099:23–8.
71. Pacific Biosciences: https://github.com/PacificBiosciences/cDNA_primer/wiki/ Understanding-PacBio-transcriptome-data#getFL. Accessed 15 April 2018. 2018.
72. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150-2.
73. Proovread: https://github.com/BioInf-Wuerzburg/proovread. Accessed 28 April 2018. 2018.
74. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35:W345-349.
75. txCDsPredict: https://github.com/bowhan/kent/tree/master/src/hg/txCds/txCdsPredict. Accessed 8 July 2018. 2018.
76. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 2013;41(17):e166.
77. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44(D1):D279-285.
78. Altschul SF, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.
79. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674-6.
80. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. Nucleic Acids Res. 2005;33:W116-20.
81. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L et al. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 2006;34:W293-7.
82. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35:W182-5.
83. Draw Venn Diagram: http://bioinformatics.psb.ugent.be/webtools/Venn/. Accessed 20 June 2018. 2018
84. Ernst, J. , & Bar-Joseph, Z. . (2006). Stem: a tool for the analysis of short time series gene expression data. BMC Bioinformatics, 7(1), 191-0.
85. Lokesh, K. , & Matthias, E. F. . (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation, 2(1), 5-7.
86. Zheng Yi., Jiao Chen., Sun Honghe., Rosli Hernan G., Pombo Marina A., Zhang Peifen., Banf Michael., Dai Xinbin., Martin Gregory B., Giovannoni James J., Zhao Patrick X., Rhee Seung Y., Fei Zhangjun., (2016). iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases., Mol Plant, 9, 1667-1670.
87. Schmittgen, T.D., Livak, KJ., 2008. Analyzing real-time PCR data by the comparative CT method. Nat Protoc. 3(6), 1101-1108.

Supplementary Files Legend

Figures S1 Length distribution of Polymerase reads and Subreads.

Figures S2 ROI length distributions of each libraries.

Figures S3 Pie chart of (ROI) reads of insert classify summary.

Table S1 Summary of overview for Pacbio Polymerase Reads Subreads.

Table S2 Statistic of trimmed and error corrected Pacbio isoforms by RNA-seq dataset.

Table S3 Detailed functional annotation results.

Table S4 Summary of RNA-seq raw reads and clean reads from twelve libraries.

Table S5 Summary of the identification of transcription factors and transcriptional regulators.

Table S6 Primers used in qRT-PCR for validation of RNA-seq data.

Figures

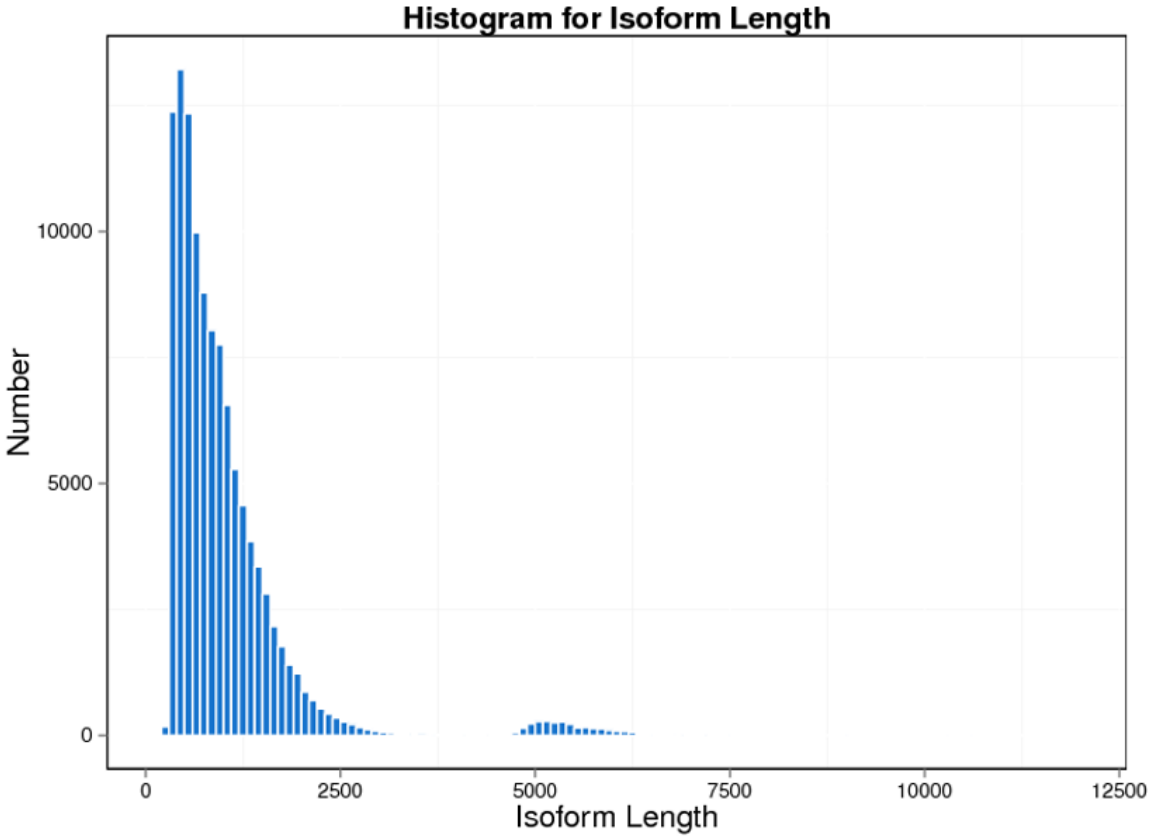

Figure 1

Length distribution of final consensus isoforms from merged libraries.

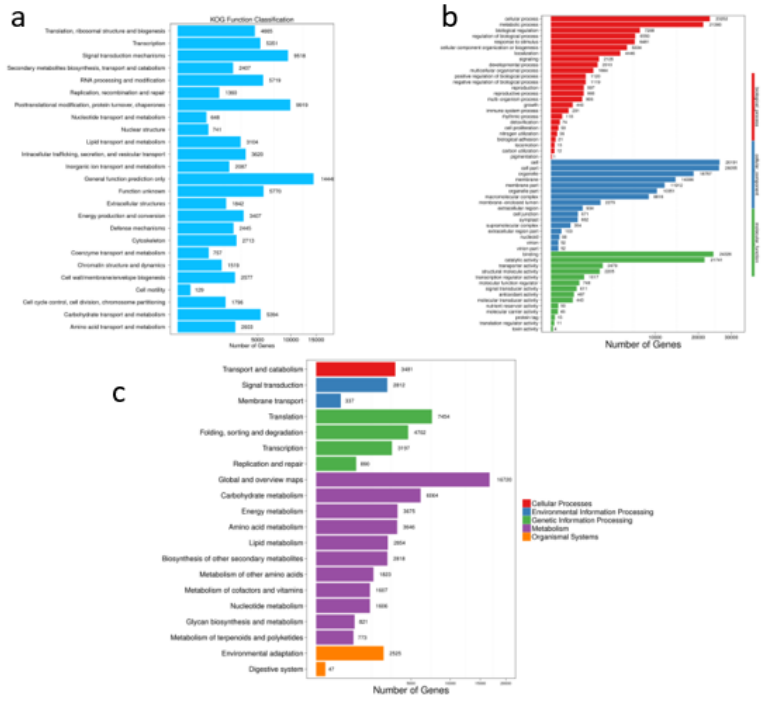

Figure 2

Functional distribution of KOG, GO, KEGG annotation.

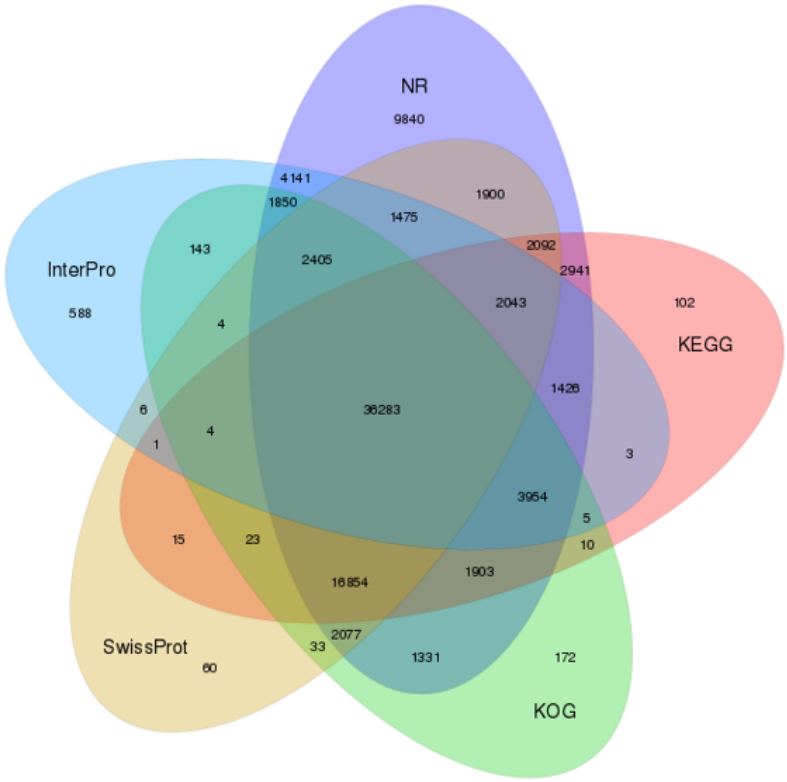

Figure 3

Annotation of venn diagram between the database of NR, KOG, KEGG, Swissprot and Interpro.

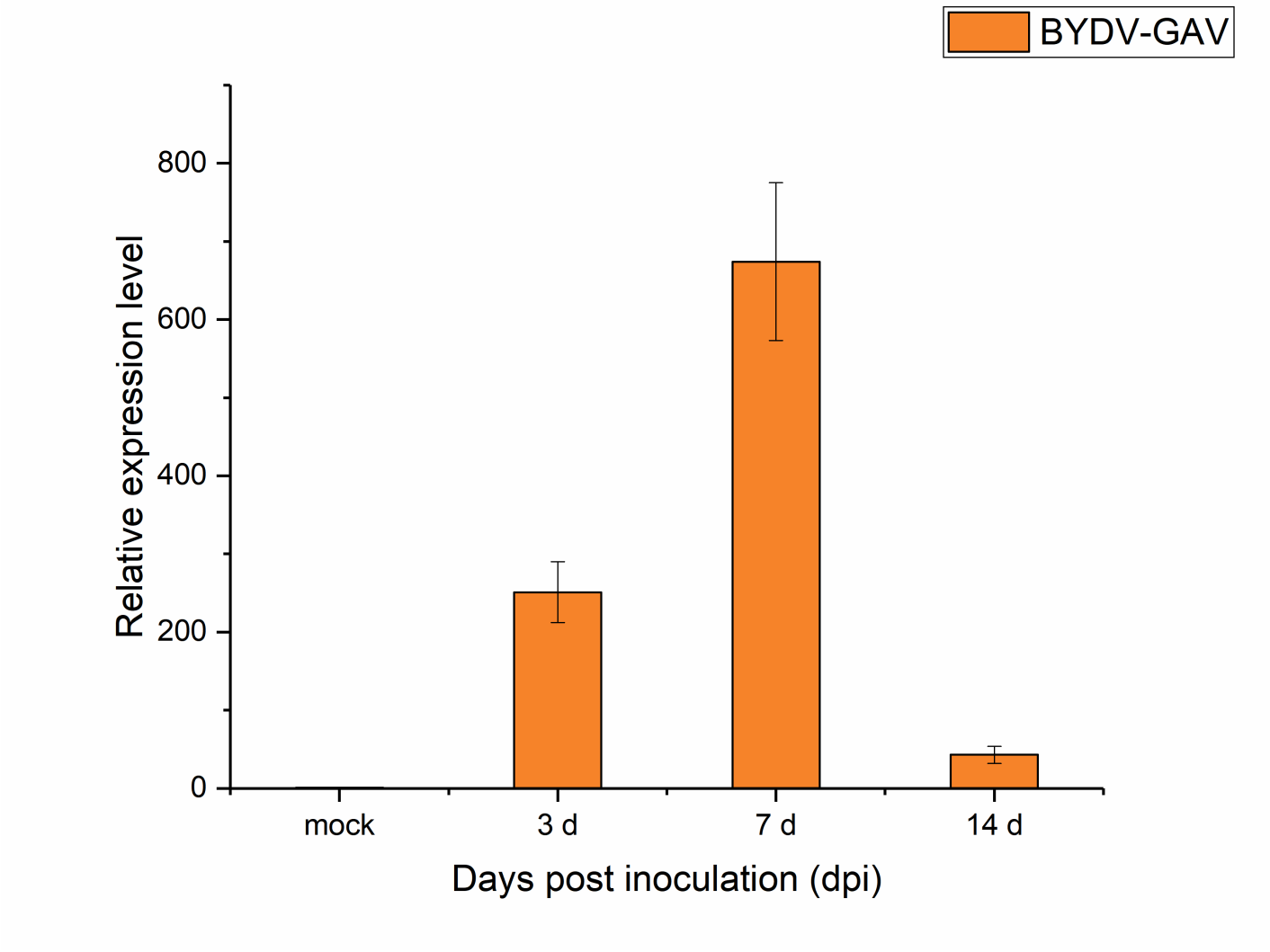

Figure 4

The titer of BYDV-GAV infected in P. huashanica at mock, 3 dpi, 7 dpi and 14 dpi.

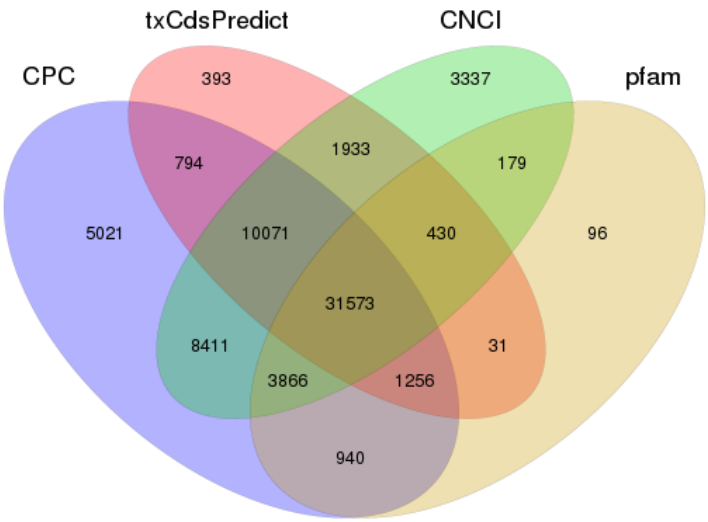

Figure 5

Venn diagram of the statistic of all coding sequences.

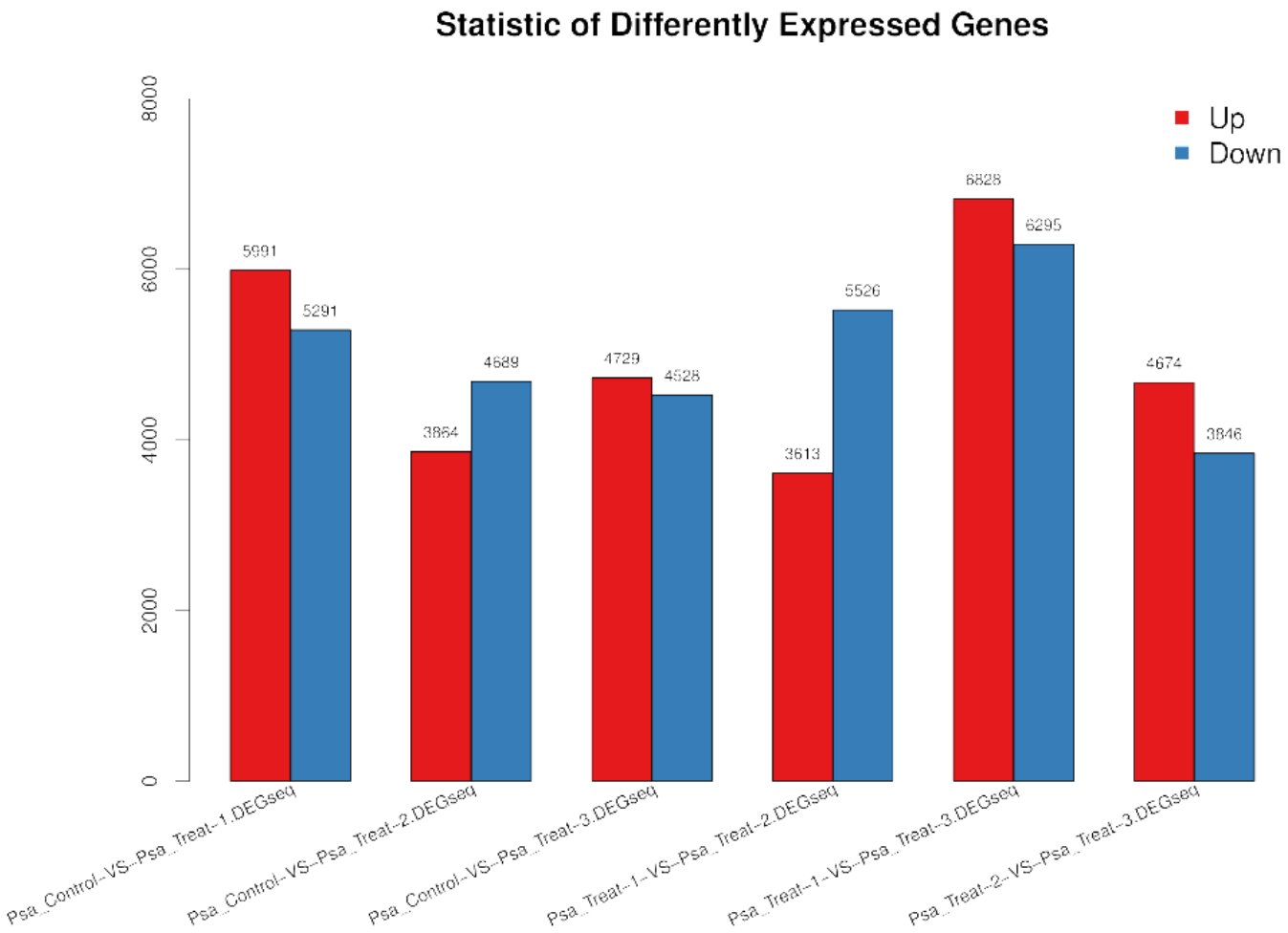

Figure 6

Statistic of DEGs of BYDV-GAV infected in P. huashanica at different time points.

Figure 7

Volcano plots of DEGs of BYDV-GAV infected in P. huashanica at 3dpi, 7dpi and 14dpi compare to mock, respectively.

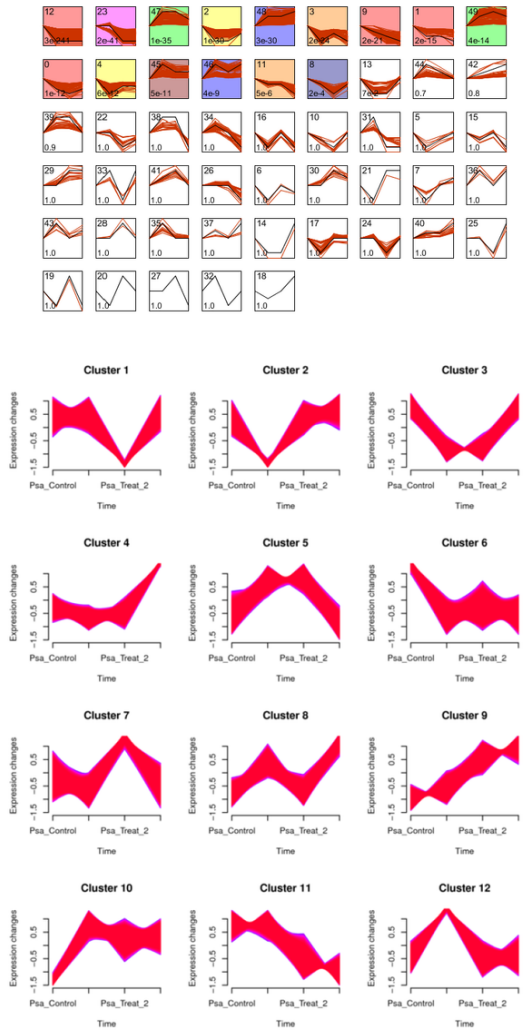

Figure 8

Cluster and STEM analyses of DEGs and all detective genes. A. Forty clusters were obtained using STEM software, and profiles are ordered based on the number of significant P-values assigned to genes (P-value≤0.05). The top number is the gene number assigned in each cluster and the bottom number is the P-value.

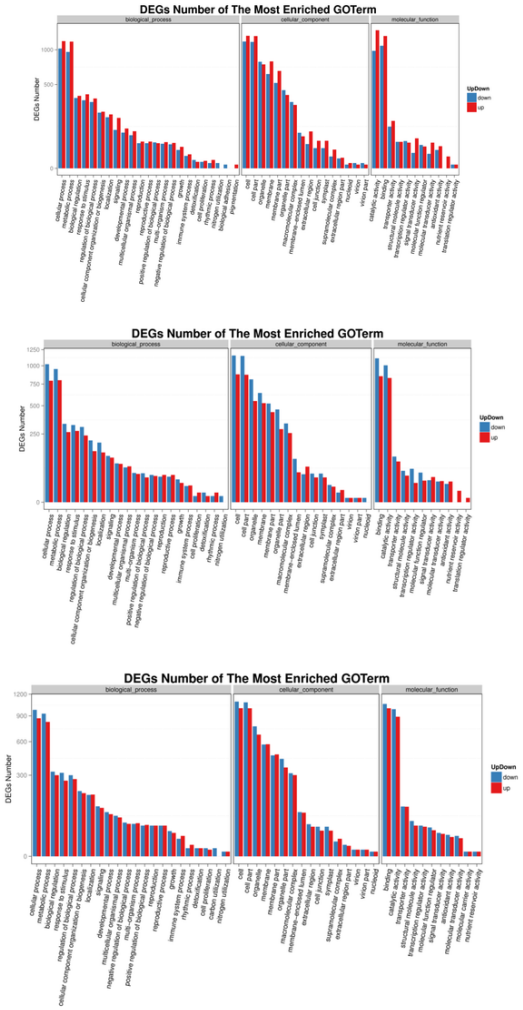

Figure 9

GO pathway enrichment of DEGs of BYDV-GAV infected in P. huashanica at 3 dpi, 7 dpi and 14 dpi compare to mock, respectively.

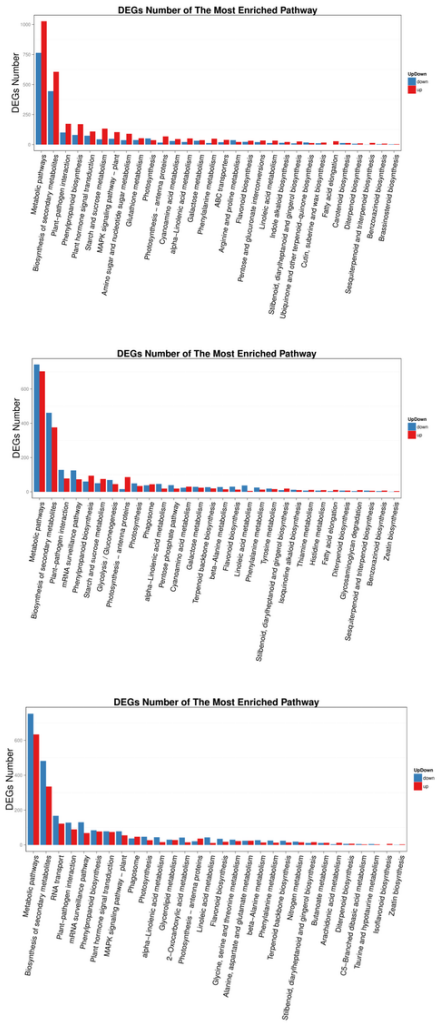

Figure 10

KEGG pathway enrichment of DEGs of BYDV-GAV infected in P. huashanica at 3 dpi, 7 dpi and 14 dpi compare to mock, respectively.

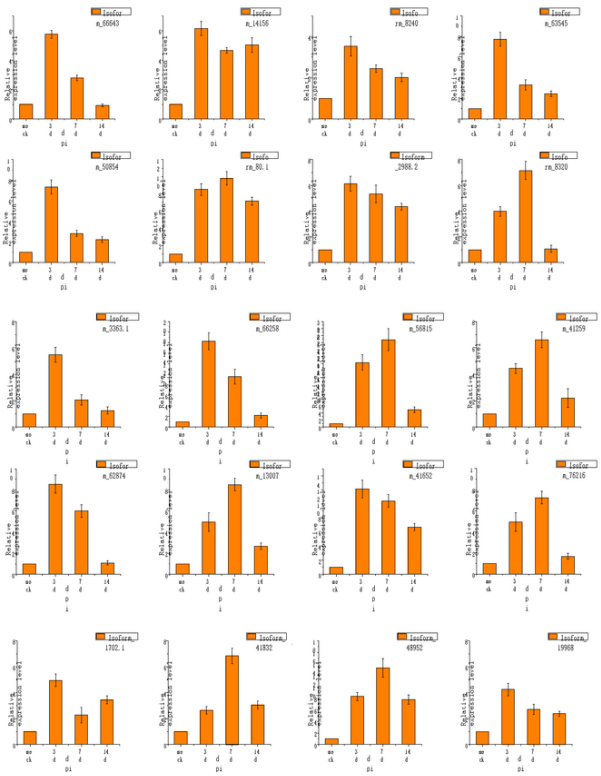

Figure 11

Validation of gene expression using qRT-PCR at mock, 3 dpi, 7 dpi and 14 dpi.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.doc
- TableS3.xls
- TableS2.xls
- TableS1.xls
- TableS4.xls
- TableS5.xls
- FigureS2.doc
- FigureS3.docx
- TableS6.xlsx