


RESEARCH

Open Access



# Integrated single-molecule long-read sequencing and Illumina sequencing reveal the resistance mechanism of *Psathyrostachys huashanica* in response to barley yellow dwarf virus-GAV

Chuan Shen, Caiyan Wei, Jingyuan Li, Xudong Zhang and Yunfeng Wu\* 

## Abstract

Although *Psathyrostachys huashanica* has excellent potential for resistance gene mining and molecular genetic breeding, no reference genome is available. To date, most studies of *P. huashanica* have been focused on the creation of translocation lines and additional lines, as well as the development of molecular markers. Therefore, research at the transcriptional level is lacking. In this study, the full-length transcriptome of *P. huashanica* was sequenced using PacBio isoform sequencing (Iso-Seq) of a pooled RNA sample to explore the potential full-length transcript isoforms. We obtained 112,596 unique transcript isoforms with a total length of 114,957,868 base pairs (bp). Subsequently, Illumina sequencing reads were used to correct and trim the PacBio isoforms. We annotated 103,875 unigenes in at least one functional database, and identified a plethora of differentially-expressed genes (DEGs) that are involved in the defense responses of *P. huashanica* against barley yellow dwarf virus-GAV (BYDV-GAV). Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis revealed that these DEGs were mostly involved in plant-pathogen interaction, plant hormone signal transduction, and the mitogen-activated protein kinase (MAPK) signaling pathway. Additionally, we selected twenty of the RNA-seq identified resistance-related up-regulated genes, including MAPKs, cysteine-rich receptor-like protein kinases (CRPKs), calcium-dependent protein kinases (CDPKs), pathogenesis-related protein (PR) proteins, WRKYs, and disease resistance proteins, and validated their up-regulation in response to BYDV-GAV by quantitative real-time PCR. Our results indicate that a series of defense-related genes were induced in *P. huashanica* during BYDV-GAV infection. The full-length transcriptome dataset will contribute to improved use of stress-resistance genes of *P. huashanica*, and serves as a reference database for the analysis of transcript expression in *P. huashanica*.

**Keywords:** *Psathyrostachys huashanica*, Full-length transcriptome, Illumina sequencing, PacBio, BYDV-GAV

## Background

*Psathyrostachys huashanica* Keng ( $2n = 2x = 14$ , NsNs) is a perennial cross-pollinating plant (tribe *Triticeae*, family *Poaceae*), which is a rare and endangered plant and a wild relative of crops in need of protection (Baden

1991). It grows on the residual soil of the steep braes and rocky slopes of the Huashan Pass in the Qinling Mountains in Shaanxi Province of northwest China (Lu 1995). *P. huashanica* is regarded as a potential omnipo-tent germplasm resource due to its excellent performance at early maturity, multiple florets and kernels, and high resistance to both abiotic (drought, barren soil, and salinity) and biotic stress (wheat fungal diseases, such as

\* Correspondence: [wuyunfeng@nwfau.edu.cn](mailto:wuyunfeng@nwfau.edu.cn)

State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University, Yangling 712100, China



© The Author(s). 2020 **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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

stripe rust, leaf rust, powdery mildew, and take-all, and virus diseases, such as the wheat yellow dwarf disease) (Chen et al. 1991; Wang and Shang 2000; Wang et al. 2011a, 2011b; Song et al. 2013; He et al. 2015; Wang et al. 2015; Kang et al. 2016).

The development of alien addition lines is vital for transferring useful genes from exotic species into common wheat (Ali et al. 2016). In 1991, scientists began to create hybrid materials between common wheat and *P. huashanica*. It is an effective and economical method to introduce agronomically desirable characteristics into available wheat cultivars (Chen et al. 1991). For many years, chromosome engineering techniques have been widely used for the targeted introgression of resistance genes from wild wheat relatives to hexaploid wheat (Zhang et al. 2017). By using chromosome engineering techniques, scientists crossed common wheat cv. 7182 and *P. huashanica* to generate the hexaploid hybrid H8911 (2n = 49, AABBDDNs) (Chen et al. 1991). To study its resistance to stripe rust, genomic in situ hybridization (GISH) and EST-SSR techniques were used to create a *Triticum aestivum*-*P. huashanica* Keng monosomic addition line, PW11, with high resistance to stripe rust. This led to the discovery that a small terminal segment of the 3Ns genome carrying new stripe rust resistance genes resulted in high levels of disease resistance (Du et al. 2014a). In the same way, the wheat of *P. huashanica* Keng 2Ns, 4Ns, 6Ns, and 7Ns disomic addition lines were generated and, in those lines, enhanced tiller numbers and disease resistance were also found (Du et al. 2013a; Du et al. 2013b; Du et al. 2014b; Du et al. 2014c). A recent study showed that *P. huashanica* exhibits high resistance to the barley yellow dwarf virus-GAV (BYDV-GAV) and can disrupt the movement of viral particles from inoculated leaves to new leaves (Song et al. 2013).

Barley yellow dwarf viruses (BYDVs) are a class of single-stranded RNA viruses, belonging to the family *Luteoviridae*, which are phloem-limited and obligately transmitted by aphids in a persistent, circulative manner (Miller and Rasochova 1997). Based on the serological properties of the viruses and the specificity of their aphid vectors, BYDVs have been classified into several distinct genus, including BYDV-PAV, -PAS, -MAV, -kerII, and -ker III that belong to the genus *Luteovirus*; CYDV-RPS, and BYDV-RPV, -RMV that belong to the genus *Polerovirus* (Liu et al. 2020). Among these, a similar serological reaction was observed between BYDV-GAV and BYDV-MAV. Hence, BYDV-GAV is identified as a Chinese isolate of BYDV-MAV and is most often found in China in recent years (Zhou et al. 1984; Jin et al. 2004; Domier 2012). BYDV-GAV 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*) (Jin et al. 2004; Zhang et al. 2009). In the field, symptoms of leaf yellowing and plant dwarfism can be observed in several wheat cultivars infected with BYDV-GAV, which may cause yield loss in severe cases (Banks et al. 1995).

Transcriptomics is the study of gene structure, expression, and regulation. Short-read second-generation sequencing (SGS) has become a powerful tool for quantifying gene expression levels and exploring transcriptional gene regulation (Buermans and den Dunnen 2014). However, SGS has its limitations, such as 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 (Rhoads and Au 2015; An et al. 2018). 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, enhancing our understanding of the transcriptome complexity (Rhoads and Au 2015; An et al. 2018).

Full-length cDNA sequencing is fundamental to structural and functional genomics studies, and is used 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 (Roberts et al. 2013; Liu et al. 2017a). Additionally, SMRT sequencing is used for the assembly of plant genomes based on its long reads, such as its application in *Triticum aestivum* (Zimin et al. 2017a), loblolly pine (Zimin et al. 2017b), *Zea mays* (Jiao et al. 2017) and rubber tree (Pootakham et al. 2017). On the other hand, full-length cDNA sequencing has been combined with Illumina-based RNA-seq datasets in *Sorghum bicolor* (Abdel-Ghany et al. 2016), *Zea mays* (Wang et al. 2016), cotton (Wang et al. 2018), sugarcane (Hoang et al. 2017), coffee bean (Cheng et al. 2017) and sweet potato (Luo et al. 2017), to characterize the complexity of transcriptomes. Moreover, SMRT sequencing has also been widely used to construct transcriptomes in some plants without a reference genome, such as *Camellia sinensis* (Xu et al. 2017), *Astragalus membranaceus* (Li et al. 2017a), and *Halogeton glomeratus* (Wang et al. 2017), but it has not been used in *P. huashanica*.

In this study, we constructed full-length cDNA libraries of *P. huashanica* and performed SMRT sequencing to generate large-scale full-length transcripts. Then, we used full-length transcripts as a reference to identify differentially-expressed genes (DEGs) in *P. huashanica* in response to BYDV-GAV infection and found plenty of defense-related genes induced by the viral invasion. Our results provide novel insights into the resistance response of *P. huashanica* to BYDV-GAV infection and may contribute to the utilization of resistance resources of *P. huashanica* in the future.

Results

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

To obtain the desirable transcript isoforms, high-quality RNA was extracted from four types of *P. huashanica* tissue to generate four RNA samples, which were subsequently pooled in equal amounts. Two size-fractionated, full-length cDNA libraries (1–5 kb and 4.5–10 kb) were constructed. These were 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 sub-reads with an average of 6.9 Gb per cell were generated based on SMRT sequencing data (Additional file 1: Table S1). After quality control, we obtained 3,677,921, 244 bp circular consensus sequences (CCS), and 1,688,359 reads of insert (ROIs), of which 1,140,528 ROIs were in the 1–5 kb library, and 547,831 in the 5–10 kb library (Table 1). The ROI length distribution of each library is shown in Additional file 2: Figure S1. Among these ROIs, 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 reads (Additional file 2: Figure S2). Only full-length non-chimeric reads and non-full-length reads were used in further analyses, while short reads with a length of < 300 bp and chimeric reads were discarded for subsequent analysis.

Functional annotation of coding genes

We further clustered the full-length non-chimeric reads into consensus groups. For each cluster with sufficient full-length and non-full-length coverage, Quiver was run to polish the consensus. Only high-Quiver output (QV) consensus sequences were used for downstream analysis. High-quality consensus isoforms of each library were merged, and redundancy was removed. We obtained 112,596 final isoforms from 177,882 isoforms (Table 2), and the length distribution of the final consensus isoforms is shown in Additional file 2: Figure S3. We used RNA-seq reads to trim all final consensus isoforms to improve the

PacBio transcript quality. After this, we obtained mapping identity, substitution, deletion, insertion, and mapping coverage (Additional file 1: Table S2). Then we used BLAST, BLAST2GO, and InterProScan5 to identify functional annotation terms from the Non-redundant Protein Sequence Database (Nr, NCBI), Nucleotide Sequence Database (Nt, NCBI), Gene Ontology (GO), Eukaryotic Orthologous Groups (KOG), Kyoto Encyclopedia of Genes and Genomes (KEGG), SwissProt, and InterPro protein databases. We annotated 103,875 (92.25%) uni-genes with at least one functional database (Table 3). Functional distribution was statistically assessed with KOG, GO, and KEGG annotation (Fig. 1).

We obtained 36,283 annotated transcripts among Nr, KEGG, SwissProt, and InterPro (Additional file 2: Figure S4).

Illumina sequencing and assessment of sequencing quality

To determine the appropriate time points for performing RNA-seq analysis of *P. huashanica* in response to BYDV-GAV infection, we analyzed the time-course of changes in the amount of virus using quantitative RT-PCR (RT-qPCR). As shown in Fig. 2, the amount of BYDV-GAV in plant leaves increased dramatically at 3 days post-inoculation (dpi) and reached a peak at 7 dpi, then declined gradually until 14 dpi when the virus was barely detectable. Therefore, we selected mock-inoculated and BYDV-GAV-inoculated leaves at 3, 7, and 14 dpi for RNA-seq analysis.

Since the genome sequence of *P. huashanica* was not available, we used our PacBio sequencing dataset as a reference for RNA-seq. High-quality RNA was extracted from the four samples, and each sample had three repetitions to generate 12 libraries. The libraries were sequenced on an Illumina HiSeq2000 platform using paired-end sequencing. Over 16 GB of raw reads and 115 million clean reads were generated from each library (Additional file 1: Table S3). After that, clean reads were aligned to the PacBio dataset, and 76,313 transcripts were detected, including 16,827 lncRNA transcripts and 47,196 mRNA transcripts, which were used for FPKM analysis. The transcript coding quality was assessed by three types of prediction software (Fig. 3).

**Table 1** Summary of reads of insert (ROIs) 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)
<i>Psathyrostachys huashanica</i>	F01	565,467	301,093 (53.25%)	316,921 (56.05%)	292,646 (51.75%)	194,333 (34.37%)	451.5
<i>Psathyrostachys huashanica</i>	G01	575,061	345,129 (60.02%)	360,354 (62.66%)	333,960 (58.07%)	232,059 (40.35%)	457.5
<i>Psathyrostachys huashanica</i>	H01	547,831	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 clustered and corrected to create final isoforms

Library	sLibrary isoforms	Final isoforms
<i>Psathyrostachys huashanica</i> F01	70,137	49,273
<i>Psathyrostachys huashanica</i> H01	21,969	3158
<i>Psathyrostachys huashanica</i> G01	85,776	60,165
Total	177,882	112,596

### Identification of DEGs

DEGs were obtained by comparing the BYDV-GAV-inoculated samples at 3, 7 and 14 dpi with the mock-inoculated sample (a composite sample), respectively. DEGs were identified with the criteria of a |fold change|  $\geq 2$  and a  $q$ -value  $\leq 0.001$  at each time points (Fig. 4). In this way, 11,282 genes were identified as DEGs at 3 dpi, of which 5991 were up-regulated, and 5291 were down-regulated. Interestingly, there were 4689 down-regulated genes, and 3864 up-regulated genes at the 7 dpi. At 14 dpi, the number of up-regulated genes increased to 4729, exceeding the number of down-regulated genes (4528). We used volcano plots to visualize DEGs in a more intuitive way (Additional file 2: Figure S5). The distribution of up-regulated and down-regulated genes was consistent with the amount of virus accumulated in *P. huashanica*. All statistically significant DEGs were further characterized by functional annotation and enrichment analysis.

### Short time-series expression miner (STEM) analysis

We performed a clustering and visualization analysis using STEM software to explore the major expression trends of DEGs from the sequencing dataset over the four assessed time points. We found that all DEGs were assigned to 50 clusters that displayed similar expression patterns. Of these, 15 clusters were statistically significantly variably expressed ( $p$ -value  $\leq 0.05$ ) (Fig. 5a). Among these, clusters 45, 46, 47, 48, and 49 exhibited up-regulated expression patterns during viral infection, while the other clusters showed down-regulated expression patterns. Cluster 45 initially had an increased expression level at 3 dpi but was down-regulated at later infection stages. Clusters 47, 48, and 49 exhibited gradual increases in expression levels from 3 dpi to 7dpi but showed down-regulated expression levels in the following periods. Those up-regulated clusters might be

induced by the BYDV-GAV infection to resist viral replication and movement.

We also used the R 'Mfuzz' package to cluster the expression levels of all detected genes and achieved 12 clusters (Fig. 5b), among which clusters 9 (7.9%) and 10 (5.2%) exhibited up-regulation during all stages of viral infection. On the contrary, clusters 6 (9.1%) and 11 (5.6%) were down-regulated during all viral infection stages. Moreover, clusters 5 (5.3%), 7 (6.2%), 8 (8.0%), and 12 (11.8%) also play critical roles in response to the BYDV-GAV infection.

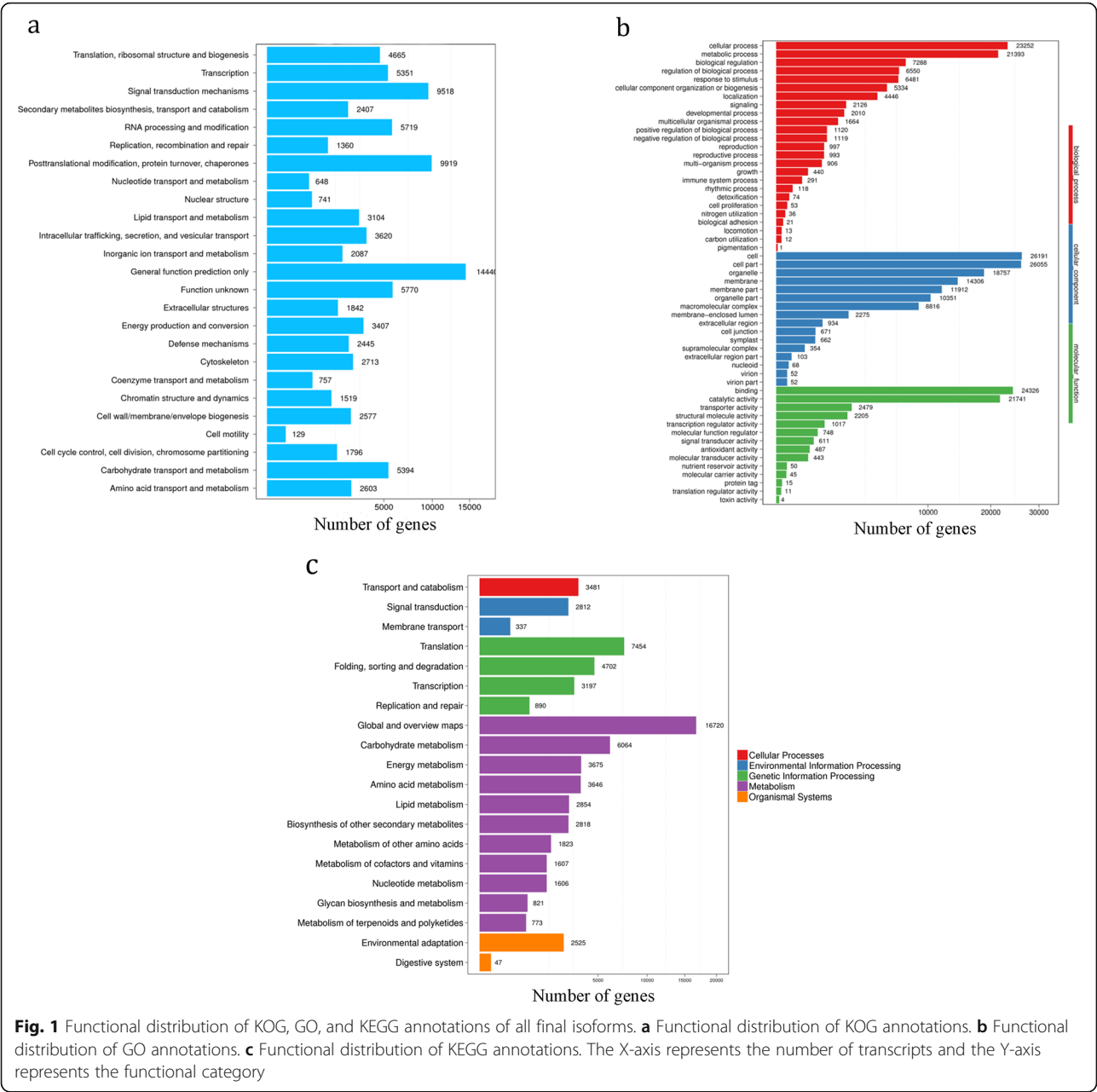
### GO and KEGG pathway enrichment analysis of DEGs

GO and KEGG pathway enrichment analysis was performed to investigate the underlying functions of BYDV-GAV-induced DEGs in *P. huashanica* at different time points. The GO enrichment results for DEGs between BYDV-GAV-inoculated samples at 3 dpi and mock-inoculated samples indicated that cellular and metabolic processes were the most significantly represented groups in the biological process category, while cells and cell parts were the most commonly enriched terms in the cellular component. Within the molecular function category, catalytic activity and binding were the predominantly enriched groups (Fig. 6a). The above results were similar to those obtained at other time points post viral infection (Fig. 6b, c), however, we also found some differences. For example, the number of up-regulated genes enriched in metabolic processes and catalytic activity was increased at 3dpi, whereas decreased at 7 dpi, and increased again at 14 dpi, which corresponded to the amount of virus present in the infected plants.

According to the KEGG database annotations, the most enriched pathway at all infection stages was the metabolic pathway, followed by biosynthesis of secondary metabolites. We found that DEGs related to plant-pathogen interaction, phenylpropanoid biosynthesis, starch and sucrose metabolism, the MAPK signaling pathway, and plant hormone signal transduction were all significantly enriched at 3 dpi (Fig. 7a). At 7 dpi, the pathways of plant-pathogen interaction, mRNA surveillance, phenylpropanoid biosynthesis, starch and sucrose metabolism, and glycolysis/gluconeogenesis were predominant (Fig. 7b). RNA transport, plant-pathogen interaction, plant hormone signal transduction, phenylpropanoid biosynthesis, mRNA surveillance pathway,

**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%

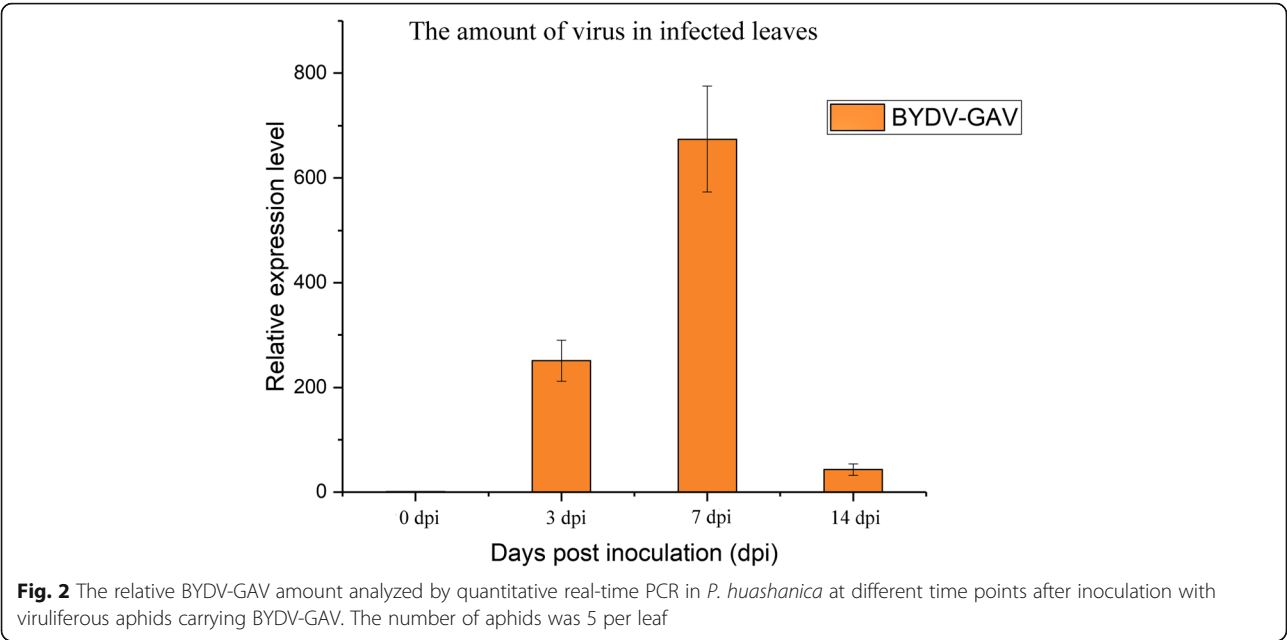


and the MAPK signaling pathway were significantly up-regulated at 14 dpi (Fig. 7c). In summary, most of the detected DEGs are putatively involved in resistance-related pathways. These genes appear to participate in the interactions between *P. huashanica* and BYDV-GAV.

**Identification of transcription factors and transcriptional regulators**

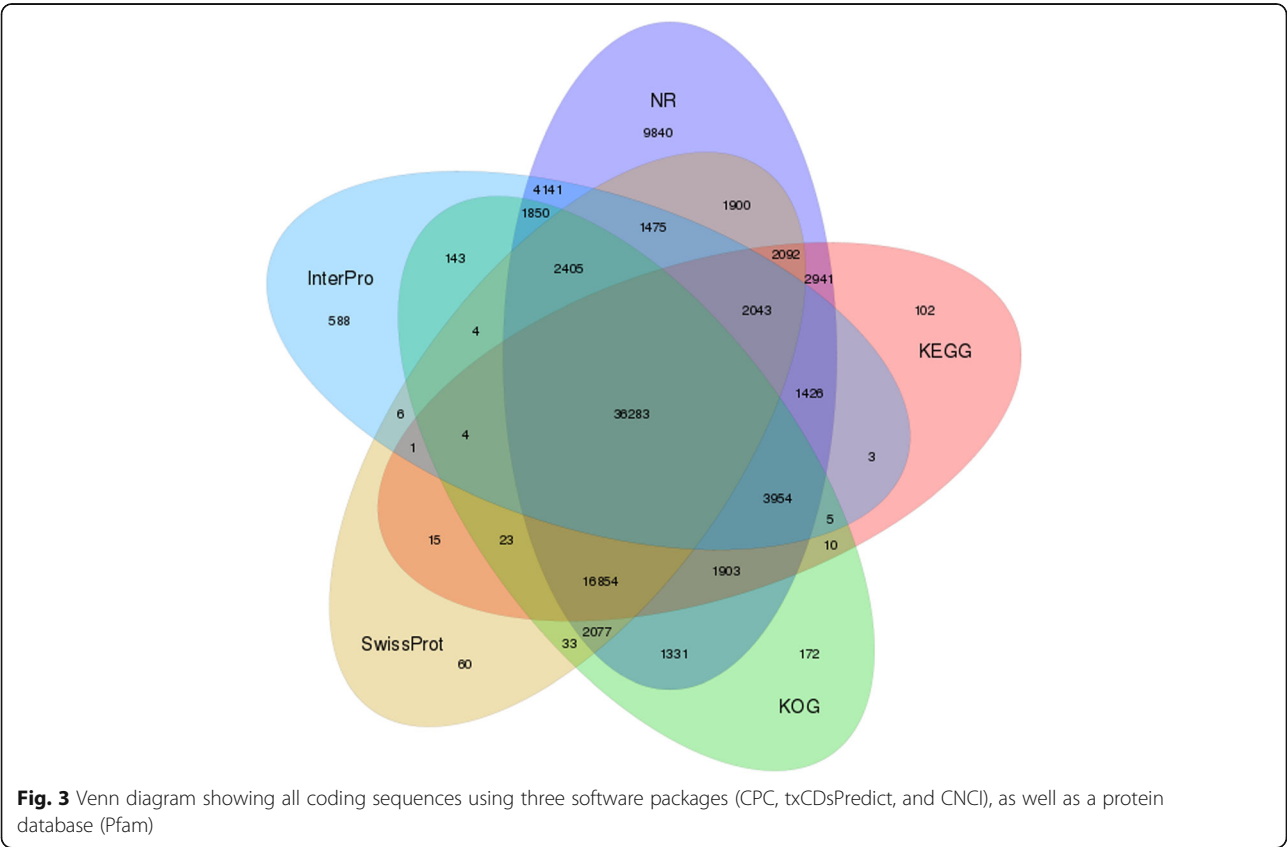
We identified 404 transcription factors and 201 transcription regulators distributed in 46 families among the 15,499 DEGs. These TFs were classified into the following families: AP2-EREBP (APETALA2/ethylene-responsive factor) (25

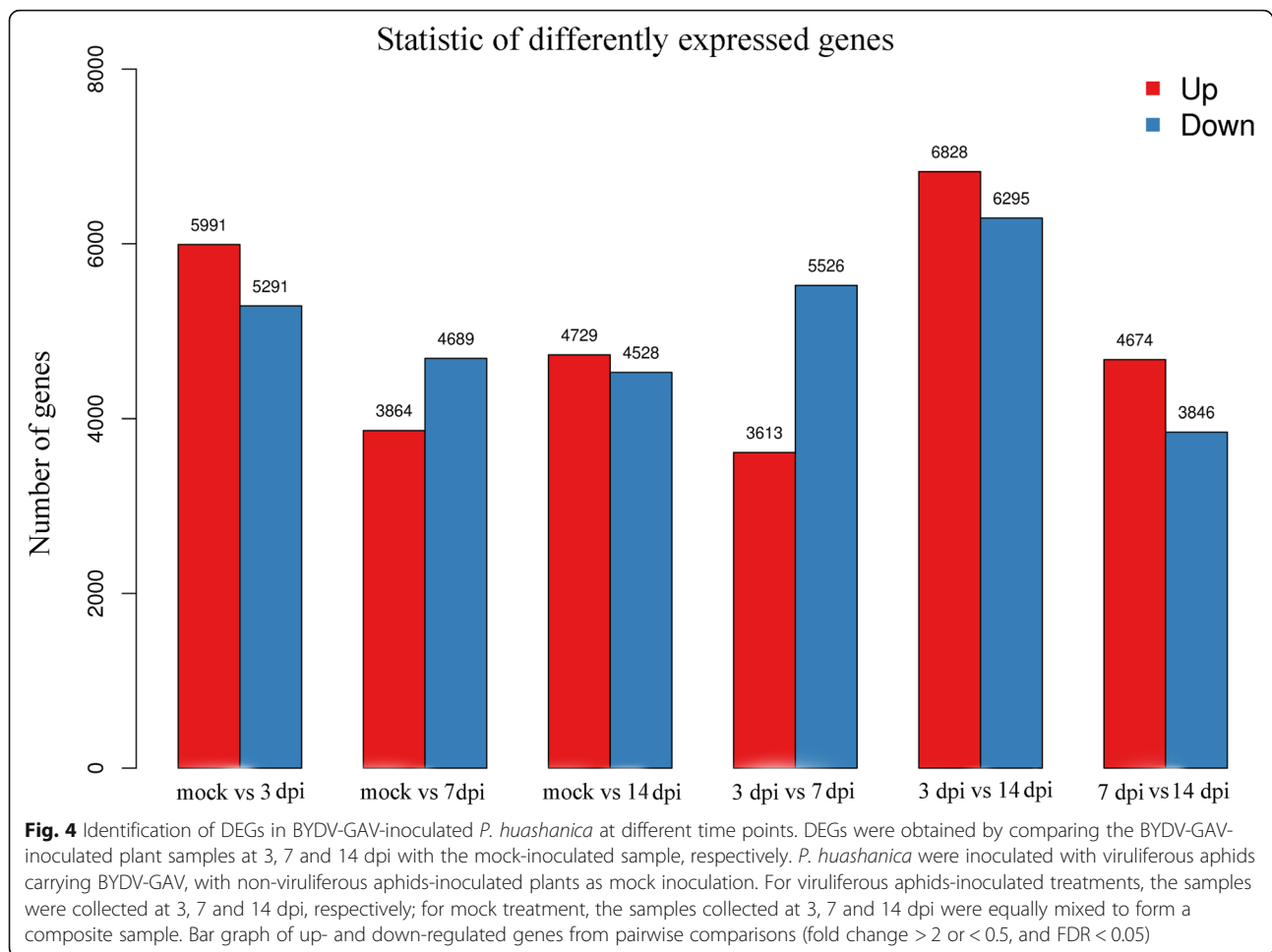
genes), WRKY (41 genes), bHLH (basic helix-loop-helix) (33 genes), MYB-related (40 genes), MYB (13 genes), NAC (NAM, ATAF1–2, and CUC2) (13 genes), C2H2 (13 genes), NF-YA (13 genes), GARP-G2-like (17 genes), and bZIP (basic region/leucine zipper motif) (22 genes) (Additional file 1: Table S4). Transcriptional regulators mainly focused on AUX/IAA (35 genes), SET (19 genes), TRAF (Tumor 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 1: Table S4). Our data indicate that TFs and TRs are widely involved in the defense response of *P. huashanica* to BYDV-GAV.



**Validation the expression profiles of DEGs using RT-qPCR**  
The expression of 20 representative genes selected from DEGs were verified by RT-qPCR. Genes related to the defense response were divided into three categories. The first class consists of 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 (WAK). The second class consists of defense-related proteins, such as pathogenesis-related protein (PR), heat shock





protein 70 (HSP70), and WRKY transcription factor (WRKY). The third class consists of resistance genes, such as those coding for disease resistance proteins. The RT-qPCR results confirmed that the expression profiles of DEGs were consistent with the results of RNA-seq (Fig. 8).

## Discussion

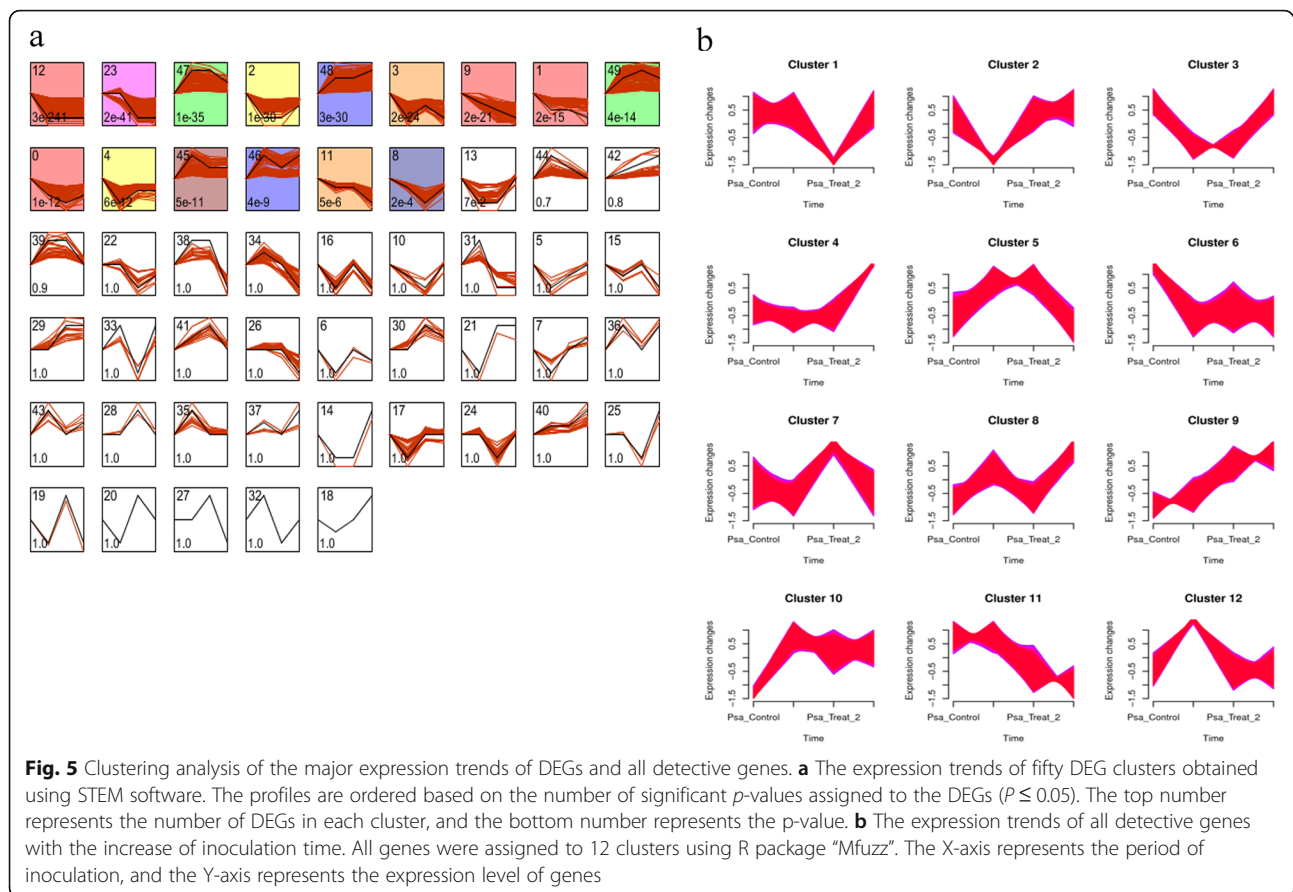
Although many studies related to *P. huashanica* have been performed, no reference genome is available for this plant, resulting in little progress in its functional genomics. Therefore, there was an urgent need to produce a reference genome or transcripts for *P. huashanica* to facilitate further study.

PacBio isoform sequencing (Iso-Seq) was developed and applied to the characterization of transcriptomes of different species, including plants and animals. It has many advantages, such as creation of long reads, and there is no need for assembly, making it especially suitable for non-model organisms that lack genomic sequences (Roberts et al. 2013; Liu et al. 2017a).

The long-standing and widespread use of *P. huashanica* has led to an intense interest in uncovering its

resistance mechanisms to numerous pathogens (Song et al. 2013; Kang et al. 2016). Because of the lack of a reference genome and reference transcripts, much of this interest has only been focused on the creation of translocation and addition lines. To tackle this problem, we performed long-read SMRT sequencing and analysis of four distinct tissues (i.e. the root, stem, leaf, and bud), from which we generated a complete transcriptome. Finally, we created an initial collection of 112,596 unique *P. huashanica* transcript isoforms. This transcriptome dataset provides a broad range of possibilities to study the multifaceted characteristics of *P. huashanica*.

The PacBio full-length isoform sequencing platform can be used to obtain transcripts without assembly to overcome the difficulty of short-reads in next-generation sequencing (Bayega et al. 2018). However, even though PacBio offers longer reads than other current platforms, it has a higher error rate (Au et al. 2012). The short-reads from the Illumina platform are 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 (Buermans and den Dunnen 2014).

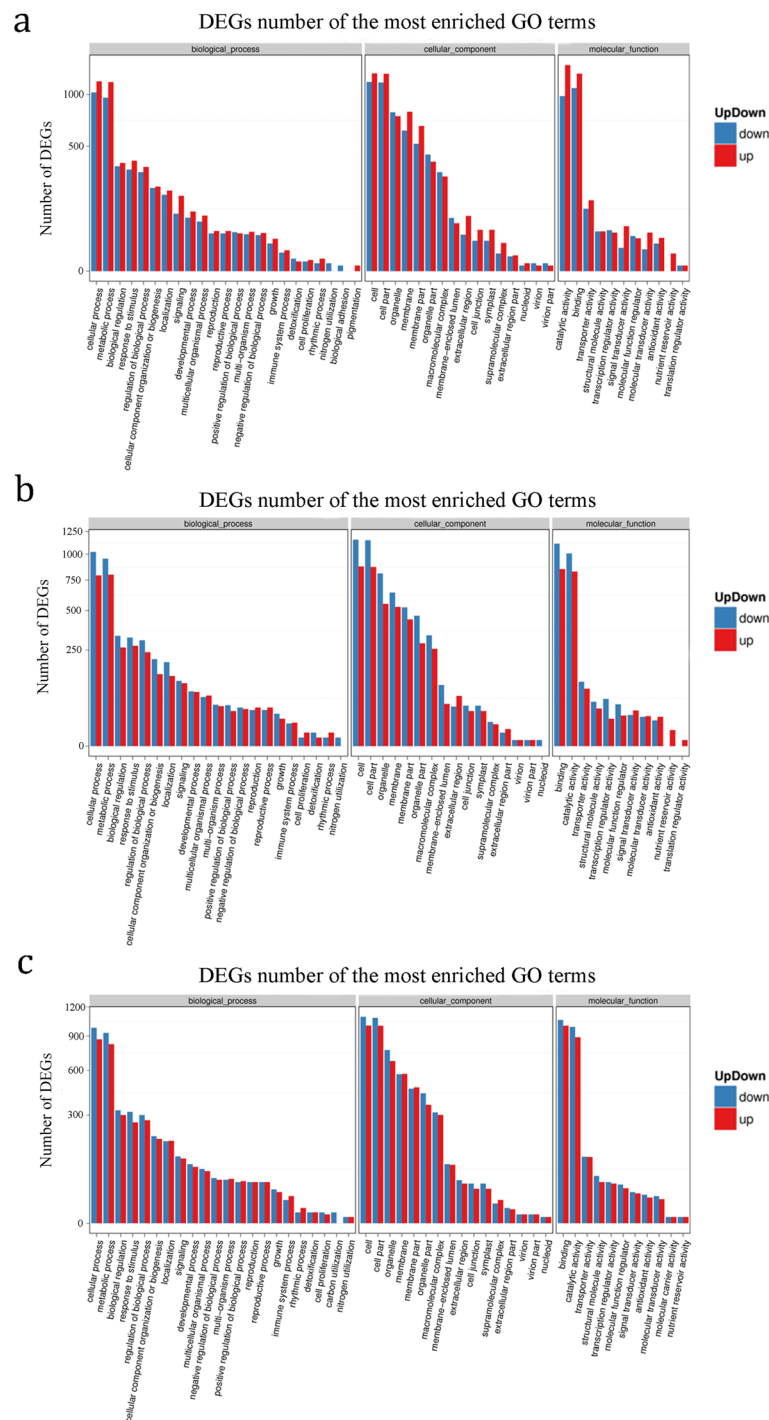


Therefore, we used clean reads created from short-read next-generation sequencing (NGS) technology to correct the PacBio transcript isoforms (Mahmoud et al. 2017).

We found that *P. huashanica* contains a huge transcription factor and resistance gene library. Previously, it was shown that using wild related species to create high-quality, durable, and resistant materials is an extraordinary way to recover the genetic diversity of wheat germplasm (Warburton et al. 2006; Nevo and Chen 2010; Wulff and Moscou 2014). Starting twenty years ago, research groups introgressed BYDV resistance genes in more than ten wild relative species belonging to *Thinopyrum*, *Agropyron*, *Elymus*, *Leymus*, *Roegneria*, and *Psathyrostachys* genera into wheat (Zhang et al. 2009). In this study, Illumina sequencing was performed to profile DEGs at several different time points for BYDV-GAV-infected and mock-inoculated *P. huashanica*. A total of 11,282, 8553, and 9257 genes were identified as DEGs at 3, 7, and 14 dpi, respectively. Most of these genes were enriched in plant-pathogen interaction, plant hormone signal transduction, and MAPK signaling pathways. In nature, plants have evolved a complicated system to recognize pathogens and activate defense responses. Plant pattern recognition receptors on the plasma membrane belong to the receptor-

like kinase or receptor-like protein family, which recognize conserved microbe-associated molecular patterns to trigger PAMP-triggered immunity (PTI) (Dodds and Rathjen 2010; Lu et al. 2010; Yu et al. 2017). In addition to PTI, plants also synthesize resistance (R) proteins to specifically recognize pathogen effectors or avirulence (Avr) factors and activate effector-triggered immunity (ETI) (Li et al. 2015; Gouveia et al. 2016). In our experiment, we found several MAPKs, CRPKs, CDPKs, WAKs, and serine/threonine-protein kinases that were significantly up-regulated during viral infection. These are potentially associated with innate immunity in plants (Kohorn and Kohorn 2012; Rayapuram et al. 2012; Meng and Zhang 2013; Liu et al. 2017b).

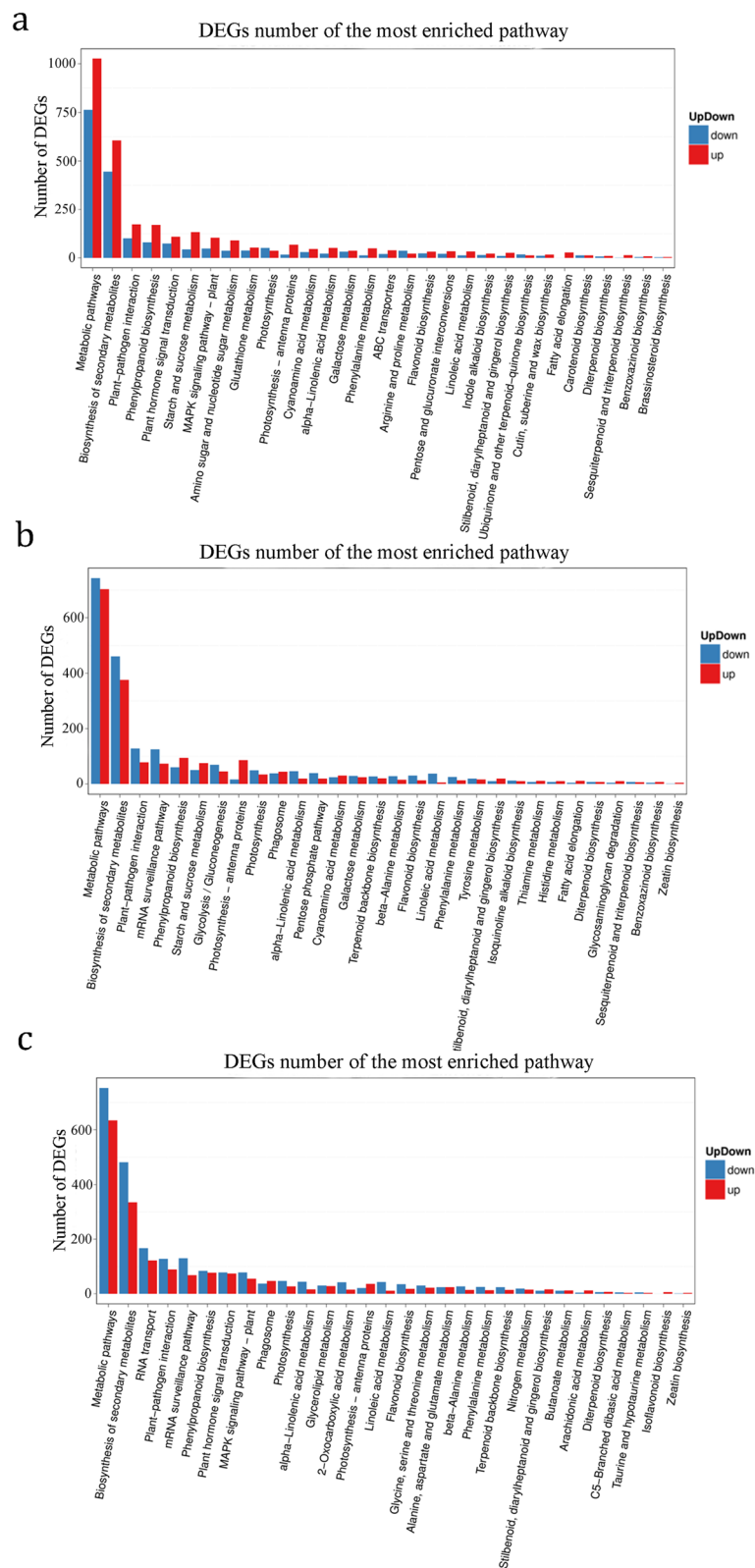
Additionally, the proteins belonging to the nucleotide-binding site-leucine-rich repeat (NBS-LRR) family, one of the largest gene families known in plants, are enriched in resistance responses and pathogen detection, in particular the effector molecules of pathogens responsible for virulence (Maekawa et al. 2011; Jacob et al. 2013; Chiang and Coaker 2015). Our sequencing results identified that many NLR resistance proteins were induced to high expression levels to defend against viral invasion at different time points. These results are suggestive of a defense response involving both ETI and PTI.



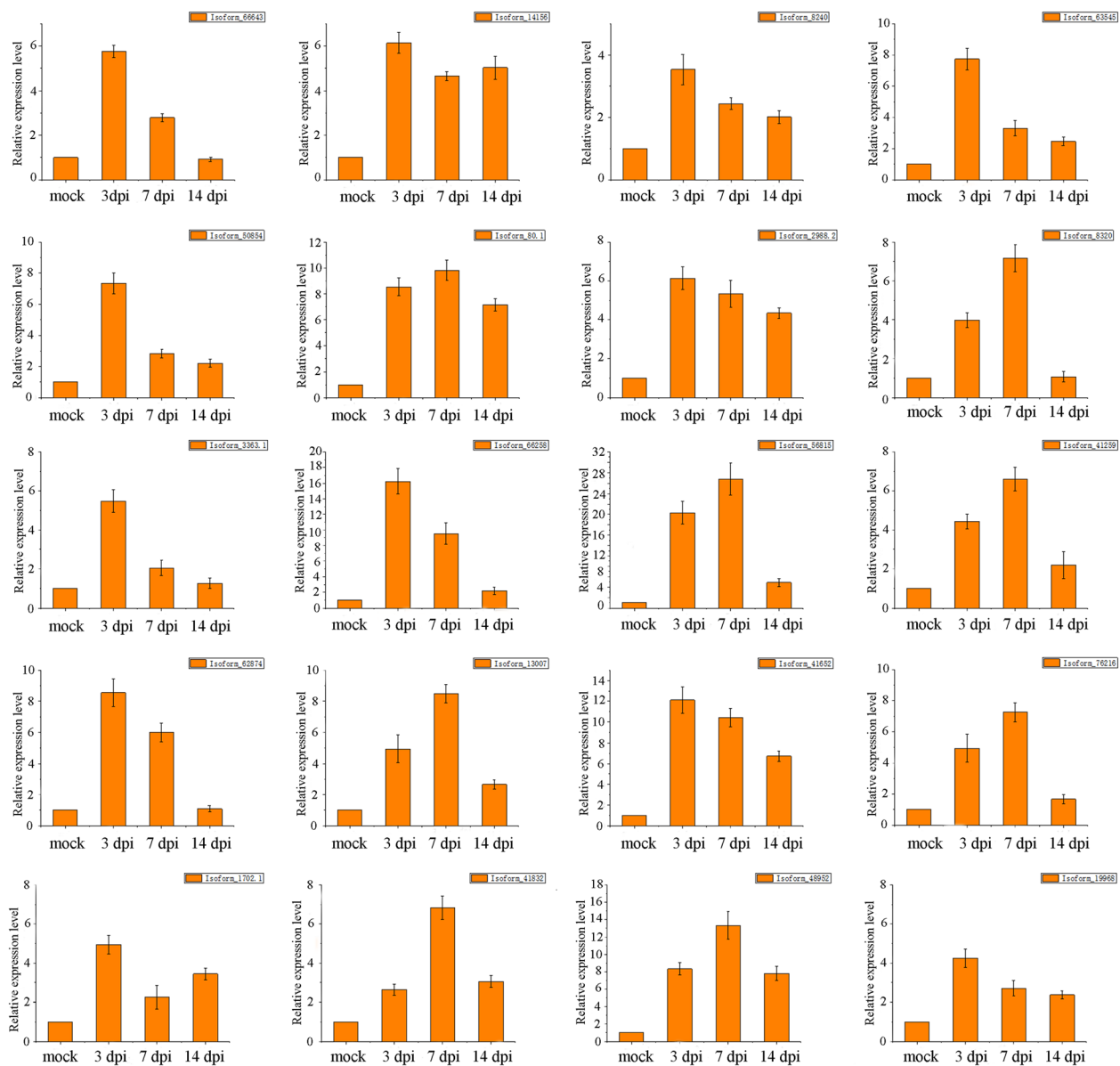
**Fig. 6** GO enrichment of DEGs between BYDV-GAV-inoculated *P. huashanica* at different time after inoculation and mock-inoculated samples, respectively. Enriched GO terms between BYDV-GAV-inoculated *P. huashanica* at 3 dpi and mock (**a**), 7 dpi and mock (**b**), 14 dpi and mock (**c**). The X-axis represents the functional category and the Y-axis represents the number of transcripts

Transcription factors (TFs) play a central role in response to multifaceted biotic stresses, such as insect attack and pathogen infection (Amorim et al. 2017). Many TF families are differentially expressed in numerous plants in response

to bacterial, fungal, and viral infections (Birkenbihl et al. 2017). Several families of TFs, such as MYB, WRKY, ERF, NAC, and bZIP, which are known critical regulators in the defense response, were found to be highly expressed in *P.*



**Fig. 7** KEGG pathway enrichment of DEGs between BYDV-GAV-inoculated *P. huashanica* at different time after inoculation and mock-inoculated samples, respectively. Enriched KEGG pathway terms between BYDV-GAV-inoculated *P. huashanica* at 3 dpi and mock **(a)**, 7 dpi and mock **(b)**, 14 dpi and mock **(c)**. The X-axis represents the functional category and the Y-axis represents the number of transcripts



**Fig. 8** Validation of the RNA-Seq results by RT-qPCR. Twenty DEGs were selected for RT-qPCR. The gene expression levels were normalized using the *P. huashanica* 18S rRNA gene as an internal reference. Error bars represent standard errors of three biological replicates

*huashanica* after viral infection (Yang et al. 1999; Chen and Chen 2000; Fischer and Dröge-Laser 2004; Yoshii et al. 2009; Li et al. 2017b).

RNA silencing-based antiviral defense is one of the primary response strategies against viral infections used by plants, in which the primary components consist of DCLs, AGOs, and RDRs (Guo et al. 2019). We also analyzed the expression of DCLs, AGOs, and RDRs and found that some DCLs and AGOs were up-regulated in response to virus infection. Therefore, we argue that RNA silencing is widely involved in the resistance of *P. huashanica* to the invasion and replication of BYDV-GAV.

In summary, the full-length transcript sequencing of *P. huashanica*, a wild relative of wheat, allowed us to identify high-quality germplasm resources for crop breeding. Moreover, we obtained a diversity of resistance genes induced by BYDV-GAV using Illumina sequencing, which could contribute to future antiviral breeding in wheat.

## Conclusions

We combined strand-specific Illumina RNA-seq and PacBio full-length cDNA sequences to achieve a reference transcriptome database with a low error rate in this study. After functional annotating the isoforms, we found that 82.17% isoforms had annotations in the Nr

database. Approximately 40.38% and 60.09% of the total isoforms could be annotated with the GO and KEGG pathway databases, respectively. The Venn diagram of protein alignment results between Nr, KOG, KEGG, SwissPort, and InterPro indicated that 36,283 isoforms were annotated in these databases. Using all transcripts, we predicted the CDS for 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 extends our understanding of the complex molecular mechanisms behind the interactions between BYDV-GAV and *P. huashanica*.

## Methods

### Sample preparation and RNA extraction

*P. huashanica* was kindly provided by Professor Jinxue Jing from the North West Agriculture and Forestry University, China. To obtain a good representation of the *P. huashanica* transcriptome, roots, stems, buds, and leaves (both inoculated and uninoculated leaves) were harvested from five independent plants at various development stages. Total RNA was extracted from each sample following the protocol described by Furtado (2014), employing a Trizol kit (Invitrogen) and a Qiagen RNeasy Plant Minikit (Qiagen), and subsequently pooled the RNA in equal amounts.

For Illumina RNA-Seq data analysis, *P. huashanica* were inoculated with viruliferous aphids carrying BYDV-GAV at the three- to four-leaf stage, 5 aphids per leaf, with non-viruliferous aphids-inoculated plants as mock inoculation. The viruliferous and non-viruliferous aphids were raised on susceptible wheat line 7182. For viruliferous aphids-inoculated treatments, the samples were collected at 3, 7 and 14 dpi, respectively; for mock treatment, the samples collected at 3, 7 and 14 dpi were equally mixed to form a composite sample. Then, all these samples were immediately frozen in liquid nitrogen for over 10 min and stored at  $-80^{\circ}\text{C}$  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 on an Agilent 2100 bioanalyzer (Agilent, USA). Only those samples with an RNA Integrity Number (RIN) value of more than 8.0 were retained for further use.

### Library preparation and PacBio sequencing

We prepared cDNA according to the PacBio Iso-seq protocol. Total RNA was used to prepare the first- and second-strand cDNA by 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 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

From the libraries produced by the Pacific Biosciences Sequel, we processed raw data using the SMRT analysis pipeline. After removing the redundant subreads, we identified insert fragments from the consistent sequences. The sequences achieved in this step were 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 tail were defined as full-length reads (Pacific Biosciences). The full-length non-chimeric sequences and non-full-length sequences were retained for further analysis.

This classified ROI was used to predict de novo consensus isoforms by performing an Iterative Clustering for Error Correction (ICE) algorithm. Then, we polished the predicted consensus isoforms using the Quiver algorithm to finally obtain the full-length polished consensus sequences. Depending on the QV, which indicates how confident the consensus calls are, the algorithm classified the polished isoforms into high-QV (expected accuracy  $\geq 99\%$ ) and low-QV isoforms, and only the 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 the final full-length isoforms (Fu et al. 2012).

### Illumina RNA-seq and PacBio data correction

Total RNA was digested with DNase I and the Ribo-Zero™ rRNA Removal Kit (Illumina, USA) was used for removing 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 to carry out end repair. The cDNA fragments with adapters were amplified by PCR, and the products were purified using Ampure XP Beads. We measured the purity and quality of the libraries using an Agilent 2100 Bioanalyzer and Qubit 2.0. BGI (BGI-Shenzhen, China) conducted the PCR amplification and

sequencing of the twelve libraries on the Illumina HiSeq platform. 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 reads generated from the Illumina HiSeq X-ten platform were used to correct the PacBio isoforms. Then, clean reads from the NGS were delivered to Proofread version 2.14.0 with the default parameters to correct single-bases and trim the chimera, and low-quality regions.

#### Identification of coding RNAs

We used three analysis tools: CPC (Kong et al. 2007), txCDsPredict, and CNCI (Sun et al. 2013), as well as a protein database, Pfam (Finn et al. 2016), to distinguish between coding and non-coding sequences based on their coding potential. Three analysis tools marked the code capacity of isoforms and set a threshold value to distinguish lncRNAs from mRNAs. Isoforms that aligned with the Pfam database were identified as mRNA; others were considered lncRNAs. We confirmed that the transcript was mRNA or lncRNA based on at least three of the four judgment results being consistent, lncRNAs 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 in Nr, Nt, SwissProt, InterPro, KOG, GO and KEGG databases with a threshold E-value  $\leq 10^{-5}$ . The annotation of Nt, Nr, KOG, KEGG, and SwissProt was performed using BLAST version 2.2.23 (Altschul et al. 1990), GO using Blast2Go version 2.5.0 (Conesa et al. 2005), and InterPro using InterProScan5 version 5.11–51.0 (Quevillon et al. 2005). GO terms were enriched and plotted by WEGO (Ye et al. 2006). KEGG pathway mapping was performed on the KEGG Automatic Annotation Server (KAAS) v2.0 (Moriya et al. 2007). Venn diagrams of the Nr, KEGG, SwissProt, and InterPro annotation results were drawn using the online Venn tool.

#### Time-series and transcription factor expression analyses

Time-series cluster analyses, based on the STEM method, were conducted using STEM software (Ernst and Bar-Joseph 2006). This was used to identify the global trends and similar temporal model patterns of the expression of the total DEGs. The R package “Mfuzz” (version = 3.8) was used to draw the cluster patterns of all expressed genes (Kumar and Futschik 2007). The prediction and classification of plant TFs and transcriptional regulators from DEGs were performed using a web-based tool iTAK, Plant Transcription Factor & Protein Kinase Identifier, and Classifier (<http://itak.feilab.net/cgi-bin/itak/index.cgi>) (Zheng et al. 2016).

#### RT-qPCR validation for DEGs

We performed RT-qPCR with the same plant materials at the same sampling time point to validate the RNA-Seq data. Total RNA extracted using a Plant Total RNA Extraction Kit (Bioer Technology, Hangzhou) from three inoculated replications was used to examine the expression patterns of the selected genes. Total RNA (1  $\mu$ g) was reverse transcribed with M-MLV Reverse Transcriptase (Promega, USA), and the resulting cDNA samples were diluted two-fold. For RT-qPCR, quantitative assays were conducted on each cDNA dilution using an UltraSYBR Mixture (Cwbio, China) and a Roche Light-Cycler480 Real-time PCR Detection System instrument (Roche, Switzerland).

We selected twenty genes that were differentially expressed in the RNA-seq data (i.e. > 2-fold change,  $P < 0.05$ ) for RT-qPCR analysis to validate the accuracy of the RNA-seq results. The gene-specific primer pairs for each gene were designed using Primer3web version 4.0.0 ([http://primer3plus.com/primer3web/primer3web\\_input.htm](http://primer3plus.com/primer3web/primer3web_input.htm)). The *P. huashanica* housekeeping gene 18S rRNA was used to normalize all data (a list of these genes primer sequences is listed in Additional file 1: Table S5). The relative expression levels of DEGs were normalized using the  $2^{-\Delta\Delta C_t}$  method (Schmittgen and Livak 2008). All RT-qPCR reactions were repeated in three biological and three technical replicates, and the means and corresponding standard errors were calculated.

#### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s42483-020-00057-8>.

**Additional file 1: Table S1.** Summary of PacBio Polymerase Reads Subreads. **Table S2.** Summary statistics of the trimmed and error-corrected PacBio isoforms by RNA-seq dataset. **Table S3.** Summary statistics of RNA-seq raw reads and clean reads from twelve libraries. **Table S4.** Summary statistics of the identification of transcription factors and transcriptional regulators from RNA-seq. **Table S5.** Primers used in RT-qPCR for validation of RNA-seq data.

**Additional file 2: Figure S1.** ROI length distribution of each library. The X-axis shows the lengths of reads calculated in our library, and the Y-axis shows the number of reads. **Figure S2.** Pie charts show the classification summary of ROI reads for each library. **Figure S3.** Length distribution of the final consensus isoforms from merged libraries. The X-axis shows the lengths of isoforms, and the Y-axis shows the number of isoforms. **Figure S4.** Venn diagram of annotation between the Nr, KOG, KEGG, SwissProt, and Interpro databases. **Figure S5.** Volcano plots of DEGs in BYDV-GAV-infected *P. huashanica* at 3, 7, and 14 dpi, compared to mock-inoculated samples. **a** Functional distribution of the KOG annotations. **b** Functional distribution of the GO annotations. **c** Functional distribution of the KEGG annotations. The X-axis represents the number of transcripts, the Y-axis represents the functional category.

#### Abbreviations

CDPK: Calcium-dependent protein kinase-related genes; CRPK: Cysteine-rich receptor-like protein kinase; DEGs: Differentially-expressed genes; GISH: Genomic in situ hybridization; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: Eukaryotic Orthologous Groups; MAPK: Mitogen-activated protein kinase; PR: Pathogenesis-related protein;

SGS: Second-generation sequencing; SMRT: Single-molecule real-time; WAK: Wall-associated receptor kinase

## Acknowledgments

Not applicable.

## Authors' contributions

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

## Funding

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

## 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.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

Received: 26 March 2020 Accepted: 16 June 2020

Published online: 30 July 2020

## References

- Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, et al. A survey of the sorghum transcriptome using single-molecule long reads. *Nat Commun*. 2016;7:11706.
- Ali N, Heslop-Harrison JSP, Ahmad H, Graybosch RA, Hein GL, Schwarzbacher T. Introgression of chromosome segments from multiple alien species in wheat breeding lines with wheat streak mosaic virus resistance. *Heredity*. 2016;117:114–23.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
- Amorim LLB, da Fonseca Dos Santos R, Neto JPB, Guida-Santos M, Crovella S, Benko-Iseppon AM. Transcription factors involved in plant resistance to pathogens. *Curr Protein Pept Sci*. 2017;18(4):335–51.
- An D, Cao HX, Li C, Humbeck K, Wang W. Isoform sequencing and state-of-art applications for unravelling complexity of plant transcriptomes. *Genes*. 2018;9:43.
- Au KF, Underwood JG, Lee L, Wong WH. Improving PacBio long read accuracy by short read alignment. *PLoS One*. 2012;7:e46679.
- Baden C. A taxonomic revision of *Psathyrostachys* (Poaceae). *Nord J Bot*. 1991;11(1):3–26.
- Banks PM, Davidson JL, Bariana H, Larkin PJ. Effects of barley yellow dwarf virus on the yield of winter wheat. *Aust J Agr Res*. 1995;46:935–46.
- 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.
- Birkenbihl RP, Liu S, Somssich IE. Transcriptional events defining plant immune responses. *Curr Opin Plant Biol*. 2017;38:1–9.
- Buermans HP, den Dunnen JT. Next generation sequencing technology: advances and applications. *Biochim Biophys Acta*. 2014;1842:1932–41.
- Chen C, Chen Z. Isolation and characterization of two pathogen- and salicylic acid-induced genes encoding WRKY DNA-binding proteins from tobacco. *Plant Mol Biol*. 2000;42:387–96.
- Chen S, Zhang A, Fu J. The hybridization between *Triticum aestivum* and *Psathyrostachys huashanica* (in Chinese). *J Genet Genomics*. 1991;18(6):508–12 <http://www.jgenetgenomics.org/EN/abstract/abstract18258.shtml>.
- Cheng B, Furtado A, Henry RJ. Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. *GigaScience*. 2017;6:1–13. <https://doi.org/10.1093/gigascience/gix086>.
- Chiang YH, Coaker G. Effector triggered immunity: NLR immune perception and downstream defense responses. *The Arabidopsis Book*. 2015;13:e0183. <https://doi.org/10.1199/tab.0183>.
- 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;21:3674–6.
- Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet*. 2010;11:539–48.
- Domier LL. Family *Luteoviridae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, editors. *Virus taxonomy*, ninth report of the international committee on taxonomy of virus. London: Elsevier Academic Press; 2012. p. 1045–53.
- Du W, Wang J, Lu M, Sun S, Chen X, Zhao J, et al. Characterization of a wheat-*Psathyrostachys huashanica* Keng 4Ns disomic addition line for enhanced tiller numbers and stripe rust resistance. *Planta*. 2014a;239:97–105.
- Du W, Wang J, Pang Y, Li Y, Chen X, Zhao J, et al. Isolation and characterization of a *Psathyrostachys huashanica* Keng 6Ns chromosome addition in common wheat. *PLoS One*. 2013a;8:e53921.
- Du W, Wang J, Pang Y, Wang L, Wu J, Zhao J, et al. Isolation and characterization of a wheat-*Psathyrostachys huashanica* 'Keng' 3Ns disomic addition line with resistance to stripe rust. *Genome*. 2014b;57:37–44.
- Du W, Wang J, Wang L, Wu J, Zhao J, Liu S, et al. Molecular characterization of a wheat-*Psathyrostachys huashanica* Keng 2Ns disomic addition line with resistance to stripe rust. *Mol Genet Genomics*. 2014c;289:735–43.
- Du W, Wang J, Wang L, Zhang J, Chen X, Zhao J, et al. Development and characterization of a *Psathyrostachys huashanica* Keng 7Ns chromosome addition line with leaf rust resistance. *PLoS One*. 2013b;8:e70879.
- Ernst J, Bar-Joseph Z. Stem: a tool for the analysis of short time series gene expression data. *BMC Bioinf*. 2006;7:191.
- Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44:D279–85.
- Fischer U, Dröge-Laser W. Overexpression of *NtERF5*, a new member of the tobacco ethylene response transcription factor family enhances resistance to *Tobacco mosaic virus*. *Mol Plant Microbe Interact*. 2004;17:1162–71.
- Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2.
- Furtado A. RNA extraction from developing or mature wheat seeds. In: Henry R, Furtado A, editors. *Cereal genomics. Methods in molecular biology*, vol. 1099. Totowa: Humana Press; 2014. p. 23–8.
- Gouveia BC, Calil IP, Machado JPB, Santos AA, Fontes EPB. Immune receptors and co-receptors in antiviral innate immunity in plants. *Front Microbiol*. 2016;7:2139.
- Guo Z, Li Y, Ding SW. Small RNA-based antimicrobial immunity. *Nat Rev Immunol*. 2019;19:31–44.
- He X, Chen Z, Wang J, Li W, Zhao J, Wu J, et al. A sucrose: fructan-6-fructosyltransferase (*6-SFT*) gene from *Psathyrostachys huashanica* confers abiotic stress tolerance in tobacco. *Gene*. 2015;570:239–47.
- Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, et al. 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;18:395.
- Jacob F, Vernaldi S, Maekawa T. Evolution and conservation of plant NLR functions. *Front Immunol*. 2013;4:297.
- Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, et al. Improved maize reference genome with single-molecule technologies. *Nature*. 2017;546:524–7.
- 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 C Life Sci*. 2004;47:175–82.
- Kang HY, Zhang ZJ, Xu LL, Qi WL, Tang Y, Wang H, et al. Characterization of wheat-*Psathyrostachys huashanica* small segment translocation line with enhanced kernels per spike and stripe rust resistance. *Genome*. 2016;59:221–9.
- Kohorn BD, Kohorn SL. The cell wall-associated kinases, WAKs, as pectin receptors. *Front Plant Sci*. 2012;3:88.
- 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:W345–9.
- Kumar L, Futschik ME. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics*. 2007;21(1):5–7.

- Li J, Harata-Lee Y, Denton MD, Feng Q, Rathjen JR, Qu Z, et al. Long read reference genome-free reconstruction of a full-length transcriptome from *Astragalus membranaceus* reveals transcript variants involved in bioactive compound biosynthesis. *Cell Discov*. 2017a;3:17031.
- Li X, Fan S, Hu W, Liu G, Wei Y, He C, et al. Two cassava basic leucine zipper (bZIP) transcription factors (MebZIP3 and MebZIP5) confer disease resistance against cassava bacterial blight. *Front Plant Sci*. 2017b;8:2110.
- Li X, Kapos P, Zhang Y. NLRs in plants. *Curr Opin Immunol*. 2015;32:114–21.
- Liu N, Hake K, Wang W, Zhao T, Romeis T, Tang D. CALCIUM-DEPENDENT PROTEIN KINASE5 associates with the truncated NLR protein TIR-NBS2 to contribute to *exo70B1*-mediated immunity. *Plant Cell*. 2017b;29:746–59.
- Liu X, Mei W, Soltis PS, Soltis DE, Barbazuk WB. Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. *Mol Ecol Resour*. 2017a;17:1243–56.
- Liu Y, Khine MO, Zhang P, Fu Y, Wang X. Incidence and distribution of insect-transmitted cereal viruses in wheat in China from 2007 to 2019. *Plant Dis*. 2020;104:1407–14.
- Lu B. Diversity and conservation of *Triticum* genetic resource. *Biodivers Sci*. 1995;3:63–8 (in Chinese). <http://www.biodiversity-science.net/CN/10.17520/biods.1995010>.
- Lu D, Wu S, Gao X, Zhang Y, Shan L, He P. A receptor-like cytoplasmic kinase, BLK1, associates with a flagellin receptor complex to initiate plant innate immunity. *Proc Natl Acad Sci U S A*. 2010;107:496–501.
- Luo Y, Ding N, Shi X, Wu Y, Wang R, Pei L, et al. Generation and comparative analysis of full-length transcriptomes in sweetpotato and its putative wild ancestor *I. trifida*. *BioRxiv*. 2017. <https://doi.org/10.1101/112425>.
- Maekawa T, Kufer TA, Schulze-Lefert P. NLR functions in plant and animal immune systems: so far and yet so close. *Nat Immunol*. 2011;12:817–26.
- Mahmoud M, Zywicki M, Twardowski T, Karlowski WM. Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. *Genomics*. 2017;111:43–9.
- Meng X, Zhang S. MAPK cascades in plant disease resistance signaling. *Annu Rev Phytopathol*. 2013;51:245–66.
- Miller WA, Rasochova L. Barley yellow dwarf viruses. *Annu Rev Phytopathol*. 1997;35:167–90.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAA5: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007;35:W182–5.
- Nevo E, Chen G. Drought and salt tolerances in wild relatives for wheat and barley improvement. *Plant Cell Environ*. 2010;33:670–85.
- Pootakham W, Sonthirod C, Naktang C, Ruang-Areerate P, Yoocha T, Sangsrakru D, et al. De novo hybrid assembly of the rubber tree genome reveals evidence of paleotetraploidy in *Hevea* species. *Sci Rep*. 2017;7:41457.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. *Nucleic Acids Res*. 2005;33:W116–20.
- Rayapuram C, Jensen MK, Maiser F, Shanir JV, et al. Regulation of basal resistance by a powdery mildew-induced cysteine-rich receptor-like protein kinase in barley. *Mol Plant Pathol*. 2012;13:135–47.
- Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*. 2015;13:278–89.
- Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. *Genome Biol*. 2013;14:A05.
- Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative  $C_T$  method. *Nat Protoc*. 2008;3:1101–8.
- Song S, Tao Y, Zhang H, Wu Y. *Psathyrostachys huashanica*, a potential resource for resistance to barley yellow dwarf virus-GAV. *Eur J Plant Pathol*. 2013;137:217–21.
- Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41:e166.
- Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun*. 2016;7:11708.
- Wang J, Yao L, Li B, Meng Y, Ma X, Wang H. Single-molecule long-read transcriptome dataset of halophyte *Halogeton glomeratus*. *Front Genet*. 2017;8:197.
- Wang L, Liu Y, Du W, Jing F, Wang Z, Wu J, et al. Anatomy and cytogenetic identification of a wheat-*Psathyrostachys huashanica* Keng line with early maturation. *PLoS One*. 2015;10:e0131841.
- Wang M, Shang H. Evaluation of resistance in *Psathyrostachys huashanica* to wheat take-all fungus. *J Northwest Agric Univ*. 2000;28(6):69–71 (in Chinese). [http://www.xnxbz.net/xnblkjdxr/ch/reader/view\\_abstract.aspx?file\\_no=200006230&flag=1](http://www.xnxbz.net/xnblkjdxr/ch/reader/view_abstract.aspx?file_no=200006230&flag=1).
- Wang M, Wang P, Liang F, Ye Z, Li J, Shen C, et al. A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytol*. 2018;217:163–78.
- Wang Y, Xie Q, Yu K, Poysa V, Lin L, Kang H, et al. Development and characterization of wheat-*Psathyrostachys huashanica* partial amphiploids for resistance to stripe rust. *Biotechnol Lett*. 2011a;33(6):1233–8.
- Wang Y, Yu K, Xie Q, Kang H, Lin L, Fan X. Cytogenetic, genomic in situ hybridization (GISH) and agronomic characterization of alien addition lines derived from wheat-*Psathyrostachys huashanica*. *Afr J Biotechnol*. 2011b;10(12):2201–11.
- Warburton ML, Crossa J, Franco J, Kazi M, Trethowan R, Rajaram S, et al. Bringing wild relatives back into the family: recovering genetic diversity in CIMMYT improved wheat germplasm. *Euphytica*. 2006;149:289–301.
- Wulff BBH, Moscou MJ. Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Front Plant Sci*. 2014;5:692.
- Xu Q, Zhu J, Zhao S, Hou Y, Li F, Tai Y, et al. 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.
- Yang P, Chen C, Wang Z, Fan B, Chen Z. 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*. 1999;18:141–9.
- Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*. 2006;34:W293–7.
- Yoshii M, Shimizu T, Yamazaki M, Higashi T, Miyao A, Hirochika H, et al. Disruption of a novel gene for a nac-domain protein in rice confers resistance to rice dwarf virus. *Plant J*. 2009;57:615–25.
- Yu X, Feng B, He P, Shan L. From chaos to harmony: responses and signaling upon microbial pattern recognition. *Annu Rev Phytopathol*. 2017;55:109–37.
- 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: Periannan S, editor. *Wheat rust diseases. Methods in molecular biology*, vol. 1659. New York: Humana Press; 2017. p. 163–72.
- Zhang Z, Lin Z, Xin Z. Research progress in BYDV resistance genes derived from wheat and its wild relatives. *J Genet Genomics*. 2009;36:567–73.
- Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, et al. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol Plant*. 2016;9:1667–70.
- Zhou GH, Cheng ZM, Qian YT, Zhang XC, Rochow WF. Serological identification of Luteoviruses of small grains in China. *Plant Dis*. 1984;68:710–3.
- 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*. 2017a;6:1–7. <https://doi.org/10.1093/gigascience/gix097>.
- Zimin AV, Stevens KA, Crepeau MW, Puiu D, Wegryn JL, Yorke JA, et al. An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *GigaScience*. 2017b;6(1):1–4. <https://doi.org/10.1093/gigascience/giw016>.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

