#### SUPPORTING INFORMATION FOR

## **Integrating transcriptome and proteome analysis to determine the lignin synthesis pathway involved in *Panax notoginseng* in fungal stress**

Qian Yang1, Jianbin Li1, Lin Xiao1, Jiani Peng1, Jialing Sun1, Xiuming Cui 1\*

1Laboratory of Sustainable Utilization of *Panax notoginseng* Resources, State Administration of Traditional Chinese Medicine, Key Laboratory of *Panax notoginseng* in Yunnan Province, *Panax notoginseng* Research Institute in Yunnan Province, Faculty of Life Science and Technology, Kunming University of Science and Technology, 650500, Kunming, Yunnan, China

Correspondence authors: Xiuming Cui (sanqi37@vip.sina.com)

#### List of Supporting information

Supplementary information includes 2 methods and 6 tables.

**Supplementary Methods S1** Transcriptome analysis

**Supplementary Methods S2** Proteome analysis

**Supplementary Table S1** Primer information of related genes in the lignin synthesis pathway.

**Supplementary Table S2** All transcripts identified from the leaves of *P. notoginseng* induced by *Alternaria panax*.

**Supplementary Table S3** Significantly different transcripts in *P. notoginseng* leaves induced by *Alternaria panax*.

**Supplementary Table S4** All proteins identified from the leaves of *P. notoginseng* induced by *Alternaria panax*.

**Supplementary Table S5** Significantly different proteins in *P. notoginseng* leaves induced by *Alternaria panax*.

**Supplementary Table S6** Differential genes/proteins screened out in transcript/protein correlation analysis.

#### Methods S-1

**Transcriptome analysis**

**RNA extraction and library preparation**

The leaves collected from Inf and Con were immediately frozen in liquid nitrogen. The samples were stored at -80°C until RNA was extracted. According to the manufacturer's protocol, total RNA was extracted from the leaves using Trizol reagent (TaKaRa). Based on the ratio of the optical density at 260 nm to the optical density at 280 nm (OD260/280), the integrity of RNA was checked using an ultraviolet spectrophotometer (Hoefer, MA, USA), and evaluated by electrophoresis in denatured formaldehyde. Agilent 2100 Bio-analyzer (Agilent Technologies, Santa Clara, CA, United States) is then used for verification. The Poly (A) mRNA was isolated from total RNA using magnetic beads with Oligo (dT) (Qiagen, Germany), and then fragmented into short fragments by adding fragmentation buffer. The first cDNA was synthesized using random hexamers, and then the second cDNA was synthesized using RNase H and DNA polymerase I. The QIA rapid PCR extraction kit was used to purify the cDNA fragments, and then used EB buffer for end repair and poly (A) addition, and connected to the sequencing connector. After agarose gel electrophoresis and extraction of cDNA from the gel, the cDNA fragments were purified and enriched by PCR to construct the final RNA-Seq library, which was sequenced on the Illumina HiSeq X Ten platform using paired-end mode.

**Sequence analysis**

The original RNA-Seq reads were filtered using trimmomatic ([Lohse et al. 2012](#_ENREF_22" \o "Lohse, 2012 #44)) to remove low quality sequences, reads with more than 5% N bases (bases unknown) and reads containing linker sequences. All downstream analysis is based on high-quality clean reads. Then, the index of *P. notoginseng* reference genome was constructed using Hista2 (NCBI Sequence ReadArchive, project ID SRP091819) ([Chen et al. 2017](#_ENREF_6" \o "Chen, 2017 #45)),and the cleaned reads were compared with the reference genome using Hista2 (Broad Institute, Boston, Massachusetts) ([Trapnell et al. 2009](#_ENREF_33" \o "Trapnell, 2009 #46)). Next, a StringTie count was used to map reads to each gene. Fragments per kilobase of exon model per million mapped fragments (FPKM) of each gene was calculated based on the length of the gene, and reads count was mapped to the gene ([Trapnell et al. 2012](#_ENREF_34" \o "Trapnell, 2012 #47)).

**Screening, clustering and functional annotation of differentially expressed genes (DEGs)**

DEGs were identified using the edgeR package ([Robinson et al. 2010](#_ENREF_30" \o "Robinson, 2010 #48)), and the original p values were corrected using the error detection rate (FDR) for multiple tests. The corrected P value <0.05 and the fold change | log2 |> 1 was used as the screening cut-off value for the significantly different gene expression between Con and Inf.

The basic local alignment search tool (BLAST) was used to identify significant homology between genes and known gene products. The BLAST search results were then classified based on gene function to identify DEGs between Con and Inf. For gene ontology (GO ) annotation ([Ashburner et al. 2000](#_ENREF_3" \o "Ashburner, 2000 #49)) and KEGG pathway enrichment analysis, DEGs were mapped to terms in Kyoto encyclopedia of genes and genomes (KEGG) and GO databases using KOBAS 3.0 (KEGG atom-based annotation system) ([Kanehisa et al. 2007](#_ENREF_17" \o "Kanehisa, 2007 #50)). The KEGG pathway with corrected p-value ≤0.05 was considered statistically significant. And through to the Nt and Nr database (NCBI) and KAAS (http://www.genome.jp/tools/kaas/) for blasting to annotate DEGs.

**Methods S-2**

**Proteome analysis**

**Lable-free proteomics sample preparation**

The samples were frozen in liquid nitrogen, and then ground into powder with a mortar. Add 5 volumes of TCA / acetone (volume ratio 1: 9) to the powder and mix by vortexing. The mixture was placed at -20 ° C for 4h, then centrifuged at 4 ° C, 6000 rpm for 40 min. After discarding the supernatant, pre-cooled acetone was added and washed three times, and then the precipitate was air-dried. 30 volumes of SDT (4% SDS, 100 mM DTT, 150 mM Tris-HCl pH 8.0) buffer was added to about 30 mg of powder, mixed and boiled for 5 min. The lysate was ultrasound and then boiled for another 15 min. After centrifugation at 14000g for 40 min, the supernatant was filtered with a 0.22 μm filter. Finally, after quantifying the filtrate with BCA protein assay kit (Solarbio, China), the samples were stored at -80 ° C until use.

**Peptide digestion**

Approximately 200 μg of protein was taken from each sample for trypsin digestion using the filter aided proteome preparation (FASP) method ([Wiśniewski et al. 2009](#_ENREF_36" \o "Wiśniewski, 2009 #51)), and digested in 40 μl of 25 mM NH4HCO3 buffer at 37 ° C overnight to collect the filtrate of the resulting peptide. Then use C18 Cartridges to desalt the peptides (Empore ™ SPE Cartridges C18 (standard density), bed ID 7 mm, volume 3 ml, Sigma). After the peptides were lyophilized, in 40 μl of 0.1% (v/v) formic acid Re-dissolve in the solution, and finally determine the content of the peptides (OD280).

**LC-MS / MS data processing and analysis**

Each sample was separated using the HPLC liquid phase system Easy nLC (Thermo Fisher Scientific) with a nanoliter flow rate. The peptide mixture was added to the nanoViper C18 liquid chromatography Column (Thermo Scientific Acclaim map100,100 m\* 2cm). In the buffer A (0.1% formic acid), using a linear gradient buffer B (84% acetonitrile and 0.1% formic acid) was separated by the C18-A2 analytical Column (Thermo Scientific Easy Column, 10cm, ID75 m, 3 m) at the flow rate of 300nl/min. Perform a 2-hour linear gradient: 0-55% buffer B for 110 min, 55-100% buffer B for 5 min, and remain in 100% buffer B for 5 min.

After chromatographic separation of the samples, LC-MS / MS analysis was performed by a Q-Exactive mass spectrometer (Thermo Scientific), which operated in positive ion mode. The specific parameters are set as follows: first-level mass spectrometry scan range: 300-1800 m/z, resolution is 70,000 (200 m/z), automatic gain control (AGC) target value is 1.0 × 106, the maximum ion implantation time (Maximum IT) is set to 50 ms, and the dynamic exclusion time is set to 60s. After each full MS scan, 20 fragment maps (MS2 scan) were collected, and a secondary mass spectrum was obtained with the resolution of m/z 200 of 17,500. MS2 activation type was higher energy collision dissociation (HCD), and the isolation width was 2m/z. The normalized collision energy is 30 eV, and the unspecified filling ratio (which specifies the minimum percentage of the target value that may be reached at the maximum filling time) is defined as 0.1%. MaxQuant software version 1.5.3.17 (Max Planck Institute of Biochemistry in Martinsried, Germany) ([Cox and Mann 2008](#_ENREF_10" \o "Cox, 2008 #52)) was used to analyze MS data, and Label Free Quantitation (LFQ) ([Cox et al. 2014](#_ENREF_9" \o "Cox, 2014 #1)) was used for protein quantitative analysis. Search using the following settings based on the *P. notoginseng* transcriptome database (http://www.ncbi.nlm.nih.gov/sra. accession number: PRJNA706472). The following parameters were used to identify the protein: the maximum number of missed cleavage sites allowed by trypsin was 2, carbamoylmethyl was used as the fixed modification, and oxidation was used as the variable modification. The charge states of the peptides are + 1, + 2 and + 3. For all obtained MS1 mass spectra, the peptide mass tolerance was 20 ppm, and for all obtained MS2 mass spectra, the mass tolerance was 20 ppm. The false discovery rate (FDR) is set to ≤0.01. Using LFQ for protein quantification, a fold change of ≥ 2 or ≤ 0.5 were considered significantly up-regulated or significantly down-regulated, respectively. Intensities <1×108 (scientific notation) and unidentified proteins were considered to have no significant changes. The pvclust R package was used for hierarchical clustering analysis.

**GO annotation and KEGG pathway annotation**

The same method described in the above transcriptome method was used to identify GO annotations and the KEGG pathway enriched in DEGs. GO annotation and KEGG pathway with corrected P value <0.05 were considered to be significantly enriched in differentially expressed proteins (DEPs).

In order to further explore the effect of DEPs on response to stress and to discover the internal relationship between DEPs, an enrichment analysis was performed. GO classification and KEGG pathway enrichment analysis were based on Fisher's exact test, and the entire quantitative protein annotation was regarded as a background data set.