RESEARCH ARTICLE

# Transcriptome Analysis of *Beta macrocarpa* and Identification of Differentially Expressed Transcripts in Response to *Beet Necrotic Yellow Vein Virus* Infection

Huiyan Fan[1,2], Yongliang Zhang[1], Haiwen Sun[1], Junying Liu[1], Ying Wang[1], Xianbing Wang[1], Dawei Li[1], Jialin Yu[1], Chenggui Han[1] *

**1** State Key Laboratory for Agrobiotechnology and Department of Plant Pathology, China Agricultural University, Beijing, 100193, China, **2** College of Pharmacy, Zhejiang Chinese Medicine University, Hangzhou, 310053, Zhejiang, China

* hanchenggui@cau.edu.cn

## Abstract

### Background

Rhizomania is one of the most devastating diseases of sugar beet. It is caused by *Beet necrotic yellow vein virus* (BNYVV) transmitted by the obligate root-infecting parasite *Polymyxa betae*. *Beta macrocarpa*, a wild beet species widely used as a systemic host in the laboratory, can be rub-inoculated with BNYVV to avoid variation associated with the presence of the vector *P. betae*. To better understand disease and resistance between beets and BNYVV, we characterized the transcriptome of *B. macrocarpa* and analyzed global gene expression of *B. macrocarpa* in response to BNYVV infection using the Illumina sequencing platform.

### Results

The overall *de novo* assembly of cDNA sequence data generated 75,917 unigenes, with an average length of 1054 bp. Based on a BLASTX search (E-value $\leq 10^{-5}$) against the non-redundant (NR, NCBI) protein, Swiss-Prot, the Gene Ontology (GO), Clusters of Orthologous Groups of proteins (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, there were 39,372 unigenes annotated. In addition, 4,834 simple sequence repeats (SSRs) were also predicted, which could serve as a foundation for various applications in beet breeding. Furthermore, comparative analysis of the two transcriptomes revealed that 261 genes were differentially expressed in infected compared to control plants, including 128 up- and 133 down-regulated genes. GO analysis showed that the changes in the differently expressed genes were mainly enrichment in response to biotic stimulus and primary metabolic process.

## Conclusion

Our results not only provide a rich genomic resource for beets, but also benefit research into the molecular mechanisms of beet- BNYVV interaction.

## Introduction

Rhizomania is a major threat to sugar beet (*Beta vulgaris*) production, through reductions in both crop yield and beet sugar content, and is distributed worldwide across regions depending on cultivation conditions [1,2]. It is caused by *Beet necrotic yellow vein virus* (BNYVV), which is transmitted by the obligate root-infecting parasite *Polymyxa betae* in a persistent manner because of resting spores retaining the virus [3]. BNYVV has a multipartite RNA genome, the larger RNA1 and RNA2 contain the housekeeping genes of the virus and are required for infection, whereas the smaller RNAs are involved in pathogenicity and vector transmission [4–6]. Up to now, the major focus in study of the host–pathogen interaction with BNYVV has been on the virus itself, and there have been very few studies of the host genomic response. Recently, gene expression profiling using Illumina RNA-Seq revealed 3,016 differentially expressed genes during BNYVV infection in the experimental host *Nicotiana benthamiana*, which provided a list of candidate genes involved in resistance to BNYVV infection [7]. However, RNA-Seq technology has not been used to analyze the natural hosts' genomic response to BNYVV infection.

Because of the lack of effective and acceptable chemical control methods for the vector, genetic resistance is the most promising approach for control of this disease [1]. A number of different sources of partial resistance to BNYVV have been developed and these have performed better than susceptible cultivars in the field [8–10]. However, this breeding success may be a short lived as the presence of virulent forms of BNYVV, such as the P-type, and resistance-breaking isolates suggest that further research is required to identity new sources of resistance for breeders to work with [11,12]. The recent genome sequencing of sugar beet provided a powerful tool to characterize the genes responsible for agronomic traits [13]. However, the transcriptome of only one species may not provide an understanding of gene/genome evolution, genome organization and genetic variations. Thus, sequencing the transcriptome of wild species should lead toward better understanding of evolutionary processes and knowledge about the impact of gene flow to domesticated plants [14]. In addition, wild species could be new sources of resistance genes. The Holly resistance source has been widely exploited in most current varieties. It is well known that BNYVV resistance genes are derived from wild beet *B. vulgaris* subsp. *maritime* accessions [8,15].

Very few genomic resources are available for wild beet relatives; and this includes *B. macrocarpa*, a widely used laboratory systemic host, which can be rub-inoculated with BNYVV, thereby avoiding variation associated with presence of the vector *P. betae*. Only nine related nucleotide sequences were available for *B. macrocarpa* in GenBank as of June 2014. Next generation sequencing technologies have enabled very rapid and cost-effective sequencing of transcriptomes. In this study, we took advantage of RNA-Seq to survey the foliar transcriptome of *B. macrocarpa* and identified a total 75,917 unigenes. Comparative analysis of the expression profiles of BNYVV-infected and non-infected *B. macrocarpa* indicated that BNYVV infection could potentially perturb primary metabolism of plants and activate immune regulatory systems. To our knowledge, this is the first report to define the *B. macrocarpa* transcriptome and

could enrich the genetic resources of beets and aid in discovery of genes related to the molecular mechanism of beet–BNYVV interactions.

## Results and Discussion

### Illumina paired-end sequencing and *de novo* assembly

Under greenhouse conditions, the leaves of *B. macrocarpa* inoculated with BNYVV showed chlorotic spots at 5–7 d, followed by systemic yellow mosaic with severe stunting after two weeks (Fig 1). To obtain an overview of *B. macrocarpa* gene expression profile and genes responding to virus in leaves, we prepared two respective RNA samples from the leaf pools of BNYVV-infected and non-infected plants for transcriptome analysis using the Illumina sequencing platform Hiseq2000.

Paired-end sequencing resulted in a total of 121,269,730 reads comprising 12,126,973,000 nucleotide bases. To improve quality of the dataset, high stringency filtering was performed, which included removal of reads containing adaptor and vector sequence as well as reads with 30% having base quality $\leq$ 20. There were 48,279,346 and 59,903,306 clean reads from BNYVV-infected and non-infected libraries (S1 Table). The raw reads are available in the Sequence Read Archive at the National Center for Biotechnology Information (NCBI) under accession number SRP033294. Using the SOAPdenovo assembly program, all high-quality reads were assembled into 44,626 and 44,903 contigs, with N50s (N50 represents median length of all contigs) of 545 and 520 bp, respectively. CAP3 software[16] allowed us to map the reads back to the contigs and connect the contigs into unigenes with the fewest Ns (N represents unknown residues between two contigs) and could not be extended on either end. Finally, we obtained a total of 75,917 unigenes, with a N50 length of 1867 bp and total length of 80.0 Mb (Table 1). The length distribution of non-redundant unigenes is shown in Table 2, implying that the Illumina sequencing solution was reliable.

### Similarity analysis and functional annotation

All unigenes were compared with the NCBI non-redundant (NR) protein and Swiss-Prot databases for functional annotation using BLASTX with an E-value threshold of $10^{-5}$ (E-value $\leq 10^{-5}$). Among the 75,917 unigenes from both the BNYVV-infected and non-infected libraries, 39,176 (51.60%) showed significant matches in the NR database. An additional 26,326 (34.68%) unigenes showed significant matches in the Swiss-Prot database (Table 3, S1 File). The species distribution of the best match result for each sequence is shown in Fig 2. The sequences of *B. macrocarpa* had a 36% match with grape (*Vitis vinifera*), followed by flowering peach (*Prunus persica*), white poplar (*Populus trichocarpa*) and castor (*Ricinus communis*) with 9%. In addition, there were 784 distinct sequences with the highest homology to genes from *B. vulgaris*, supporting the robustness and validity of our RNA-Seq based approach. The sequencing in this study provides more detail and general genetic data that will facilitate large-scale discovery and utilization of genetic resources for *B. macrocarpa*.

Based on NR annotation, 16,332 unigenes were assigned gene ontology (GO) terms. GO-annotated unigenes belonged to the cellular components, molecular function and biological processes clusters and were further divided into 39 subcategories. Cell or cell part, binding and metabolic processes were the vast majority of the categories from each GO cluster (Fig 3). This suggested that the life cycle of *B. macrocarpa* is prominently governed by genes related to cellular structure, molecular interaction and metabolism In contrast, among the subcategories with the fewest members were 'extracellular region part 'of the cellular components ontology, 'nutrient reservoir' of the molecular function ontology and 'developmental process' of the biological processes ontology.
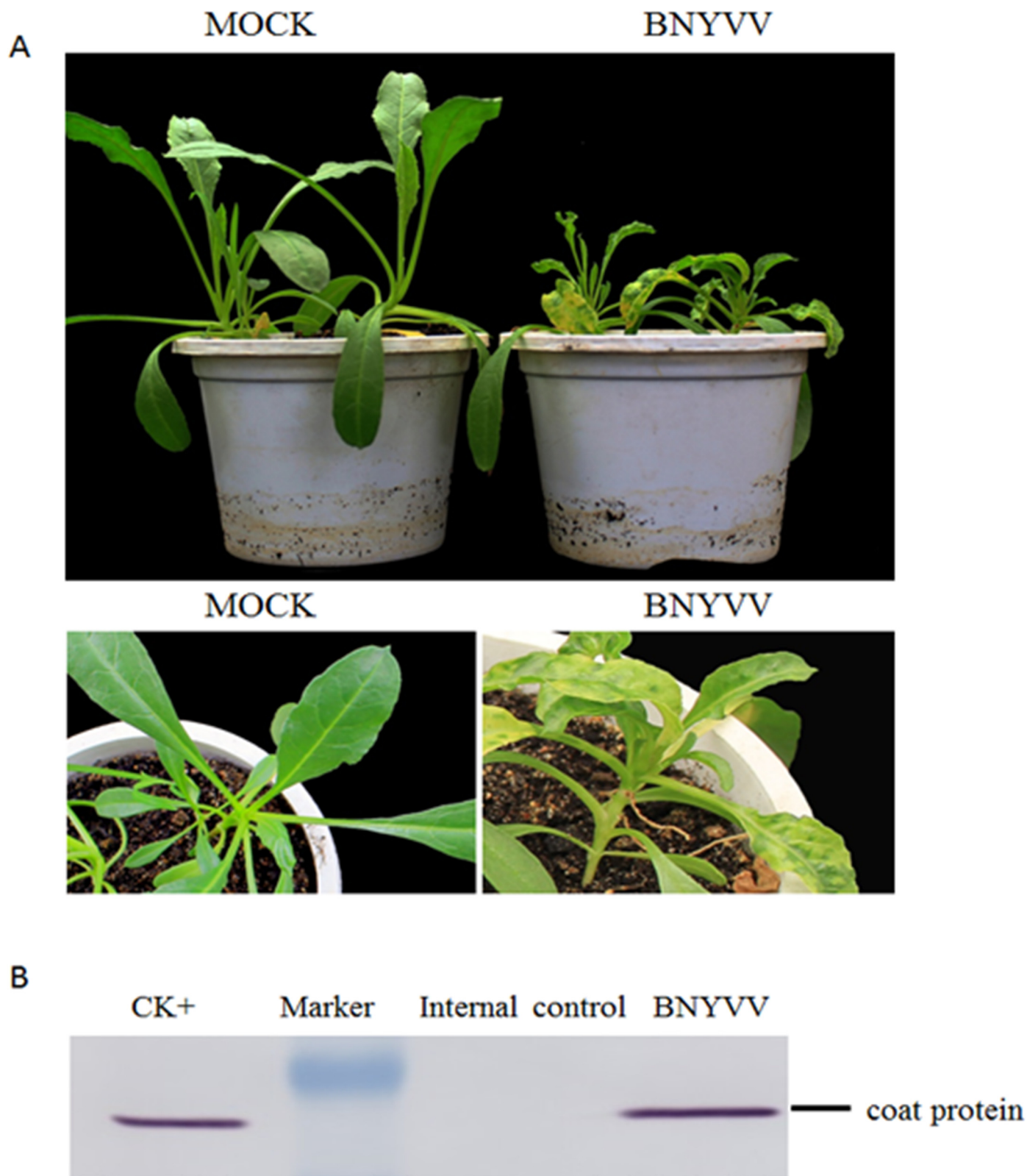
**Fig 1. BNYVV phenotypes and detection of virus in systemically infected *B. macrocarpa* leaves.** (A) Systemic symptoms induced in *B. macrocarpa* plants by BNYVV. Mock-infection was plants rubbed with buffer. Mock and BNYVV (B) Western blotting for analysis of BNYVV CP expression. BNYVV-infected leaves of *Tetragonia expansa* were the positive control (CK+ lane) and healthy *B. macrocarpa* leaves (mock) were the negative control.

doi:10.1371/journal.pone.0132277.g001

**Table 1. Overview of the sequencing and assembly.**

| Statistics of data production | Mock | BNYVV infected |
|---|---|---|
| Clean reads | 48,279,346 | 59,903,306 |
| Clean bases (bp) | 4,693,645,595 | 5,830,613,062 |
| Contigs | | |
| No. of contigs | 44,626 | 44,903 |
| Total nucleotides(nt) in contigs | 21,231,334 | 20,853,654 |
| Length of N50 (bp) | 545 | 520 |
| Average length of contigs (bp) | 476 | 464 |
| GC percentage (%) | 41.03 | 41.00 |
| All-Unigenes | | |
| No. of All-unigenes | 75,917 | |
| Total nucleotides(nt) in All-unigenes | 79,982,383 | |
| Length of N50 (bp) | 1,867 | |
| Average length of unigenes (bp) | 1,054 | |
| GC percentage (%) | 39.82 | |

doi:10.1371/journal.pone.0132277.t001

To further evaluate the completeness of our transcriptome library and effectiveness of our annotation process, we also searched the annotated sequences for the genes involved in the clusters of orthologous group (COG) classifications. In total, of 39,179 NR hits, there were 14,116 sequences with a COG classification (Table 3). Among the 24 COG categories, the cluster for 'General function prediction only' (20.53%) was the largest group, followed by 'Post-translational modification, protein turnover, chaperones' (9.13%) and 'Replication, recombination and repair' (7.57%) clusters, while the category 'Nuclear structure' (0.04%) was the smallest group (Fig 4).

Pathway-based analysis helps in further understanding the biological functions of genes. To identify the biological pathways active in *B. macrocarpa*, we mapped all the unigenes to the referential canonical pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Of the 75,917 unigenes, 7,209 were matched in the databases and assigned to 261 KEGG pathways (Table 3 and S2 Table). Among them, spliceosome pathway (492 members) was the largest group, followed by plant–pathogen interaction pathway (271) and plant hormone signal transduction pathway (246).

**Table 2. Length distribution of unigenes in the assembled transcriptomes.**

| Unigene length (bp) | Total number | Percentage (%) |
|---|---|---|
| 0–500 | 34368 | 45.27 |
| 500–1000 | 13751 | 18.11 |
| 1000–1500 | 8613 | 11.35 |
| 1500–2000 | 7124 | 9.38 |
| 2000–3000 | 7457 | 9.82 |
| 3000–4000 | 2966 | 3.91 |
| 4000–5000 | 954 | 1.26 |
| 5000–10000 | 660 | 0.87 |
| ≥10000 | 24 | 0.03 |
| Total | 75917 | 100.00 |

doi:10.1371/journal.pone.0132277.t002

**Table 3. Summary of functional annotation of all B. macrocarpa unigenes.**

| Database | 75,917 all unigenes | |
|---|---|---|
| | Number of annotated sequences | Percentage of annotated sequences (%) |
| NR | 39176 | 51.60 |
| SwissProt | 26326 | 34.68 |
| KEGG | 7209 | 9.50 |
| COG | 14116 | 18.59 |
| GO | 16332 | 21.51 |

doi:10.1371/journal.pone.0132277.t003

## Putative molecular markers

Molecular markers have proven to be valuable tools for various applications in genetics and breeding [17]. Therefore, to develop a novel set of functional simple sequence repeats (SSRs), the *B. macrocarpa de novo* assembly transcriptome sequences were mined for the presence of microsatellite motifs using the microsatellite identification tool (MISA). In total, 3,564 sequences containing 4,834 SSRs were identified from 75,917 consensus sequences, with 761 sequences containing more than one SSR (S3 File). The most abundant type of repeat motif were tri-nucleotides (80.39%), followed by di- (14.07%), tetra- (3.23%), hexa- (1.47%) and penta-nucleotide repeats (0.85%)(Table 4). The SSR frequency is dependent on several factors such as genome structure or composition, and arithmetical method for SSR detection and parameters for exploration of microsatellites [18]. In general, it is expected that the frequency of di-, tri-, tetra-, penta- and hexa-nucleotide repeats should simultaneously decrease [19]. However, in *B. macrocarpa*, tri-nucleotides were the most abundant SSR type, consistent with reports for lentil, radish, purple sweet potato and other plants [20–22]. Moreover, in mining of SSRs in sugar beet, tri-nucleotides were the most frequently occurring motif type, whereas di-nucleotide repeats were the most abundant SSRs in genomic DNA libraries [23]. Overall, among all types of SSRs, the dominant repeat motif was TGG/CAA (8.45%), followed by GA/TC (6.83%), TGA/TCA (6.77%), GAT/ATC (5.60%) and ATG/CAT (5.33%). The frequencies of SSRs based on number of motifs revealed that SSRs with five tandem repeat motifs (48.45%) were the most common, followed by six tandem repeat motifs (20.96%), and > 10 tandem repeat motifs was least (1.65%; Table 4). Up to now, although some genomic SSRs were available for some wild beets and sugar beet [23–26], there were only a limited number of microsatellite sequences for *B. macrocarpa*. The large numbers of potential molecular markers found in our study will be particularly useful for future genetic mapping, parentage analysis, genotyping and resistant breeding of beet species.

## Genes responding to BNYVV infection in *B. macrocarpa*

To identify genes showing a significant expression change in *B. macrocarpa* upon virus infection, the differentially expressed genes between healthy and BNYVV-inoculated leaves were identified by an algorithm [27] based on the criteria of significance [False Discovery Rate (FDR) $\leq$ 0.05]. Comparison of gene expression revealed 261 genes differentially expressed between the two samples, including 128 up-regulated and 133 down-regulated genes (Fig 5A). Among the 261 differentially expressed genes, 208 (79.69%) genes were well annotated, whereas the remaining 53 (20.31%) genes had low sequence homology to known sequences in public databases (S2 File), suggesting that they might be putative novel genes in *B. macrocarpa* involved in response to BNYVV infection.
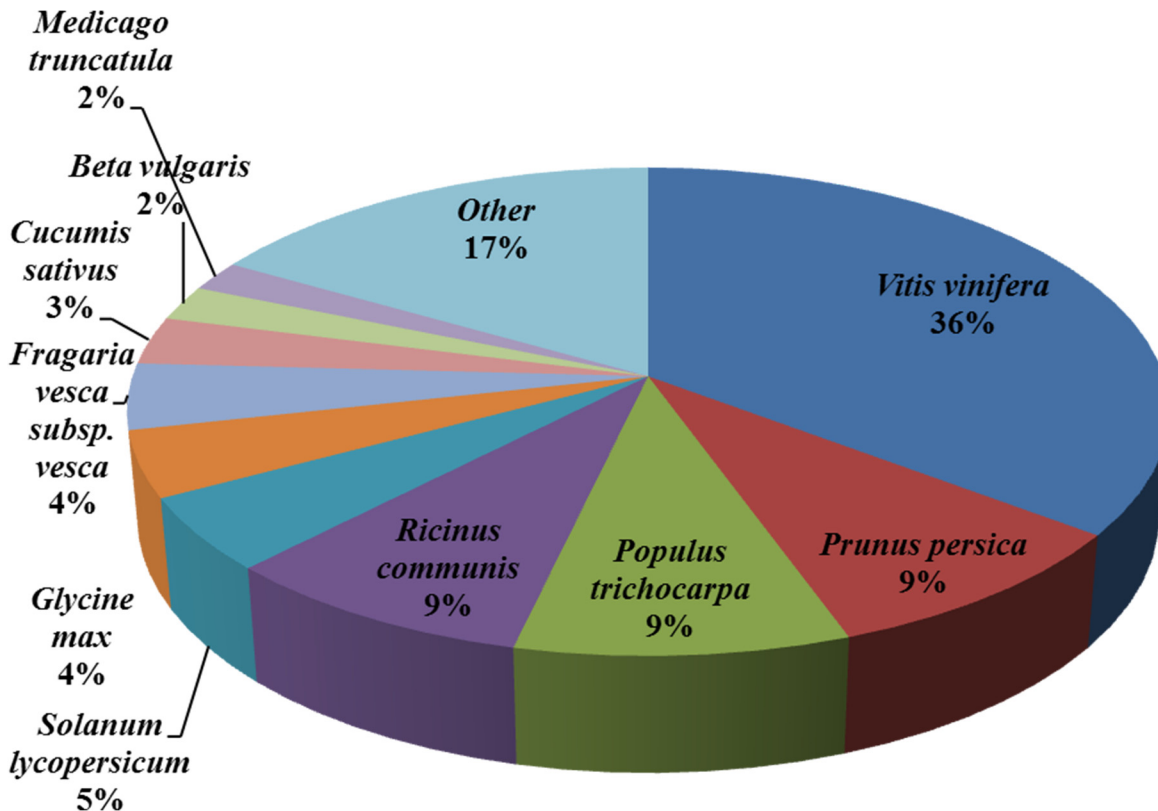
**Fig 2. Species distribution of unigene BLASTX results against the NCBI NR protein database, with a cut-off E-value $\leq 10^{-5}$ and proportions of each unigene species.** Different colors represent different species.

RNA-Seq provides a platform for a more sensitive measuring of differences in gene expression than traditional microarray hybridization experiments [28]. We used our RNA-Seq data to analyze the expression of all previously annotated genes, as well as a set of novel transcripts that were uncovered in this study. Fragments per kilobase per million reads (FPKM) values were determined for all genes in each sample, and the resulting data were transformed by first dividing each value for a gene at a particular condition by that gene's mean FPKM value across two samples and then taking $\log_{10}$ of the resulting values. These data were then subjected to hierarchical clustering using Pearson's correlation coefficient as the distance metric (Fig 5B). This enabled us to determine the similarity in relative change for each transcript in response to BNYVV infection and how these changes were similar or different between transcripts.

## Functional categorization of the identified genes regulated by BNYVV

GO analysis showed that differentially expressed genes in BNYVV-infected *B. macrocarpa* were significantly (FDR $\leq 0.01$) enriched in two main categories—molecular function and biological process—with four and five functional groups, respectively (Fig 6). The most enriched functional groups of the BNYVV-infected population had altered transcripts in response to biotic stimulus, primary metabolic process, oxidoreductase activity and hydrolase activity.

A common aspect of plant virus or viroid infection shown by expression profiling experiments is the induction of biotic and abiotic stress response genes [29–34]. Consistent with these findings, > 50% of the differentially expressed genes were involved in biotic stress and wound responses, reactive oxygen species (ROS) metabolism, pathogenesis-related (PR)
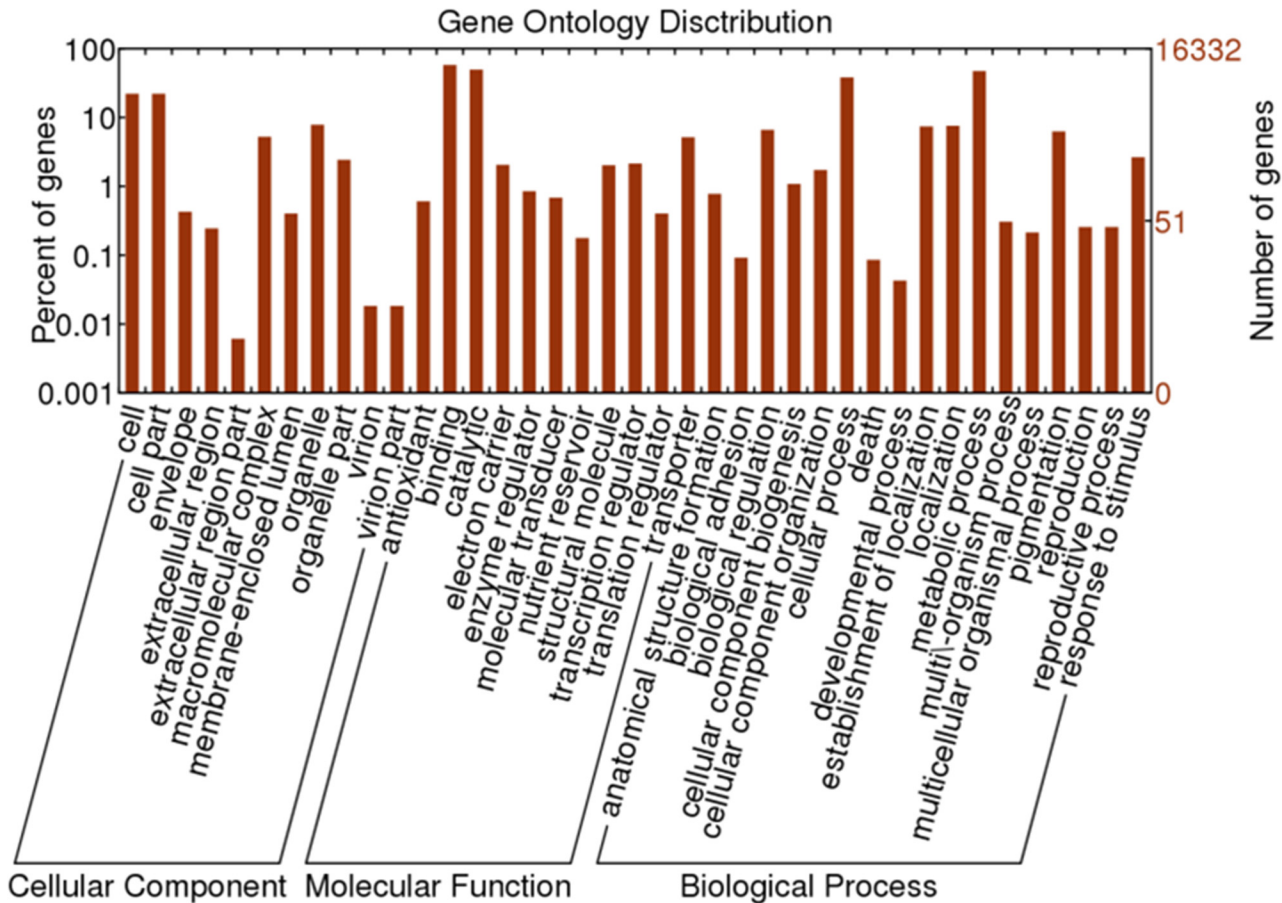
Fig 3. **Gene ontology classification of *B. macrocarpa* leaves transcriptome.** The functions of unigenes are divided into three main categories: biological processes, cellular components and molecular functions. In total, 16,332 unigenes with BLASTX matches were assigned to gene ontology.

proteins, transcription factors and putative disease resistance genes (S2 File). Plant transcription factors are often involved in response to biotic stimulus [35–39]. During BNYVV
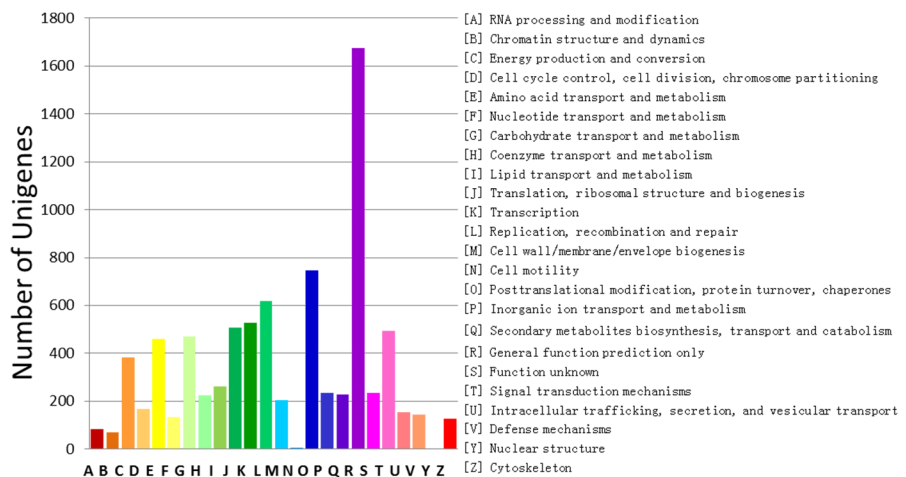


Fig 4. **Histogram classifying COG.** A total of 14,116 unigenes were grouped into 24 COG categories.

**Table 4. Frequency of occurrence of SSRs in *B. macrocarpa* transcriptome.**

| Motif length | Repeat numbers | | | | | | | Total | % |
|---|---|---|---|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 | 10 | >10 | | |
| Di | - | - | - | 353 | 127 | 130 | 70 | 680 | 14.07 |
| Tri | 2160 | 951 | 700 | 64 | 1 | 3 | 7 | 3886 | 80.39 |
| Tetra | 117 | 39 | 0 | 0 | 0 | 0 | 0 | 156 | 3.23 |
| Penta | 34 | 6 | 1 | 0 | 0 | 0 | 0 | 41 | 0.85 |
| Hexa | 33 | 18 | 9 | 1 | 4 | 3 | 3 | 71 | 1.47 |
| Total | 2344 | 1014 | 710 | 418 | 132 | 136 | 80 | | |
| % | 48.45 | 20.96 | 14.68 | 8.64 | 2.73 | 2.81 | 1.65 | | |

doi:10.1371/journal.pone.0132277.t004

infection, regulation of transcription factors and DNA- or RNA-related genes was observed; among them, six were up-regulated and 10 down-regulated (S2 File). The transcription factors, e.g. WRKY and AP2, are regulated by plant hormones such as jasmonic acid, salicylic acid and ethylene as well as by pathogen challenge, and are consistent with the existence of a highly complex regulatory network underlying the physiological response to viral infection [36,39]. In particular, WRKY6 is important for its ability to regulate the expression of defense gene PR-1 in *Arabidopsis* and *N. benthamiana* [37]. In addition, the changes of transcription factor genes could be related to expression of BNYVV protein P14 and P31, which are silencing suppressors [6,40] and are both located in nuclei [41].

PR proteins are often triggered during the early response to pathogen attack. In our study, several classes of PR proteins—including PR-1a, PR-2 (beta-1,3-glucanase), PR-3 (chitinase)
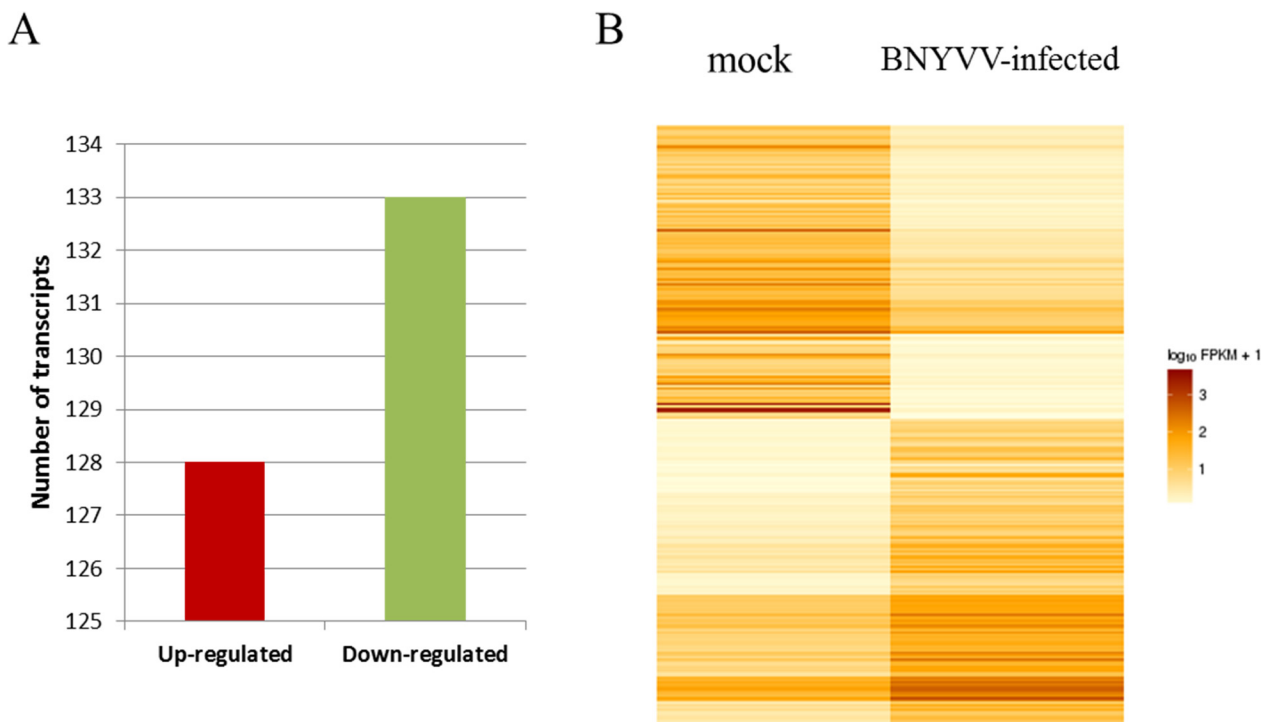


**Fig 5. Analysis of differentially expressed unigenes in BNYVV- and mock-infected *B. macrocarpa*.** (A) Numbers of differentially expressed unigenes between the mock and BNYVV-infected plant libraries. (B) Hierarchical clustering of differential expression profiles for 261 genes between the mock- and BNYVV-infected plant libraries (FDR $\leq$ 0.05 and absolute value of the log$_2$ ratio $\geq$ 2).

doi:10.1371/journal.pone.0132277.g005
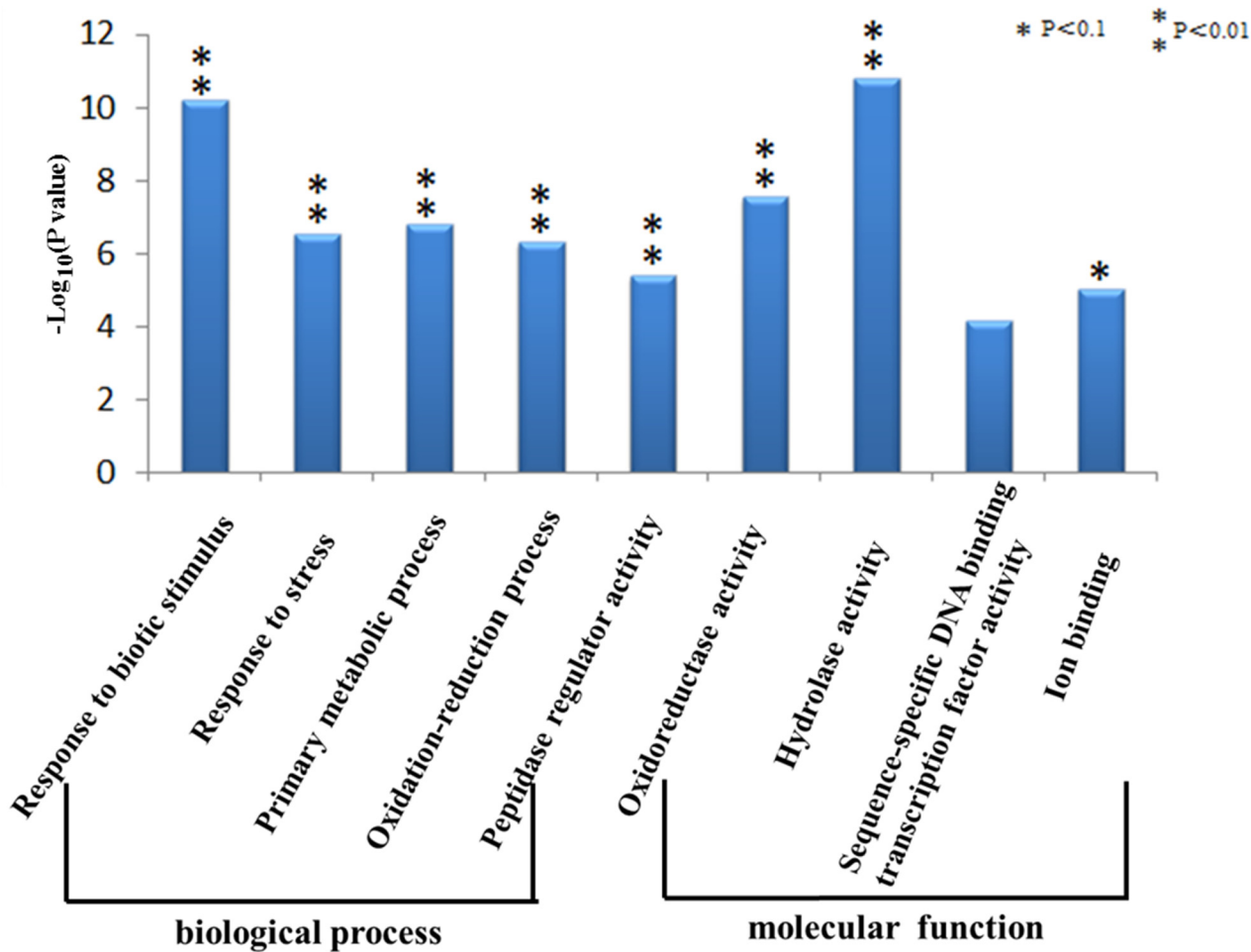
**Fig 6. GO enrichment differentially expressed genes between BNYVV-infected and non-infected *B. macrocarpa* leaves.** The Y-axis is $-\log_{10}$ transformation of the P-value calculated in enrichment test.

and PR-5 (thaumatin family proteins)—accounted for nearly 10% of up-regulated defense transcripts in response to BNYVV infection. The expressions of both PR-2 and PR-3 were both induced at a high level by BNYVV in *B. vulgaris* and *N. benthamiana* [42,43]. This suggests that these genes might have similar expression patterns in response to BNYVV infection, regardless of plant species.

For the primary metabolism, nine genes were up-regulated and 14 down-regulated in BNYVV-infected plants. Most genes involved in ATP-dependent proteolysis, amino acid metabolism and carbohydrate metabolism were down-regulated in BNYVV-infected plants, indicating that protein synthesis and amino acid and carbohydrate metabolism were inhibited by BNYVV infection. A similar effect has been reported for other plants infected by this virus [29,31,32]. The ability of viruses to interfere with these basic host functions offers opportunities not only to probe basic plant functions but also to decipher the molecular events associated with symptom development [44]. Thus, primary metabolism of *B. macrocarpa* can be disturbed by BNYVV, which might be correlated with development of systemic symptoms (arrest of growth and yellow mosaic). A previous study in *N. benthamiana* showing that suppression of CESA genes and decreases in GA accumulation act in concert to contribute to the stunted growth occuring during BNYVV ($RNA_{1+2+3+4}$) infectious. In addition, altered expressions of

genes involved in the RNA silencing pathway (AGO 4, AGO 5,AGO 10, Rnase III-like protein) were observed in BNYVV-infected *N. bethamiana*, but no substantial difference in the expression levels of putative RNA silencing related genes was found in *B. macrocarpa* after infection with BNYVV.These differents may be caused by the diveristy among the host defense systems.

The expansin gene was up-regulated in BNYVV-infected tissues and is an interesting case—it is involved in cellular expansion of cell walls and was found important in determining root phenotype, with a link described between its expression and root hair initiation [45]. Moreover, expansin genes were previously reported as typically up-regulated in infected sugar beet roots and p25 protein-expressing transgenic *Arabidopsis* plants [42,46]. We postulated that the root proliferation observed on rhizomania-infected sugar beet was possibly due to over-expression of the expansin gene.

## Verification of differential gene expression by quantitative real-time PCR (qRT-PCR)

In order to validate the reliability of Solexa sequencing, we compared the expression profiles of BNYVV-infected and non-infected *B. macrocarpa* using qRT-PCR. Twenty unigenes with annotations were selected randomly for testing using specific primers (S3 Table, see supporting information). There were 13 of 17 unigenes (76.47%) examined with the same expression profiles as deduced from the original Solexa sequencing (Fig 7A). Inconsistencies among the remaining four genes could be artificially caused by mutations with primer sites or possibly a lower sensitivity of qRT-PCR than RNA-Seq [47,48]. Nevertheless, a high correlation ($R^2$ = 0.8644) between RNA-Seq and qRT-PCR (Fig 7B) suggests that the transcriptome data is reliable and reflects the actual direction of gene expression in response to BNYVV infection.

## Conclusion

This study represents the first application of Illumina sequencing technology to obtain the transcriptome of *B. macrocarpa* challenged with BNYVV at an unprecedented depth (10.52 Gb) and produced 75,917 assembled unigenes with 39,372 unigenes obtaining annotation. These findings provide a substantial contribution to existing sequence resources for wild beet, and will likely accelerate research on the sugar beet resistance mechanism to BNYVV. Comparative transcriptome analysis between BNYVV-infected and control plants revealed significant differences in gene expression. Although the molecular functions of some genes and their associated pathways remain largely unknown, this study provides valuable information on the role of the differentially expressed genes in response to BNYVV infection. Future functional analysis of these potential virus defense genes is expected to aid a better understanding of the molecular mechanisms of pathogen-defense in beet. Furthermore, the large number of transcripts and molecular markers obtained in this study offers a strong basis for future genomic research on beets.

## Materials and Methods

### Plant, virus inoculation and detection

*Beta macrocarpa* plants were grown in a controlled-environment chamber at 24 ± 1°C under a 16/8 h light/dark regimen. BN (RNAs 1+2+3+4+5) was a mixture of total RNAs from the BN3 (RNAs 1+2+3) inoculated leaves of *Tetragonia expansa* and *in vitro* transcripts of RNA4 and RNA5. Virus inoculum supplemented with an equal volume of inoculation buffer (50 mM glycine, 30 mM $K_2HPO_4$, 1% bentonite and 1% celite at pH 9.2) was rubbed onto *B. macrocarpa* leaves as described previously [49].
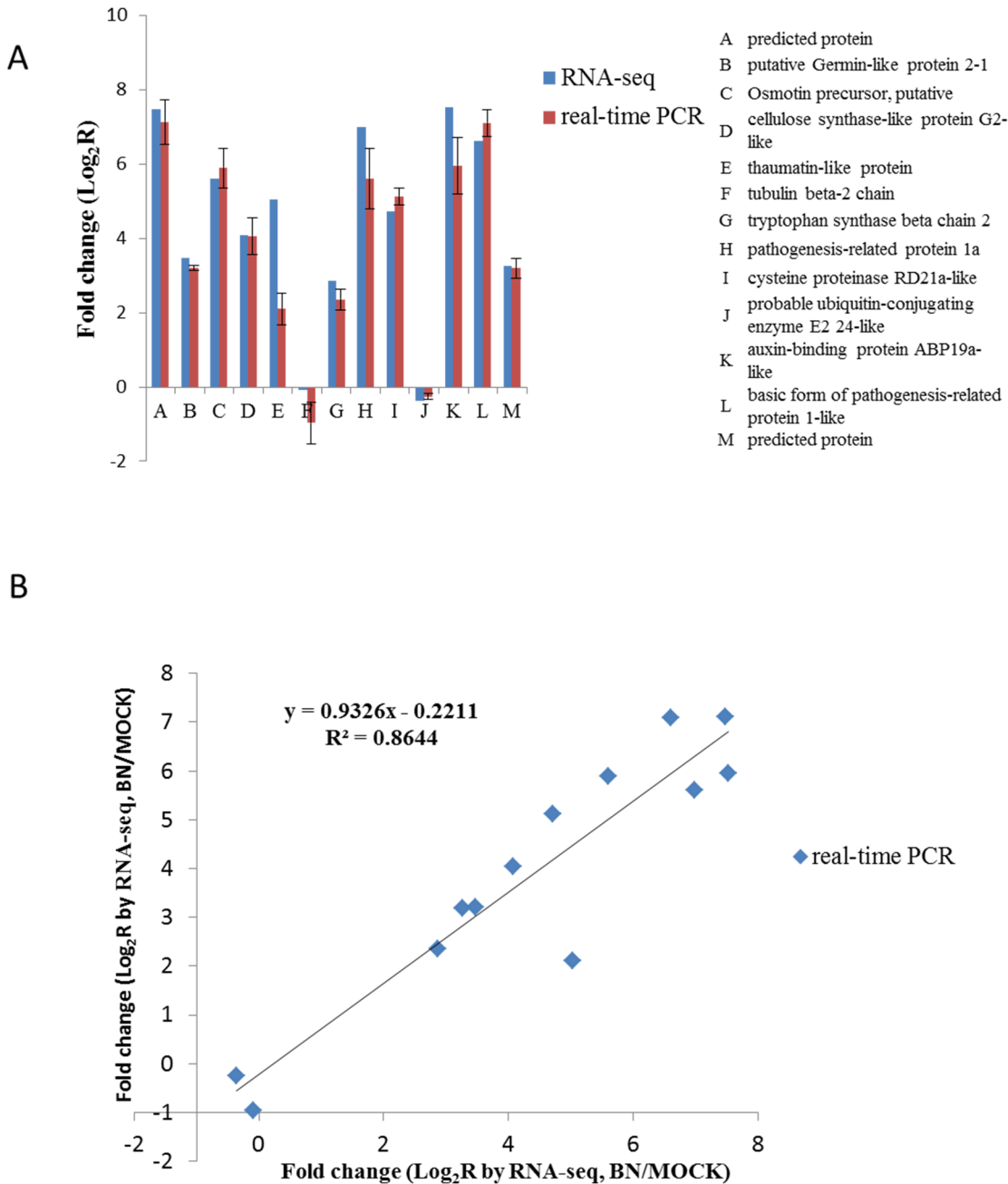
Fig 7. **Verification of the relative expression levels of genes by qRT-PCR.** (A) Expression patterns of selected *B. macrocarpa* genes in response to BNYVV, as determined by qRT-PCR (Red) and RNA-Seq (Blue). The X-axis shows the annotations of selected genes. The Y-axis shows the normalized expression levels of the transcripts. (B) Correlation of the expression ratio of selected genes measured by qRT-PCR and RNA-Seq.

Inoculated or uninoculated upper leaves from BNYVV-infected and mock-infected *B. macrocarpa* at 15 day post inoculation (dpi) were collected and homogenized in liquid nitrogen for protein extractions with equal volumes of gel loading buffer (100 mM Tris base, pH 6.8; 20% glycerol; 4% SDS; 200 mM β-mercaptoethanol; 0.2% bromophenol blue). For western blot, BNYVV coat protein specific polyclonal antibodies (obtained from rabbit immunized with purified BNYV virions) were applied at a dilution of 1:500. Alkaline phosphatase

conjugated to goat anti-rabbit antiserum (Sigma, USA) was used as secondary antibody at a dilution of 1:5000.

## cDNA library preparation and deep sequencing

Total RNAs were extracted from BNYVV-infected and non-infected *B. macrocarpa* using TRI-ZOL reagent (Takara, Dalian, China) according to the manufacturer's instructions. Integrity and size distribution were checked with Bioanalyzer 2100 (Agilent Technologies, USA). The RNA pool was prepared by mixing together equal quantities of five RNA samples per group. cDNA library preparation and Illumina-Solexa sequencing were performed as previously described. The normalized cDNA libraries were sequenced using an Illumina HiSeq 2000, and all obtained data were submitted to the NCBI database Short Read Archive.

## Analyses of Illumina sequencing results

Raw sequencing reads were quality trimmed, and dirty raw reads (i.e. reads with adaptors or reads with unknown nucleotides > 5% low quality reads) were discarded. Processed reads were assembled using CAP3 software with default parameters [16]. The overall assembly was performed using the combined sequence data for both the BNYVV-infected and non-infected samples. The contigs and singletons were generally referred to as unigenes. Subsequently, unigenes were subjected to BLASTX similarity search against the NCBI NR protein and Swiss-Prot databases with a significant threshold of E-value $\leq 10^{-5}$. Functional categorization by GO terms was performed using Blast2GO software [50]. COG and KEGG pathway annotations were performed using Blastall software against the COG [51] and KEGG databases [52], respectively.

## SSR analysis

All assembled cDNA contigs from both the infected and control libraries were used for identification of SSRs. All types of microsatellites from di- to hexa-nucleotides were detected using MISA software [53] (http://pgrc.ipk-gatersleben.de/misa/) with a minimum repeat number of eight, five, five, five and five for di-, tri-, tetra-, penta- and hexa-nucleotide microsatellites, respectively. A Perl script was designed to allow the identification and characterization of microsatellites in a comparative genomic context.

## Evaluation of different expressed genes

FPKM was used to evaluate expressed value and quantify transcript levels. We used FDR $\leq 10^{-5}$ as the threshold and |$\log_2$ ratio (BNYVV-infected/ non-infected)| $\geq 1$ to judge the significance of gene expression difference. For enrichment analysis, we mapped all differentially expressed genes to terms in GO and compared this with the genome background (the whole *B. macrocarpa* transcriptome in this study).

## qRT-PCR analysis

To confirm the results of transcriptome sequencing analysis, the relative mRNA expression levels of several randomly selected genes in RNA from mock- and BN-infected leaves were evaluated. Three RNA samples were detected for each group; whereas one was the same as that used for RNA-Seq analysis, the other two replicates were derived from different plant samples. qRT-PCR was performed in 96-well plates using the CFX96 real-time PCR detection system (Bio-Rad, Hercules, CA, USA) with the following temperature program: 95°C for 15 s, followed by 40 cycles of 95°C for 15 s, and then annealing at 60°C for 30 s. For relative quantification of

mRNA, 1 μg of total RNA was extracted from leaves, treated with DNaseI (Takara) and reverse transcribed following the manufacturer's instructions. Each reaction mixture consisted of 1 μl of cDNA, 7 μl of SsoFast EyaGreen Supermix (Bio-Rad), 1.5 μl (3 pmol/μl) of both forward and reverse primers, and 3 μl of PCR-grade water, as recommended by the manufacturer (Takara). Each reaction included amplification of PP2A transcripts, which provided an internal reference. All primers used in this study are listed in S3 Table. All PCR experiments were performed in triplicate and the results were calculated using the CFX Manage Version 1.6 software (Bio-Rad, Hercules, CA, USA) with the default parameters.

## Supporting Information

**S1 File. Summary of functional annotations of non-redundant unigenes, showing the results of searches of reference sequences using BLASTX against the NCBI NR, GO and KEGG databases with a cut-off E-value of $10^{-5}$.**
(XLSX)

**S2 File. Differentially expressed transcripts between healthy and BNYVV-infected samples.**
(XLSX)

**S3 File. Frequency of SSRs occurred in *B.macrocarpa*.**
(XLSX)

**S1 Table. Summary of data production.**
(DOCX)

**S2 Table. Unigenes KEGG pathway analysis.**
(XLSX)

**S3 Table. Primers used in qRT-PCR.**
(XLSX)

## Acknowledgments

## Author Contributions

Conceived and designed the experiments: CH DL JY. Performed the experiments: HF JL. Analyzed the data: HF HS. Contributed reagents/materials/analysis tools: YW YZ XW. Wrote the paper: HF CH.

## References

1. McGrann GR, Grimmer MK, MUTASA-GÖTTGENS ES, Stevens M (2009) Progress towards the understanding and control of sugar beet rhizomania disease. Mol Plant Pathol 10: 129–141. doi: 10.1111/j.1364-3703.2008.00514.x PMID: 19161359

2. Chiba S, Kondo H, Miyanishi M, Andika IB, Han C, et al. (2011) The evolutionary history of *Beet necrotic yellow vein virus* deduced from genetic variation, geographical origin and spread, and the breaking of host resistance. Mol Plant Microbe Interact 24: 207–218. doi: 10.1094/MPMI-10-10-0241 PMID: 20977309

3. Rush CM (2003) Ecology and epidemiology of benyviruses and plasmodiophorid vectors. Annu Rev Phytopathol 41: 567–592. PMID: 14527334

4. Richards K, Tamada T (1992) Mapping functions on the multipartite genome of *Beet necrotic yellow vein virus*. Annu Rev Phytopathol 30: 291–313.

5. Tamada T, Uchino H, Kusume T, Saito M (1999) RNA 3 deletion mutants of *Beet necrotic yellow vein virus* do not cause rhizomania disease in sugar beets. Phytopathology 89: 1000–1006. doi: 10.1094/PHYTO.1999.89.11.1000 PMID: 18944654

6. Rahim MD, Andika IB, Han C, Kondo H, Tamada T (2007) RNA4-encoded p31 of *Beet necrotic yellow vein virus* is involved in efficient vector transmission, symptom severity and silencing suppression in roots. J Gen Virol 88: 1611–1619. PMID: 17412994

7. Fan H, Sun H, Wang Y, Zhang Y, Wang X, Li D, et al. (2014) Deep sequencing–based transcriptome profiling reveals comprehensive insights into the responses of *Nicotiana benthamiana* to *Beet necrotic yellow vein virus* infections containing or lacking RNA4. PloS One 9: e85284. doi: 10.1371/journal.pone.0085284 PMID: 24416380

8. Biancardi E, Lewellen RT, De Biaggi M, Erichsen AW, Stevanato P (2002) The origin of rhizomania resistance in sugar beet. Euphytica 127: 383–397.

9. Scholten OE, Lange W (2000) Breeding for resistance to rhizomania in sugar beet: a review. Euphytica 112: 219–231.

10. Gidner S, Lennefors B-L, Nilsson N-O, Bensefelt J, Johansson E, Gyllenspetz U, et al. (2005) QTL mapping of BNYVV resistance from the WB41 source in sugar beet. Genome 48: 279–285. PMID: 15838550

11. Heijbroek W, Musters P, Schoone A (1999) Variation in pathogenicity and multiplication of *Beet necrotic yellow vein virus* (BNYVV) in relation to the resistance of sugar-beet cultivars. European Journal of Plant Pathol J 105: 397–405.

12. Liu H-Y, Sears J, Lewellen R (2005) Occurrence of resistance-breaking *Beet necrotic yellow vein virus* of sugar beet. Plant disease 89: 464–468.

13. Dohm JC, Lange C, Holtgräwe D, Sörensen TR, Borchardt D, Schul B, et al. (2012) Palaeohexaploid ancestry for Caryophyllales inferred from extensive gene-based physical and genetic mapping of the sugar beet genome (*Beta vulgaris*). Plant Pathol J 70: 528–540.

14. Filipa M, Maria M R, Dora B, Maria Cristina D (2013) Biodiversity assessment of sugar beet species and its wild relatives: linking ecological data with new genetic approaches. Am J Plant Sci 4: 21.

15. Geyl L, Heriz MG, Valentin P, Hehn A, Merdinoglu D (1995) Identification and characterization of resistance to rhizomania in an ecotype of *Beta vulgaris subsp*. *maritima*. Plant Pathol J 44: 819–828.

16. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9: 868–877. PMID: 10508846

17. Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, et al. (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences.Am J Bot 99: 193–208. doi: 10.3732/ajb.1100394 PMID: 22186186

18. Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10: 967–981. PMID: 10899146

19. Zeng S, Xiao G, Guo J, Fei Z, Xu Y, Roe B, et al. (2010) Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. BMC genomics 11: 94. doi: 10.1186/1471-2164-11-94 PMID: 20141623

20. Verma P, Shah N, Bhatia S (2013) Development of an expressed gene catalogue and molecular markers from the *de novo* assembly of short sequence reads of the lentil (*Lens culinaris Medik*.) transcriptome. Plant Biotechnol J 11: 894–905. doi: 10.1111/pbi.12082 PMID: 23759076

21. Wang S, Wang X, He Q, Liu X, Xu W, Li L, et al. (2012) Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish. Plant Cell Rep 31: 1437–1447. doi: 10.1007/s00299-012-1259-3 PMID: 22476438

22. Xie F, Burklew CE, Yang Y, Liu M, Xiao P, Zhang B, et al. (2012) *De novo* sequencing and a comprehensive analysis of purple sweet potato (*Impomoea batatas L*.) transcriptome. Planta 236: 101–113. doi: 10.1007/s00425-012-1591-4 PMID: 22270559

23. Laurent V, Devaux P, Thiel T, Viard F, Mielordt S, Touzet P, et al. (2007) Comparative effectiveness of sugar beet microsatellite markers isolated from genomic libraries and GenBank ESTs to map the sugar beet genome. Theor Appl Genet 115: 793–805. PMID: 17646961

24. Cureton A, Burns M, Ford-Lloyd B, Newbury H (2002) Development of simple sequence repeat (SSR) markers for the assessment of gene flow between sea beet (*Beta vulgaris ssp*. *maritima*) populations. Mol Ecol Notes 2: 402–403.

25. Li J, Schulz B, Stich B (2010) Population structure and genetic diversity in elite sugar beet germplasm investigated with SSR markers. Euphytica 175: 35–42.

26. Viard F, Arnaud JF, Delescluse M, Cuguen J (2004) Tracing back seed and pollen flow within the crop–wild *Beta vulgaris* complex: genetic distinctiveness vs. hot spots of hybridization over a regional scale. Mol Ecol 13: 1357–1364. PMID: 15140082

27. Audic S, Claverie J-M (1997) The significance of digital gene expression profiles. Genome Res 7: 986–995. PMID: 9331369

28. Wilhelm BT, Landry J-R (2009) RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48: 249–257. doi: 10.1016/j.ymeth.2009.03.016 PMID: 19336255

29. Catoni M, Miozzi L, Fiorilli V, Lanfranco L, Accotto GP (2009) Comparative analysis of expression profiles in shoots and roots of tomato systemically infected by *Tomato spotted wilt virus* reveals organ-specific transcriptional responses. Mol Plant Microbe Interact 22: 1504–1513. doi: 10.1094/MPMI-22-12-1504 PMID: 19888816

30. Dardick C (2007) Comparative expression profiling of *Nicotiana benthamiana* leaves systemically infected with three fruit tree viruses. Mol Plant Microbe Interact 20: 1004–1017. PMID: 17722703

31. Zhang Y, Pei X, Zhang C, Lu Z, Wang Z, Jia S, et al. (2012) De novo foliar transcriptome of *Chenopodium amaranticolor* and analysis of its gene expression during virus-induced hypersensitive response. PloS One 7: e45953. doi: 10.1371/journal.pone.0045953 PMID: 23029338

32. Lu J, Du Z-, Kong J, Chen L-, Qiu Y-, Li G, et al. (2012) Transcriptome analysis of *Nicotiana tabacum* infected by *Cucumber mosaic virus* during systemic symptom development. PloS One 7: e43447. doi: 10.1371/journal.pone.0043447 PMID: 22952684

33. Hanssen IM, van Esse HP, Ballester A-R, Hogewoning SW, Parra NO, Paeleman A, et al. (2011) Differential tomato transcriptomic responses induced by *Pepino mosaic virus* isolates with differential aggressiveness. Plant Physiol J 156: 301–318.

34. Itaya A, Matsuda Y, Gonzales RA, Nelson RS, Ding B (2002) *Potato spindle tuber viroid* strains of different pathogenicity induces and suppresses expression of common and unique genes in infected tomato. Mol Plant Microbe Interact 15: 990–999. PMID: 12437296

35. Whitham SA, Yang C, Goodin MM (2006) Global impact: elucidating plant responses to viral infection. Mol Plant Microbe Interact 19: 1207–1215. PMID: 17073303

36. Eulgem T, Somssich IE (2007) Networks of WRKY transcription factors in defense signaling. Curr Opin Plant Biol 10: 366–371. PMID: 17644023

37. Gutterson N, Reuber TL (2004) Regulation of disease resistance pathways by AP2/ERF transcription factors. Curr Opin Plant Biol 7: 465–471. PMID: 15231271

38. Fukazawa J, Sakai T, Ishida S, Yamaguchi I, Kamiya Y, Takahashia Y (2000) Repression of shoot growth, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. Plant Cell 12: 901–915. PMID: 10852936

39. Dai S, Wei X, Alfonso AA, Pei L, Duque UG, Zhang Z, et al. (2008) Transgenic rice plants that overexpress transcription factors RF2a and RF2b are tolerant to rice tungro virus replication and disease. Proc Natl Acad Sci USA 105: 21012–21016. doi: 10.1073/pnas.0810303105 PMID: 19104064

40. Zhang L, Wang Z, Wang X, Li D, Han C, Chai Y, et al. (2005) Two virus-encoded RNA silencing suppressors, P14 of *Beet necrotic yellow vein virus* and S6 of *Rice black streak dwarf virus*. Chin Sci Bull 50: 305–310.

41. Chiba S, Hleibieh K, Delbianco A, Klein E, Ratti C, Ziegler-Graff V, et al. (2013) The benyvirus RNA silencing suppressor is essential for long-distance movement, requires both zinc-finger and NoLS basic residues but not a nucleolar localization for its silencing-suppression activity. Mol Plant Microbe Interact 26: 168–181. doi: 10.1094/MPMI-06-12-0142-R PMID: 23013437

42. Schmidlin L, De Bruyne E, Weyens G, Lefebvre M, Gilmer D (2008) Identification of differentially expressed root genes upon rhizomania disease. Mol Plant Pathol 9: 741–751. doi: 10.1111/j.1364-3703.2008.00498.x PMID: 19019003

43. Larson RL, Wintermantel WM, Hill A, Fortis L, Nunez A (2008) Proteome changes in sugar beet in response to *Beet necrotic yellow vein virus*. Physiol Mol Plant Pathol 72: 62–72.

44. Culver JN, Padmanabhan MS (2007) Virus-induced disease: altering host physiology one interaction at a time. Annu Rev Phytopathol 45: 221–243. PMID: 17417941

45. Cho H-T, Cosgrove DJ (2002) Regulation of root hair initiation and expansin gene expression in *Arabidopsis*. Plant Cell 14: 3237–3253. PMID: 12468740

46. Peltier C, Schmidlin L, Klein E, Taconnat L, Prinsen E, Erhardt M, et al. (2011) Expression of the *Beet necrotic yellow vein virus* p25 protein induces hormonal changes and a root branching phenotype in *Arabidopsis thaliana*. Transgenic Res 20: 443–466. doi: 10.1007/s11248-010-9424-3 PMID: 20602166

47. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Adiconis X, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnol 29: 644–652.

48. AC't Hoen P, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RH, Menezes1 R X, et al. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab

portability over five microarray platforms. Nucleic Acids Res  36: e141–e141. doi: 10.1093/nar/gkn705 PMID: 18927111

49.  Wang Y, Fan H, Wang X, Li M, Han C, Li D, et al. (2011) Detection and characterization of spontaneous internal deletion mutants of *Beet Necrotic yellow vein virus* RNA3 from systemic host *Nicotiana benthamiana*. Virol J  8: 1–9.

50.  Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics  21: 3674–3676. PMID: 16081474

51.  Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, et al. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res  29: 22–28. PMID: 11125040

52.  Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res  36: D480–D484. PMID: 18077471

53.  Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare L.*). Theor Appl Genet  106: 411–422. PMID: 12589540