Genetics and Molecular Research

Discovery of clubroot-resistant genes in *Brassica napus* by transcriptome sequencing

S.W. Chen[1]*, T. Liu[1]*, Y. Gao[1], C. Zhang[2], S.D. Peng[3], M.B. Bai[1], S.J. Li[1], L. Xu[1], X.Y. Zhou[1] and L.B. Lin[1]

[1]College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming, China
[2]College of Tropical Crops, Yunnan Agricultural University, Kunming, China
[3]Agricultural Experimental Teaching Center, Yunnan Agricultural University, Kunming, China

*These authors contributed equally to this study.
Corresponding author: L.B. Lin
E-mail: linliangbin-63@163.com

Genet. Mol. Res. 15 (3): gmr.15038243
Received December 10, 2015
Accepted April 15, 2016
Published August 12, 2016
DOI http://dx.doi.org/10.4238/gmr.15038243

ABSTRACT. Clubroot significantly affects plants of the Brassicaceae family and is one of the main diseases causing serious losses in *B. napus* yield. Few studies have investigated the clubroot-resistance mechanism in *B. napus*. Identification of clubroot-resistant genes may be used in clubroot-resistant breeding, as well as to elucidate the molecular mechanism behind *B. napus* clubroot-resistance. We used three *B. napus* transcriptome samples to construct a transcriptome sequencing library by using Illumina HiSeq™ 2000 sequencing and bioinformatic analysis. In total, 171 million high-quality reads were obtained, containing 96,149 unigenes of N50-value. We aligned the obtained unigenes with the Nr, Swiss-Prot, clusters of orthologous groups, and gene ontology databases and annotated their functions. In the Kyoto encyclopedia of

genes and genomes database, 25,033 unigenes (26.04%) were assigned to 124 pathways. Many genes, including broad-spectrum disease-resistance genes, specific clubroot-resistant genes, and genes related to indole-3-acetic acid (IAA) signal transduction, cytokinin synthesis, and myrosinase synthesis in the Huashuang 3 variety of *B. napus* were found to be related to clubroot-resistance. The effective clubroot-resistance observed in this variety may be due to the induced increased expression of these disease-resistant genes and strong inhibition of the IAA signal transduction, cytokinin synthesis, and myrosinase synthesis. The homology observed between unigenes 0048482, 0061770 and the Crr1 gene shared 94% nucleotide similarity. Furthermore, unigene 0061770 could have originated from an inversion of the Crr1 5'-end sequence.

Key words: *B. napus*; Clubroot; Transcriptome; Clubroot-resistance genes; Homology

INTRODUCTION

Clubroot is a soil-borne disease caused by the clubroot fungus (*Plasmodiophora brassicae*) in *Brassica*. It affects many plants of the Brassicaceae family, including *B. napus*, and bok choy. In recent years, this disease has become increasingly severe across the world, especially in Europe, North America, and East Asia. In many regions of China, crops belonging to the Brassicaceae family are affected by clubroot and the incidence and severity of this disease are increasing every year. Clubroot has become one of the primary diseases in plants of the Brassicaceae family, resulting in remarkably decreased (20-90%) production of crops. Breeding clubroot-resistant varieties is the most effective way to control and eradicate clubroot. However, excessive genetic mutations (Tanaka et al., 2006) and complicated pathogenesis (Strelkov et al., 2006) of the clubroot fungus render it difficult to breed clubroot-resistant varieties. This is especially true for varieties with a broad resistance spectrum and long-lasting resistance against the clubroot fungus. The most effective way to solve this problem is to select materials with different clubroot-resistant genes, and then use these materials to breed varieties with broad-spectrum and long-lasting resistance, using pyramid breeding.

Suwabe et al. (2003) identified two clubroot-resistant genes, Crr1 and Crr2, in *Brassica rapa* Siloga, and mapped them to the eighth and first linkage groups, respectively, using simple sequence repeat (SSR) markers. Suwabe et al. (2006) mapped a clubroot micro-resistant gene Crr4 in *B. rapa* Wakayama-01 also by using the SSR marker. Likewise, Hirai et al. (2004) identified a resistance gene Crr3 in *B. rapa* Milan White, and mapped this on the third linkage group, using sequenced tagged site (STS) markers. Ueno et al. (2012) first reported the molecular characteristics of *B. rapa* clubroot-resistant gene CRa, which encodes a TIR-NBS-LRR protein and has seven structural differences between the sequences of susceptible and resistant varieties. Matsumoto et al. (1998) mapped CRa in *B. rapa* on the third linkage group, using linkage marker. Piao et al. (2009) mapped CRb in DH lines from Chinese cabbage Shinki (*B. rapa* L. ssp *pekinensis*) on the third linkage group, by using sequence characterized amplified region markers. Sakamoto et al. (2008) mapped CRk and CRc in *B. napus* Debra

on the third and second linkage groups using STS markers. Manzanares-Dauleux et al. (2000) mapped clubroot-resistant gene *Pb-Bn1* on the linkage group using DH group from the hybrid between *B. napus* Darmor-bzh and Yudal. By using map-based cloning technology, Hatakeyama et al. (2013) identified gene Crr1, which encodes a TIR-NBS-LRR protein, with clubroot-resistance function.

Next-generation sequencing technologies, such as pyrosequencing, circumvent the lengthy and relatively low-throughput steps associated with Sanger sequencing, as well as provide rapid and economical technologies for transcriptomics. During the last decade, a large number of transcriptomic sequences have been generated and collected in both model and non-model organisms. This has greatly accelerated our understanding of the complexity of gene expression, regulation, and networks in higher plants. Furthermore, the large number of expressed sequence tags generated from transcriptome sequencing has provided valuable genetic resources for functional genomics and molecular marker development. Using these methods, Devos and Prinsen (2006) showed that blocking the indole-3-acetic acid (IAA) signal transduction, downregulating cytokinin synthase, and myrosinase expression enhanced *B. napus* plant resistance to clubroot. Siemens et al. (2006) found changes of 20 proteins in *B. napus* during the process of clubroot fungus infection. These proteins were mainly related to cytokinin metabolism, antioxidant protein expression, and glycolysis. Jubault et al. (2013) also found that some genes related to light reaction, lignin synthesis, and flavonoid synthesis are related to *B. rapa* clubroot-resistance.

Clubroot fungus is an obligatory parasite that cannot be cultured *in vitro* in artificial culture media. Furthermore, long-term storage of clubroot is very difficult, which results in different results in the identification of the clubroot fungus strain, and prevents the study of clubroot-resistant breeding and molecular biology of *Brassica*-resistance to clubroot. Although many clubroot-resistant genes and resistant quantitative trait loci have been identified in *Brassica* and the functional regions of proteins encoded by Crr1 and CRa have been analyzed, the molecular mechanism of *Brassica* defensive response against clubroot is not yet known. In this study, we used high-throughput sequencing technology to analyze gene expression differences in resistant and susceptible material at the transcriptional level. We selected clubroot-resistant genes to elucidate their expression characteristics induced by clubroot fungus infection. These findings may provide some evidence for revealing the molecular mechanism of clubroot-resistance in *B. napus* as well as in the breeding of clubroot-resistant *B. napus*.

MATERIAL AND METHODS

Material cultivation and clubroot inoculation

In our study, we used seedlings of Zhongshuangyou 8, which is highly susceptible to clubroot fungus, from the Economic Crops Research Institute of Henan Academy of Agricultural Sciences. As the highly clubroot-resistant variety, we used Huashuang 3 seedlings from Huazhong Agricultural University. Their resistance indices reached 85.19 and 22.29, respectively.

To inoculate clubroot fungus, we used a slightly modified version of the inoculation method suggested by Johnston (1968). When mature galled roots showed greyish internal

mottling or partial surface decomposition, they were collected from *B. napus* from Xiaoshao village, Qingzhu village, and Shangduilong village, in Kunming, China, respectively. Mature galled roots of the same weight were soaked in an 4 volume amount of sterilized water, and then squeezed gently into homogenate. The homogenates were placed in sealed bags and fermented for 48 h at 25°C, followed by filtration through a double cheesecloth. The clubroot fungus liquid, without *B. napus* root residues, was collected and kept in a 4°C refrigerator until further analysis. The spore concentration of the clubroot fungus liquid was counted using a hemocytometer Invitrogen Life Technologies, New York, NY, USA) and adjusted to 2 x 10^8/mL.

The treatment, consisting of injecting 20 mL clubroot fungus liquid, and control (20 mL sterilized water; only applied to Huashuang 3) were applied at the 5-leaf stage. The injections were done directly into the taproots in the seedlings' root hair zone and were applied to 15 seedlings planted at a density of three seedlings/pot.

Sampling and RNA extraction

RNA was isolated from the roots at 20 days after inoculation. At this time, seedlings of the different treatments were randomly sampled. Elongation and maturation zones 0.5-1 cm length were cut from the seedling's main root, and used to extract total RNA by the Trizol method (Ueno et al., 2012). The material was treated with DNaseI (Takara Shuzo Co., Kyoto, Japan) for 4 h, and its quality was assessed using Agilent Technologies 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Total RNA from different seedlings in the same treatment was mixed at equal concentrations to create a sample pool; Y1 (inoculated Zhongshuangyou 8), Y2 (inoculated Huashuang 3), and Y3 (un-inoculated Huashuang 3).

Preparation and sequencing analysis of RNA-Seq sequencing library

The mRNAs of the sample pools were enriched using magnetic beads conjugated with oligo (dT). Fragmentation buffer was added to the enriched mRNA solution to obtain small mRNA fragments that were then used as templates. Random hexamer primers were used to synthesize the first and second cDNA chains. The cDNA was then purified using QiaQuick PCR kit and washed with EB buffer. Following ends repair, addition of base A, and serial sequencing connectors, the fragments with sequencing connectors were electrophoresed and recovered from the agarose gel. The recovered DNA fragments were then amplified again, to construct the cDNA library. The sequencing of the DNA fragments in the library and the subsequent sequence analysis were performed in Gene Denovo, Guangzhou, China, using Illumina HiSeq™ 2000.

De novo assembly and gene annotation

The raw data were filtered to select clean reads. Using the comparison between the sequencing data and the reference data, the sequencing data were evaluated and the expressed genes were annotated. At the same time, a structure optimization and alternative splicing analysis of the genes were conducted by aligning the sequences of known and expressed genes. Based on the above results, new transcripts were predicted. Next, analysis of the differential gene expression from the different samples, function annotation, gene ontology (GO)

functional enrichment analysis, and pathway significance enrichment analysis were conducted. We used the short oligonucleotide alignment software SOAPaligner/soap (Li et al., 2009) to compare the obtained sequences with those of reference genes, and then statistically compiled the results. Unigene expression was calculated by RPKM (reads per kilobase per million reads method) (Mortazavi et al., 2008). We aligned the unigene sequences to the protein sequences of protein databases Nr, Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), GO, and clusters of orthologous groups (COG) by BLASTx, respectively. Furthermore, we aligned the mRNA sequences to gene sequences obtained from the NCBI database.

Florescent quantitative polymerase chain reaction (PCR) validation

Based on sequences obtained from the *B. napus* digital gene-expression spectrum sequences and the experimental primer design principle of fluorescent quantitative PCR, we designed primers for differential gene expression using Primer 3.0. A 20-μL reverse transcription reaction system containing 1 μg total RNA template was prepared according to Bestar qPCR RT Kit instructions (DBI Bioscience, USA). Using this reverse transcription reaction system, the first cDNA chain was synthesized and collected for further use. The real-time (RT) PCR amplification system included a 20-μL reaction mixture (DBI Bestar® SybrGreen qPCRmasterMix), and the amplification program was as follows: denaturation at 94°C for 2 min, followed by 40 cycles of 30 s at 94°C, 30 s at 58°C, 30 s at 72°C, and a final step at 72°C for 5 min. The amplification reaction of fluorescent quantitative PCR was replicated three times and was performed using Agilent Stratagene fluorescent quantitative PCR Mx3000P.

B. napus genomic DNA extraction and PCR amplification

The young leaves of Zhongshuangyou 8 and Huashuang 3 plants were used to extract genomic DNA using the CTAB method (Stewart and Via, 1993; Porebski et al., 1997). The extracted DNA was dissolved in 50 μL TE buffer and then stored at -20°C. By using the sequences of gene Crr1 and unigenes 0048482 and 0061770, PCR primers were designed using the Primer 5.0 software. The 20-μL PCR amplification system contained 2.0 μL 10X PCR buffer, 2.0 μL 25 mM $MgCl_2$, 0.5 μL 2.5 mM dNTPs, 0.5 μL 5 U/μL Taq polymerase, 1.0 μL 5 μmol each primer, 6 μL template DNA, and 7 μL ddH_2O. The PCR conditions were as follows: denaturation at 95°C for 5 min, followed by 30 cycles of 30 s at 95°C, 30 s at 55°C, 1 min at 72°C, and a final step at 72°C for 10 min. The amplification products were mixed with an equal volume of loading buffer, loaded onto a 1.5% agarose gel containing Goldview, and then separated by agarose gel electrophoresis. During the electrophoresis, the agarose gel was soaked in 1X TBE buffer, and 3-5 V/cm voltage was used.

RESULTS

Sequencing and sequence assembly

To elucidate the transcriptional level response mechanism of *B. napus* to clubroot fungus infection, we used the Illumina HiSeq™ 2000 platform (Pertea et al., 2003; Yassour et al., 2009) to conduct paired-end sequencing of three RNA samples (Y1, Y2, and Y3). We found

that 171.1 million clean reads could be assembled successfully. The total nucleotide number was 17.1 billion and the GC content was 47.36% (Table 1). The percentage of reads with Q20 (nucleotide recognition accuracy of 99%) was 98.68%. Among the reads from samples Y1, Y2, and Y3, 35.3, 43.0, and 32.2 million reads, respectively, were mapped, representing 63.97, 65.43, and 64.59% of the total reads, respectively. After removing the reads that contained only linker sequences and reads of low quality, the number of clean reads obtained from the samples were 27.6, 32.8, and 24.9 million for Y1, Y2, and Y3, respectively (Table 2).

Among the clean reads, 126,072 contigs were obtained with average and N50 lengths of 1024 and 1441 bp, respectively (Table 1). The contig length ranged from 201 to 15,565 bp, but most of the contigs fell within the 200-500 bp range (Figure 1A). From the contigs, 96,149 unigenes were assembled. The mean and N50 lengths of the unigenes were 973 and 1422 bp, respectively (Table 1). The unigene length ranged from 201 to 15,562 bp, but most of the unigenes fell in the 200 to 500 bp range (Figure 1B).

Table 1. Summary of Illumina paired-end sequencing and assembly for *Brassica napus.*

	Total number	Total number of nucleotides	Average length (bp)	N50	GC percentage
Reads	171,098,708	17,109,870,800	-	-	47.36
Contigs	126,072	129,186,863	1,024	1,441	45.07
Unigenes	96,149	93,608,375	973	1,422	45.24

Table 2. Summary of sequencing and *de novo* assembly of Illumina paired-end reads for *Brassica napus* in three treatments.

	Y1	Y2	Y3
Raw reads	55,287,102	65,695,840	49,831,790
Total base pairs	5,528,710,200	6,569,584,000	4,983,179,000
Total mapped reads	35,270,121 63.97% (35,270,121/55,287,102)	42,987,259 65.43% (42,987,259/65,695,840)	32,186,025 64.59% (32,186,025/49,831,790)
Clean reads	27,643,551	32,847,920	24,915,895
Perfect match	23,209,201 41.98%[a] 65.8%[b]	30,239,828 46.03%[a] 70.35%[b]	21,897,180 43.94%[a] 68.03%[b]
≤2 bp mismatch	12,060,920 21.82%[a] 34.2%[b]	12,747,431 19.40%[a] 29.65%[b]	10,288,845 20.65%[a] 31.97%[b]
Unique match	22,566,937 40.82%[a] 63.98%[b]	28,229,811 42.97%[a] 65.67%[b]	20,308,058 40.75%[a] 63.1%[b]
Multi-position match	12,703,184 22.98%[a] 36.02%[b]	14,757,448 22.46%[a] 34.33%[b]	11,877,967 23.84%[a] 36.9%[b]
Total unmapped reads	20,016,981 36.21%[a] 56.75%[b]	22,708,581 34.57%[a] 52.83%[b]	17,645,765 35.41%[a] 54.82%[b]

[a]Percent sequences to that of raw reads. [b]Percent sequences to that of aligned reads.

Unigene function annotation

Using BLASTx, the unigenes were aligned to the protein sequences of four public protein databases (Nr, Swiss-Prot, KEGG, and COG). Thus, protein function annotation information of the unigenes was obtained. The results showed that, among all unigenes, 78,398 (81.54%), 59,896 (62.29%), 29,007 (30.17%), and 25,033 (26.04%) unigenes were annotated in Nr, Swiss-Prot, COG, and KEGG, respectively. The unigenes annotated in the four databases are illustrated in a Venn diagram (Figure 2). We found that 15.52% of the unigenes had a function annotation in all four databases, whereas 17.43, 0.23, 0.03, and 0.05% unigenes had a function annotation in only Nr, Swiss-Prot, COG, or KEGG, respectively.

The percentage of unigenes with an e-value lower than $1E^{-50}$ of the total number of unigenes annotated in the Nr, Swiss-Prot, COG, and KEGG databases were 59, 48, 21, and 37%, respectively, whereas the unigene sequence similarity in the four databases were 75, 71, 82, and 80%, respectively (Figure 3). In the Nr database, 59 and 18% of the unigenes showed a high similarity to the protein sequences of *Arabidopsis thaliana* and *A. lyrata*, respectively. In addition, 3% of the unigenes showed similarity to the protein sequences of *Oryza sativa* (rice), *Theobroma cacao* (coco tree), and *B. rapa*, whereas 2% of the unigenes showed similarity to the protein sequences of *B. napus* and *Cucumis sativus* (Figure 4).

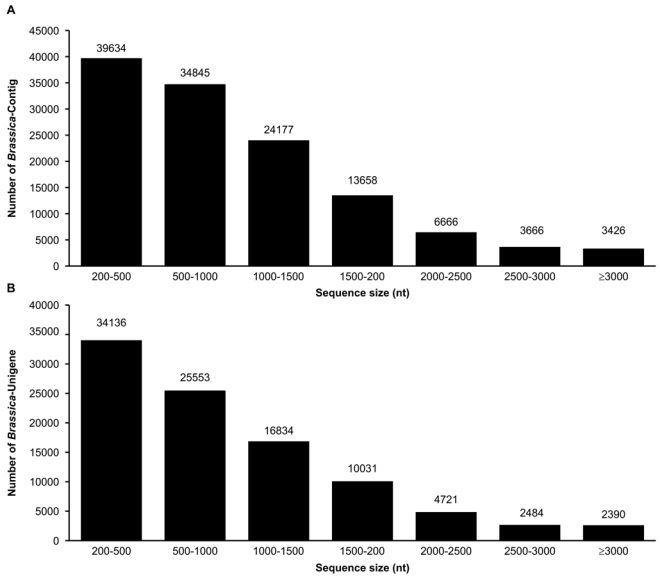

Figure 1. Assessment of assembly quality showing the length distribution of contigs (**A**) and unigenes (**B**), respectively.

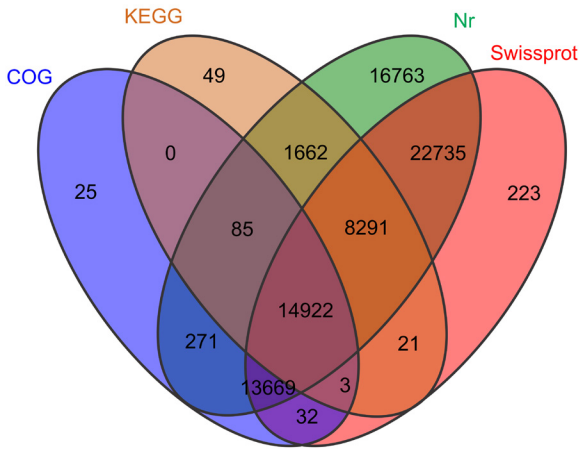

Figure 2. Venn diagram illustrating the number of unigenes with functional annotations in four protein data bases.

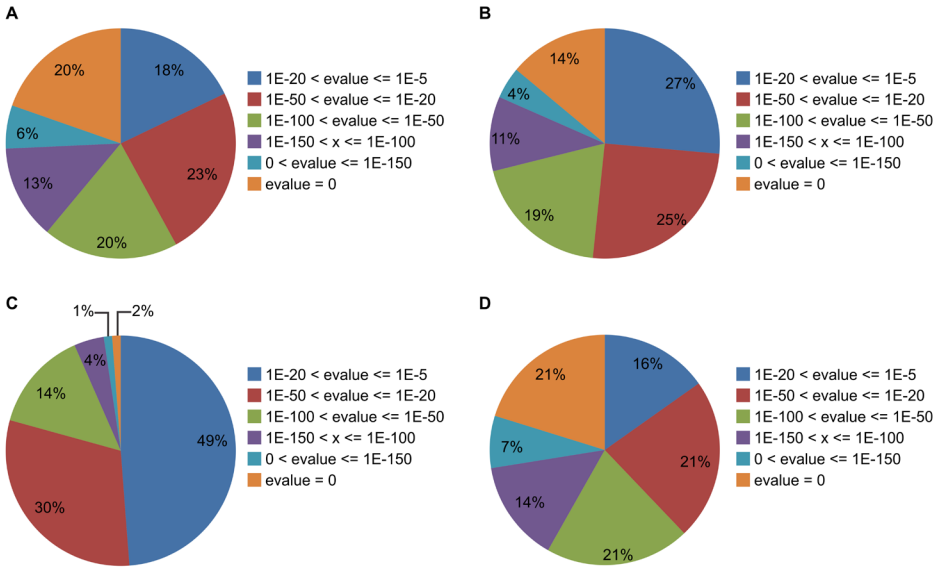

Figure 3. Characteristics of the e-value distributions of BLAST hits for each unigene with an e-value threshold of 10^{-5} in Nr (**A**), Swiss-Prot (**B**), COG (**C**), and KEGG (**D**) databases.

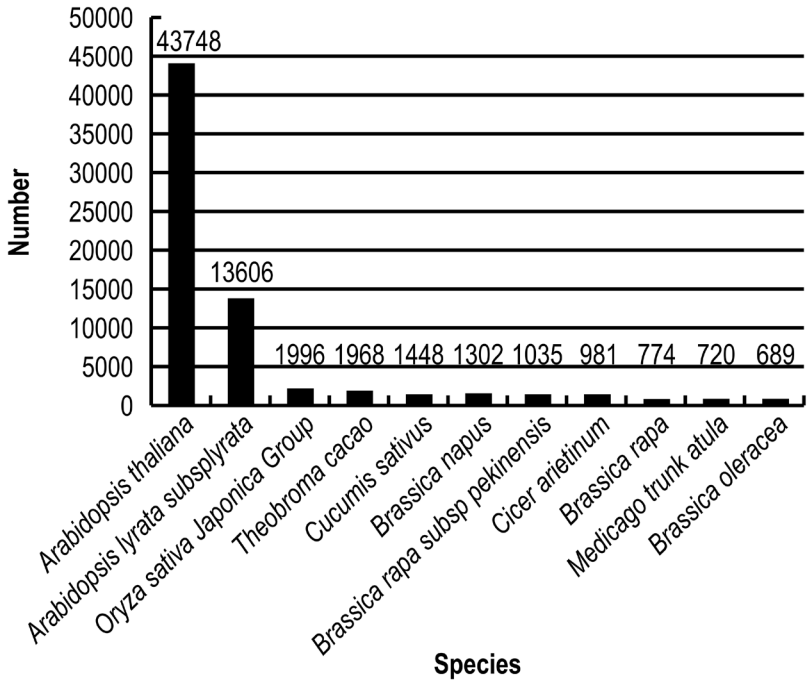

Figure 4. Distribution of the number of predicted proteins by unigenes in different species.

GO classification, COG classification, and KEGG metabolic pathways

The GO classification indicated that 34,609 of all the unigenes could be divided into three main functional groups and 43 sub-functional groups (Table 3). The three main functional groups were biological process (78,341 unigenes; 42.09%), cellular component (70,022 unigenes; 37.62%), and molecular function (37,778 unigenes; 20.3%), respectively. Within the biological process functional group, unigenes were most commonly assigned to cellular process (16,612 unigenes; 21.2%) and metabolic process (17,036 unigenes; 21.75%). Within the cellular component functional group, unigenes were mainly grouped in cell processes (22,153 unigenes; 31.64%), cellular sections (22,153 unigenes; 31.64%), and organelle (14,805 unigenes; 21.14%). Finally, the unigenes within the molecular function group, were mainly assigned to protein binding (18,636 unigenes; 49.33%) and catalytic activity (15,273 unigenes; 40.43%). The results of the COG classification showed that 29,007 unigenes were divided into 25 functional groups (Figure 5); of these, the largest group was the general function prediction group (10,219 unigenes; 17%), and the two second largest groups were the transcriptome group (5779 unigenes, 9.6%) and the replication-homologous combination-repair group (4899 unigenes; 8.1%). The result of the KEGG biological pathway analysis indicated that 25,033 unigenes were assigned to 124 biological pathways; of these, 5716 unigenes (22.83%) were assigned to the metabolic pathways group, 2782 unigenes (11.11%) were assigned to the secondary metabolite synthesis pathway, 888 unigenes (3.55%) were assigned to the plant hormone signal transduction pathway, and 813 unigenes (3.25%) were assigned to ribosomal pathways.

Table 3. Gene ontology (GO) functional classification.

GO	Class	No.	GO	Class	No.
Biological process	Anatomical structure formation	471	Cellular component	Cell	22,153
	Biological adhesion	4		Cell part	22,153
	Biological regulation	7,254		Envelope	1,523
	Cell killing	2		Extracellular region	635
	Cellular component biogenesis	667		Extracellular region part	37
	Cellular component organization	2,402		Macromolecular complex	3,436
	Cellular process	16,612		Membrane-enclosed lumen	363
	Death	146		Organelle	14,805
	Developmental process	3,977		Organelle part	4,917
	Establishment of localization	3,287	Molecular function	Antioxidant activity	144
	Growth	513		Auxiliary transport protein activity	8
	Immune system process	317		Binding	18,636
	Localization	3,660		Catalytic activity	15,273
	Locomotion	20		Electron carrier activity	14
	Metabolic process	17,036		Enzyme regulator activity	497
	Multi-organism process	1,125		Molecular transducer activity	409
	Multicellular organismal process	2,841		Structural molecule activity	587
	Pigmentation	6,272		Transcription regulator activity	32
	Reproduction	1,455		Translation regulator activity	283
	Reproductive process	1,424		Transporter activity	1,895
	Response to stimulus	8,643			
	Rhythmic process	174			
	Viral reproduction	39			

Differential gene expression and clubroot-resistant unigene

In the material that was inoculated or not inoculated with clubroot fungus, there were

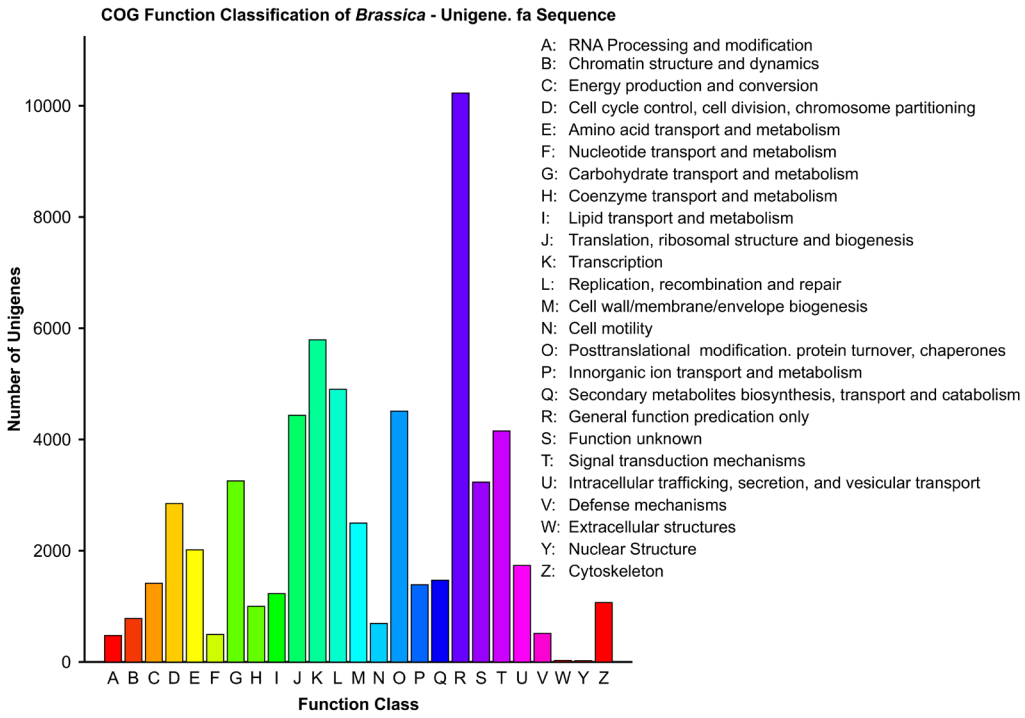

Figure 5. Orthologous groups based on COG classification of unigenes from the *Brassica napus* transcriptome.

large amounts of differentially expressed unigenes (Figure 6). Between treatments Y2 and Y1 (inoculated Huashuang 3 and Zhongshuangyou 8), the number of up-regulated unigenes (18,567) was two times higher than the number of down-regulated unigenes. Among the up- and down-regulated unigenes, 9381 (50.53%) and 2565 (28.95%), respectively, were induced specifically in Y2 (Table 4). Between treatments Y2 and Y3 (inoculated and un-inoculated Huashuang 3), the number of up-regulated unigenes was 10,844, which was almost three time higher than the number of down-regulated unigenes. Among the up- and down-regulated unigenes, 9053 (83.48%) and 240 (6.18%), respectively, were induced specifically in Y2 (Table 5). Thus, there were considerably fewer up- and down-regulated unigenes between Y2 and Y3 than between Y2 and Y1. However, among the up-regulated unigenes, the number of specifically expressed unigenes between Y2 and Y3 were similar to those between Y2 and Y1. Moreover, the number of specifically expressed unigenes among the down-regulated unigenes between Y2 and Y3 were 10 times higher than that between Y2 and Y1 (Table 5). This indicates that the differentially expressed genes between Y2 and Y1 mainly relate to species differences. The specifically expressed down-regulated genes were mainly related to the *B. napus* biological processes during clubroot fungus infection, whereas the specifically expressed up-regulated genes could be disease-resistant genes induced by clubroot fungus infection.

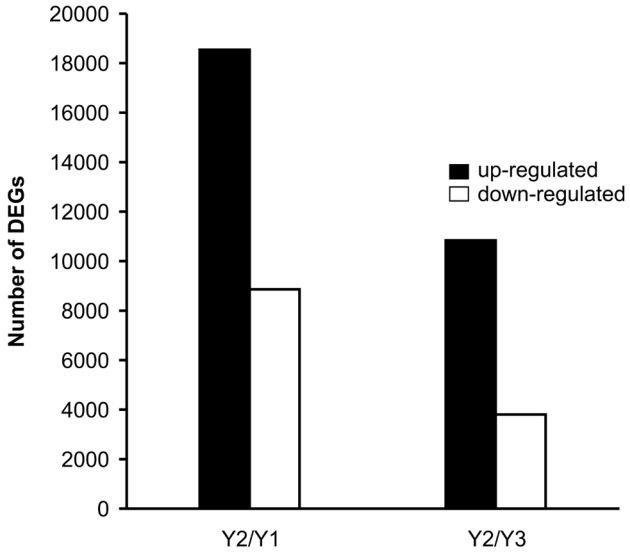

Figure 6. Number of differentially expressed genes (DEGs) in the different treatments.

Table 4. Differential gene expression between treatments Y2 and Y1.

Log2 ratio (Y2/Y1)	Up-regulation (Y2/Y1)		Down-regulation (Y2/Y1)	
	No. of differentially expressed genes	No. of genes with expression specificity	No. of differentially expressed genes	No. of genes with expression specificity
Log2 ratio >18	28	28	1	1
16< Log2 ratio ≤18	147	147	21	21
14< Log2 ratio ≤16	934	934	179	178
12< Log2 ratio ≤14	4,969	3,965	816	814
10< Log2 ratio ≤12	4,995	3,912	1,321	1,307
5< Log2 ratio ≤10	778	395	798	244
1< Log2 ratio ≤5	6,716	0	5,724	0
Total	18,567	9,381	8,860	2,565

Table 5. Differential gene expression between treatments Y2 and Y3.

Log2 ratio (Y2/Y3)	Up-regulation (Y2/Y3)		Down-regulation (Y2/Y3)	
	No. of differentially expressed genes	No. of genes with expression specificity	No. of differentially expressed genes	No. of genes with expression specificity
Log2 ratio >18	58	58	0	0
16< Log2 ratio <18	208	208	0	0
14< Log2 ratio <16	1,199	1,197	11	11
12< Log2 ratio <14	4,573	4,567	42	42
10< Log2 ratio <12	2,938	2,911	137	134
5< Log2 ratio <10	187	112	209	53
1< Log2 ratio <5	1,681	0	3,482	0
Total	10,844	9,053	3,881	240

In order to verify whether the unigenes were related to *B. napus* clubroot-resistance, we aligned the sequences of unigenes with an FDR-value ≤0.001 and an absolute Log2 ratio-value ≥2 to the sequences of the R gene database (http://prgdb.crg.eu/wiki/Main_Page). The unigenes with broad-spectrum or specific resistance, such as Crr1 and CRa were identified, and the expression of these resistance genes was found to be induced specifically in Y2. Their

RPKM-values ranged from 0.023 to 90.634 (Table 6). Based on the relationship between clubroot-resistance in Brassicaceae and the inhibition of the IAA signal transduction, cytokinin synthesis, and myrosinase synthesis, which has been reported in previous studies, we selected the resistance unigenes with down-regulated expression and specifically induced expression from these biological pathways, and aligned their sequences with the NCBI database gene sequences. We found relatively longer unigenes with greater coverage depth and high homology (Table 7), indicating that Huashuang 3 contains many genes related to clubroot-resistance. The high clubroot-resistance of Huashuang 3 was due to the highly induced expression of broad-spectrum disease-resistant genes, specific clubroot-resistant genes, as well as a strong inhibition of the IAA signal transduction, cytokinin synthesis, and myrosinase synthesis. The result of the NCBI online sequence alignment illustrated that at 88% coverage, the base sequences of unigene 0048482 are similar with the 3' end of the Crr1 sequence (base 2686 to 3699) with 94% similarity. At 97% coverage, the sequence of unigene 0061770 was similar with the negative chain sequence of Crr1 5' end (base 1 to 2126) with 94% similarity. These findings indicate that unigene 0061770 originates from an inversion of the 5' end sequence of Crr1, and that unigenes 0048482 and 0061770 together constitute a clubroot-resistant functional gene in Huashuang 3.

Table 6. Unigenes related to disease-resistance and their induced expression.

Disease-resistant gene	No. of unigenes	RPKM max in Y2	RPKM min in Y2	Mean RPKM in Y2	GenBank accession No.
Crr1	4	8.596	0.520	4.360	AB605024
CRa	2	7.084	6.605	6.845	AB751517
Bs2	1	52.526	52.526	52.526	6456754
Bs4	2	3.685	0.926	2.305	38489218
I2	5	1.754	0.849	1.194	4689222
RY-1	3	2.734	0.494	1.716	16944810
KR1	10	5.032	0.148	2.830	18033110
Lr34	24	90.634	0.083	18.528	301130794
M	1	2.623	2.623	2.623	1842250
Pi9	5	2.065	0.397	0.962	83571777
R3a	1	2.873	2.873	2.873	57233496
RAC1	11	9.221	0.638	4.281	41387773
SSI4	28	9.859	0.254	2.010	27466163
RCY1	8	19.216	0.041	4.174	29603481
Rdg2a	1	56.879	56.879	56.879	301015479
RLM3	2	2.539	0.854	1.697	79325134
Rpi-blb2	1	11.322	11.322	11.322	74040323
RPM1	2	2.160	0.596	1.378	30680118
RPP1	66	13.540	0.021	3.428	30692150
RPP4	2	2.523	0.350	1.437	42566890
Sw-5	1	0.439	0.439	0.439	15418708
RPP5	10	3.469	0.131	1.626	186511938
RPP8	8	3.677	0.166	1.865	145358807
Rps1-k-2	1	5.332	5.332	5.332	183396496
Rps4	16	17.997	0.164	6.223	18422530
RPS5	28	14.409	0.023	4.370	145335420
RRS1	11	9.186	0.744	3.077	145334738
Rx2	2	0.776	0.024	0.400	5918253

RPKM: reads per kilobase per million reads.

RT-PCR validation of digital gene expression

In order to verify the accuracy of the digital gene expression data, six unigenes related

to clubroot-resistance were selected for qRT-PCR analysis. In treatment Y2, the expression of unigenes 0035403 and 0091816 (IAA synthase genes), unigene 0027251 (cytokinin synthase gene), and unigene 0011332 (myrosinase gene) were down-regulated, whereas the expression of unigene 0007461 (Crr1 gene) was specifically induced. The qRT-PCR analysis revealed that the expression of unigene 0061770 (Crr1 gene) was more than 3370 times more up-regulated in Y2 than in Y3 (Figure 7 and Table 8). These results confirmed the validity of our transcriptome sequencing data.

Table 7. Main characteristics of unigenes related to the resistance to clubroot in *Brassica napus*.

Unigene No.	Protein	Length	Y2 RPKM	Y3 RPKM	Log2 ratio (Y2/Y3)	Query cover (%)	Identity (%)	E value
0035403	IAA synthase	730	8.057	50.530	-2.51	91	95	$5.00E^{-42}$
0036121		1642	5.024	237.855	-5.57	81	87	$1.00E^{-101}$
0091816		792	2.964	261.812	-6.46	82	89	$8.00E^{-58}$
0027251	Cytokinin synthase	826	4.682	78.158	-4.21	82	90	$2.00E^{-12}$
0085190		733	3.242	115.927	-5.29	82	89	$4.00E^{-12}$
0011332	Myrosinase	1099	4.902	276.409	-4.00	84	88	$1.00E^{-156}$
0036814		1337	2.736	58.4294	-5.40	80	86	$1.00E^{-99}$
0087544		1837	8.572	152.571	-5.01	81	87	$3.00E^{-56}$
0007461	Crr1	820	8.596	0.001	13.07	83	80	$6.00E^{-98}$
0061770		2184	4.179	0.001	12.02	97	94	$0E^{+00}$
0093499		1691	0.520	0.001	9.02	98	78	$3E^{-78}$
0048482		1206	4.158	0.001	12.01	88	94	$0E^{+00}$
0006907	CRa	963	7.084	0.001	12.79	92	86	$1.00E^{-35}$
0061763		1002	6.605	0.001	12.69	85	83	$4.00E^{-28}$

RPKM: reads per kilobase per million reads.

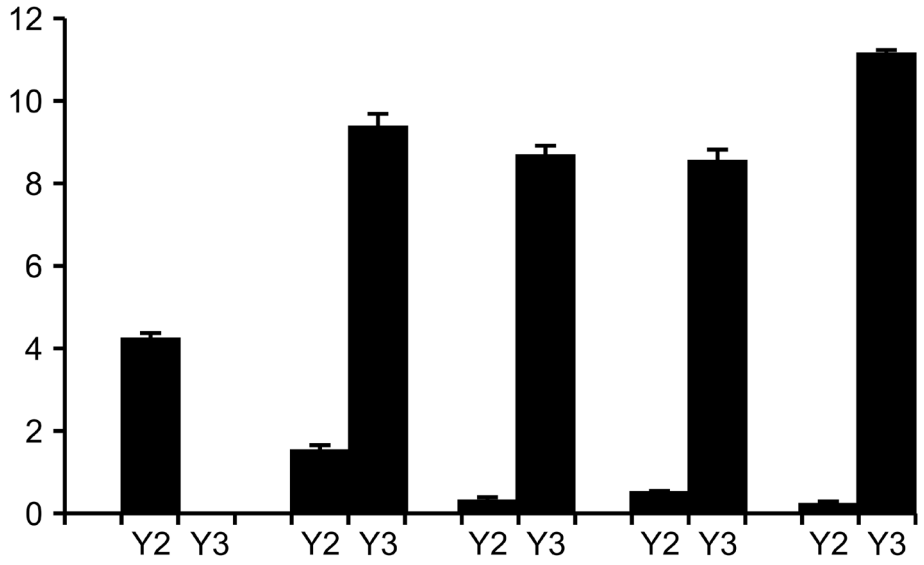

Figure 7. Results of relative quantification for the expression of five unigenes in different treatments.

Table 8. Results of absolute quantification of unigene 0007461.

Unigene 0007461	Well	Ct (dR)	Quantity (copies)	Average
Y2	B10	23.57	$6.06E^{+05}$	$7.69E^{+05}$
	B11	23.2	$7.92E^{+05}$	
	B12	23.02	$9.10E^{+05}$	
Y3	C10	35.34	$2.20E^{+02}$	$2.28E^{+02}$
	C11	34.14	$1.20E^{+02}$	
	C12	36.44	$3.45E^{+02}$	

Crr1 unigenes in the genome of clubroot-resistant variety Huashuang 3

Based on the BLASTx alignment, the sequence of unigene 0048482 shared 100% similarity to that of *B. oleracea* clubroot-resistance gene on chromosome 3 (54,090,181-54,091,385 bp), whereas no sequence was found that was similar to the unigene 0048482 sequence. To verify that unigenes 0048482 and 0061770 are two components of the same gene in the genome of the resistant variety Huashuang 3, we designed primers based to the unigene sequences and the middle sequence of Crr1 (not common between Crr1 and the unigenes), and then performed PCR amplification of genomic DNA (Table 9). The agarose gel electrophoresis results (Figure 8) indicated that the target DNA fragment was amplified from Huashuang 3 genomic DNA, but not from Zhongshuangyou 8. The lengths of the amplified DNA fragments using primer combinations F1/R3, F2/R3, and F2/R4 were longer than the expected fragment from Huashuang 3 genomic DNA. This indicates that there are introns in the sequences between the primer pairs, which can be proved by PCR sequencing. Using primer combinations F3/R3 and F3/R4, the target DNA fragments were not amplified from Huashuang 3 genomic DNA. This was probably due to the sequence characteristics of primer F3: the sequences of primer F3 may include the 3' sequences of the upstream exon and the 5' sequences of the downstream exon. It is possible that two exons are adjacent and separated

Table 9. Primers and their sequences designed based on unigene sequences and the middle sequence of the Crr1 gene.

Gene	Primer code	Primer sequence (5'-3')	Nucleotide site
Unigene 0061770	F1	ATTTCCCTACCCAAACGTGCTA	466-487
	F2	AATAATCCGACTTCCACTGCC	984-1004
	F3	AATGGTAACCAGCGGGACAC	1496-1515
	F4	ACCAACTCGTCAAGGCACCAC	1688-1708
	R1	GCTATGATGCCTTATGCGATGA	681-660
	R2	CTCCTGACTTTCAACTTAGCG	1258-1238
	R3	TCACCTCCACCTTCTCCTTCA	1995-1975
	R4	CAGAACCTGAGCCAAATCCAAG	2024-2003
	F8	CTCAAGTGGATGGATCTGTCTCA	2132-2154
	F9	GCTTCCTAATCTTTCAACTGCCA	2115-2137
Unigene 0048482	F5	CCCTCTTCTATTGGGAATCTTCA	1-23
	F6	ATTATTTCAGTTACAGAGCCACGC	612-635
	F7	CGTTCTGAGCATCTGGTCGTA	845-865
	R5	AAGCATGATGCGATTTCCTTAGG	287-265
	R6	ATTACGACCAGATGCTCAGAACG	867-845
	R7	CTCAACGAGTAGGAAACAAAGGT	1184-1162
	R8	CTTCAAATTATGAAGATTCCCAATAG	33-8
Unigene 0061770 and Crr1	F9	GCTTCCTAATCTTTCAACTGCCA	2115-2137
Crr1	F8	CTCAAGTGGATGGATCTGTCTCA	2132-2154
	R9	CATCCGGTGAGATCCAACCTC	2595-2575

by an intron in the genome DNA of *B. napus*. Primer combinations F8/R8 and F9/R9 were designed based on the 3'-end sequence of unigene 0061770, the 5'-end sequence of unigene 0048482, and the middle sequence of Crr1. These primers amplified DNA fragments from Huashuang 3 genomic DNA, but not from Zhongshuangyou 8. In the process of *P. brassicae* infection, the expression of two genes, unigenes 0061770 and 0048482, are up-regulated in Huashuang 3 (Table 7). Together, these results indicate that these two unigenes are two components of a single functional gene.

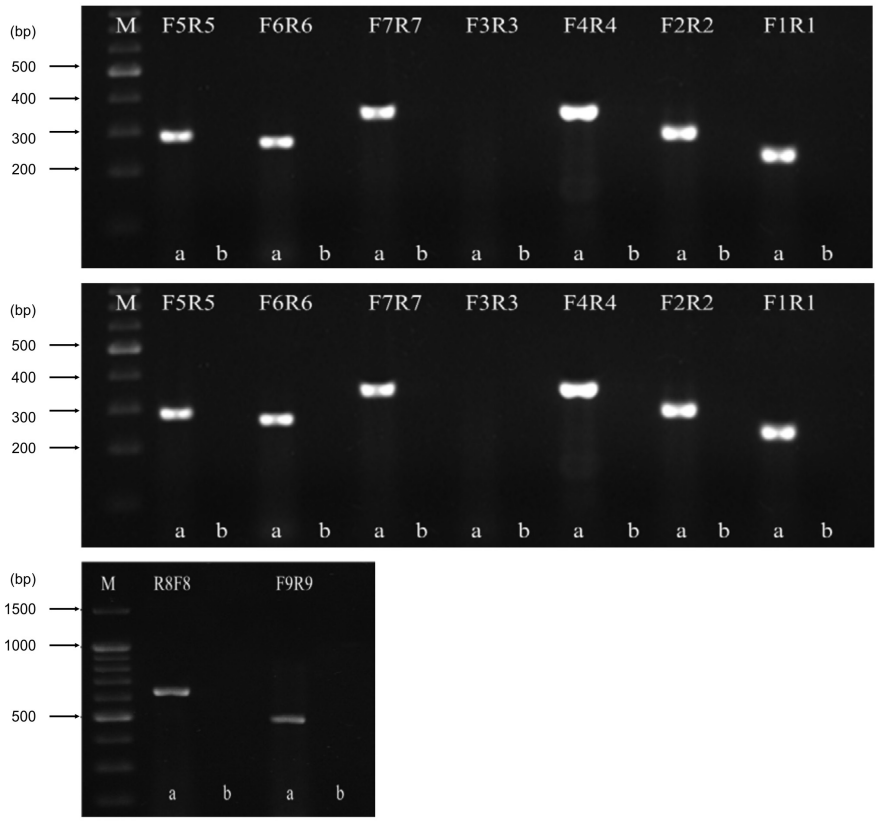

Figure 8. Agarose gel electrophoresis diagram of PCR products amplified from genomic DNA of Huashuang 3 and Zhongshuangyou 8. M, a, and b indicate the agarose gel electrophoresis results of the DNA marker, the products amplified from Huashuang 3, and Zhongshuangyou 8 genomic DNA, respectively.

DISCUSSION

Since the development of RNA-Seq technology, researchers have been able to analyze the transcriptome of many biological organisms, including yeast (Yassour et al., 2009), nematodes (Ng et al., 2009), mice (Mortazavi et al., 2008), human (Wang et al., 2008), and some prokaryotes (Yoder-Himes et al., 2009). This has led to significant advances in the field of molecular research. In this study, over 171 million high-quality reads were obtained using

a high-throughput Illumina sequencing technique. The rate of reads with Q20-value reached 98.68%. Based on these reads, 96,149 unigenes were assembled with an average length of 973 bp, which was larger than the mean unigene lengths found in butterfly (197 bp; Vera et al., 2008), sweet potato (581 bp; Guo et al., 2011), *Salvia miltiorrhiza* (331 bp; Zhang et al., 2012), safflower (446 bp; Lulin et al., 2012), and wax gourd (709 bp; Jiang et al., 2013). We aligned the unigene sequences to the Nr, Swiss-Prot, GO, COG, and KEGG protein databases and found that most unigene sequences could be matched to unique protein sequences in the various databases. This suggests that there are many homologous genes between the transcriptome of *B. napus* and other species. Numerous unigenes were assigned to detailed GO and COG function classifications, indicating that our *B. napus* transcriptome database has both universality and diversity. By aligning the unigene sequences to the KEGG database, 25,033 unigenes were found to belong to 124 biological processes. Interestingly, many unigenes were related to unsaturated fatty acid biosynthesis and vitamin metabolism.

Hao et al. (2011) conducted transcriptome sequencing on different *Taxus* tissues (root, stem, and leaf), and validated the gene expression differences among tissues by differential gene expression. Wang et al. (2010) sequenced the transcriptome from the tuber of sweet potato (*Ipomoea batatas*) and analyzed the expression of genes related to tuber development. Li et al. (2010) sequenced and analyzed the transcriptome from maize leaves, and investigated the expression of genes related to leaf angle and photosynthesis. To our knowledge, the present study is the first to perform large-scale transcriptome sequencing to analyze the physiological and biochemical processes associated with clubroot-resistance and the induced gene expression in *B. napus*. Our study identified some genes related to clubroot-resistance, and, thus, we were able to provide some evidence for the molecular mechanisms of *B. napus* resistance to clubroot fungus related to physiology, biochemistry, and molecular biology.

Next-generation Illumina transcriptome sequencing provides excessive data with high accuracy, high efficiency, and at low cost, and is especially advantageous for functional gene identification. Tang et al. (2011) conducted transcriptome sequencing in *Siraitia grosvenorii* and identified many genes related to the biosynthesis of *S. grosvenorii* glycoside. Crawford et al. (2010) performed sequencing and analysis of the transcriptome on *Anopheles funestus*, and found some unigenes related to immune genes *APL1* and *LRIM1*. Lin et al. (2012) used transcriptome sequencing technology to obtain the complete cDNA sequence of the camellia ubiquitin-conjugating enzyme gene, which was named UBE2S and encoded a 270-amino acid protein. Zenoni et al. (2010) used transcriptome sequencing technology to study the complicated transcription features during fruit development in *Vitis vinifera* (grape), and found 53 new genes of the glutathione-*S*-transferase gene family and 28 new genes of the myoglobin B transcription factor gene family. In this study, we identified some broad-spectrum disease-resistant genes as well as specific resistant genes (Crr1 and CRa) related to clubroot-resistance. We found that their expression levels were remarkably up-regulated. These results provide new information and theoretical evidence that future studies may use in the breeding of clubroot-resistant *B. napus*. Using proteome analysis techniques, Cao et al. (2008) found that the contents of 20 enzymes in infected *B. napus*, which participate in the lignin synthesis, cell division, and glycolysis, are all different during clubroot fungus infection. These enzymes are related to the resistance to *B. napus* clubroot. By using a complete *Arabidopsis* transcriptome microarray chip, Agarwai et al. (2011) found that the expression levels of genes encoding enzymes that participate the synthesis process of lignin, cytokinin, and hormones, are different

in *Arabidopsis*. The results reported by Devos and Prinsen (2006) showed that blocking the IAA signal transduction could hinder clubroot development, and that a reduction of cytokinin content can inhibit the expansion of tumors in plant roots. Devos and Prinsen (2006) conducted a gene microarray analysis of *A. thaliana* infected with clubroot, and found that myrosinase expression was down-regulated. We found that the trend in gene expression levels related to IAA, cytokinin synthesis, and myrosinase in resistant material were similar to that reported in the studies. This indicates that the inhibition of gene expression related to IAA and cytokinin synthesis is related to clubroot-resistance.

Crr1 has been located in turnip (*B. rapa* variety *rapa*) genome (Suwabe et al., 2006). Using transcriptome sequencing and bioinformatic analyses, we found two homologous unigenes of the Crr gene in Huashuang 3: unigenes 0061770 and 0048482. The sequences of these unigenes were consistent with the 5'-end negative chain sequences (base 1-2126) and the 3'-end sequences (base 2686-3699) of Crr1, respectively. The PCR amplification products from Huashuang 3 genomic DNA confirmed that unigenes 0061770 and 0048482 existed in the Huashuang 3 genome and formed a functional gene. Although unigene 0061770 originated from the inverse 5'-end sequence of Crr1, it may still be functional.

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

Agarwal A, Kaul V, Faggian R, Rookes JE, et al. (2011). Analysis of global host gene expression during the primary phase of the *Arabidopsis thaliana-Plasmodiophora brassicae* interaction. *Funct. Plant Biol.* 38: 462-478.

Cao T, Srivastava S, Rahman MH, Kav NNV, et al. (2008). Proteome-level changes in the roots of *Brassica napus* as a result of *Plasmodiophora brassicae* infection. *Plant Sci.* 174: 97-115. http://dx.doi.org/10.1016/j.plantsci.2007.10.002

Crawford JE, Guelbeogo WM, Sanou A, Traoré A, et al. (2010). *De novo* transcriptome sequencing in *Anopheles funestus* using Illumina RNA-seq technology. *PLoS One* 5: e14202. http://dx.doi.org/10.1371/journal.pone.0014202

Devos S and Prinsen E (2006). Plant hormones: a key in clubroot development. *Commun. Agric. Appl. Biol. Sci.* 71 (3 Pt B): 869-872.

Guo S, Liu J, Zheng Y, Huang M, et al. (2011). Characterization of transcriptome dynamics during watermelon fruit development: sequencing, assembly, annotation and gene expression profiles. *BMC Genomics* 12: 454. http://dx.doi.org/10.1186/1471-2164-12-454

Hao C, Ge G, Xiao P, Zhang Y, et al. (2011). The first insight into the tissue specific *taxus* transcriptome via Illumina second generation sequencing. *PLoS One* 6: e21220. http://dx.doi.org/10.1371/journal.pone.0021220

Hatakeyama K, Suwabe K, Tomita RN, Kato T, et al. (2013). Identification and characterization of Crr1a, a gene for resistance to clubroot disease (*Plasmodiophora brassicae* Woronin) in *Brassica rapa* L. *PLoS One* 8: e54745. http://dx.doi.org/10.1371/journal.pone.0054745

Hirai M, Harada T, Kubo N, Tsukada M, et al. (2004). A novel locus for clubroot resistance in *Brassica rapa* and its linkage markers. *Theor. Appl. Genet.* 108: 639-643. http://dx.doi.org/10.1007/s00122-003-1475-x

Jiang B, Xie D, Liu W, Peng Q, et al. (2013). *De novo* assembly and characterization of the transcriptome, and development of SSR markers in wax gourd (*Benicasa hispida*). *PLoS One* 8: e71054. http://dx.doi.org/10.1371/journal.pone.0071054

Johnston TD (1968). Clubroot in *Brassica* a standard inoculation technique and the specification of races. *Plant Pathol.* 17: 184-187. http://dx.doi.org/10.1111/j.1365-3059.1968.tb00452.x

Jubault M, Lariagon C, Taconnat L, Renou JP, et al. (2013). Partial resistance to clubroot in *Arabidopsis* is based on changes in the host primary metabolism and targeted cell division and expansion capacity. *Funct. Integr. Genomics* 13: 191-205. http://dx.doi.org/10.1007/s10142-013-0312-9

Li P, Ponnala L, Gandotra N, Wang L, et al. (2010). The developmental dynamics of the maize leaf transcriptome. *Nat.*

Genet. 42: 1060-1067. http://dx.doi.org/10.1038/ng.703
Li R, Yu C, Li Y, Lam TW, et al. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25: 1966-1967. http://dx.doi.org/10.1093/bioinformatics/btp336
Lin P, Yao X, Wang K, Cao Y, et al. (2012). Characterization of ubiquitin-conjugating enzyme E2S from *Camellia oleifera. Mol. Plant Breed.* 10: 1376-1382.
Lulin H, Xiao Y, Pei S, Wen T, et al. (2012). The first Illumina-based *de novo* transcriptome sequencing and analysis of safflower flowers. *PLoS One* 7: e38653. http://dx.doi.org/10.1371/journal.pone.0038653
Manzanares-Dauleux MJ, Delourme R, Baron F and Thomas G (2000). Mapping of one major gene and of QTLs involved in resistance to clubroot in *Brassica napus. Theor. Appl. Genet.* 101: 885-891. http://dx.doi.org/10.1007/s001220051557
Matsumoto E, Yasui C, Ohi M and Tsukada M (1998). Linkage analysis of RFLP markers for clubroot resistance and pigmentation in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Euphytica* 104: 79-86. http://dx.doi.org/10.1023/A:1018370418201
Mortazavi A, Williams BA, McCue K, Schaeffer L, et al. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5: 621-628. http://dx.doi.org/10.1038/nmeth.1226
Ng SB, Turner EH, Robertson PD, Flygare SD, et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461: 272-276. http://dx.doi.org/10.1038/nature08250
Pertea G, Huang X, Liang F, Antonescu V, et al. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19: 651-652. http://dx.doi.org/10.1093/bioinformatics/btg034
Piao ZY, Ramchiary N and Lim YP (2009). Genetics of clubroot resistance in *Brassica* species. *J. Plant Growth Regul.* 28: 252-264. http://dx.doi.org/10.1007/s00344-009-9093-8
Porebski S, Bailey LG and Baum BR (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Report.* 15: 8-15. http://dx.doi.org/10.1007/BF02772108
Sakamoto K, Saito A, Hayashida N, Taguchi G, et al. (2008). Mapping of isolate-specific QTLs for clubroot resistance in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Theor. Appl. Genet.* 117: 759-767. http://dx.doi.org/10.1007/s00122-008-0817-0
Siemens J, Keller I, Sarx J, Kunz S, et al. (2006). Transcriptome analysis of *Arabidopsis* clubroots indicate a key role for cytokinins in disease development. *Mol. Plant Microbe Interact.* 19: 480-494. http://dx.doi.org/10.1094/MPMI-19-0480
Stewart Jr CN and Via LE (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques* 14: 748-750.
Strelkov SE, Tewari JP and Smith-Degenhardt E (2006). Characterization of *Plasmodiophora brassicae* populations from Alberta, Canada. *Can. J. Plant Pathol.* 28: 467-474. http://dx.doi.org/10.1080/07060660609507321
Suwabe K, Tsukazaki H, Iketani H, Hatakeyama K, et al. (2003). Identification of two loci for resistance to clubroot (*Plasmodiophora brassicae* Woronin) in *Brassica rapa* L. *Theor. Appl. Genet.* 107: 997-1002. http://dx.doi.org/10.1007/s00122-003-1309-x
Suwabe K, Tsukazaki H, Iketani H, Hatakeyama K, et al. (2006). Simple sequence repeat-based comparative genomics between Brassica rapa and Arabidopsis thaliana: the genetic origin of clubroot resistance. *Genetics* 173: 309-319. http://dx.doi.org/10.1534/genetics.104.038968
Tanaka S, Mido H and Ito S (2006). Colonization by two isolates of *Plasmodiophora brassicae* with differing pathogenicity on a clubroot-resistant cultivar of Chinese cabbage (*Brassica rapa* L. subsp. *pekinensis*). *J. Gen. Plant Pathol.* 72: 205-209. http://dx.doi.org/10.1007/s10327-006-0276-x
Tang Q, Ma X, Mo C, Wilson IW, et al. (2011). An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics* 12: 343. http://dx.doi.org/10.1186/1471-2164-12-343
Ueno H, Matsumoto E, Aruga D, Kitagawa S, et al. (2012). Molecular characterization of the CRa gene conferring clubroot resistance in *Brassica rapa. Plant Mol. Biol.* 80: 621-629. http://dx.doi.org/10.1007/s11103-012-9971-5
Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, et al. (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol. Ecol.* 17: 1636-1647. http://dx.doi.org/10.1111/j.1365-294X.2008.03666.x
Wang ET, Sandberg R, Luo S, Khrebtukova I, et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470-476. http://dx.doi.org/10.1038/nature07509
Wang Z, Fang B, Chen J, Zhang X, et al. (2010). *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics* 11: 726. http://dx.doi.org/10.1186/1471-2164-11-726
Yassour M, Kaplan T, Fraser HB, Levin JZ, et al. (2009). *Ab initio* construction of a eukaryotic transcriptome by massively

parallel mRNA sequencing. *Proc. Natl. Acad. Sci. USA* 106: 3264-3269. http://dx.doi.org/10.1073/pnas.0812841106

Yoder-Himes DR, Chain PS, Zhu Y, Wurtzel O, et al. (2009). Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc. Natl. Acad. Sci. USA* 106: 3976-3981. http://dx.doi.org/10.1073/pnas.0813403106

Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, et al. (2010). Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiol.* 152: 1787-1795. http://dx.doi.org/10.1104/pp.109.149716

Zhang J, Liang S, Duan J, Wang J, et al. (2012). De novo assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). *BMC Genomics* 13: 90. http://dx.doi.org/10.1186/1471-2164-13-90