Transcriptome Analysis of Coriander: A Dual Purpose Crop Unravels Stem Gall Resistance Genes

Sharda Choudhary1*, Mahantesha B. N. Naika2*, Radheshyam Sharma3, R. D. Meena1, Ravindra Singh1, Gopal Lal1

1 ICAR-National Research Centre on Seed Spices, Tabiji-305206, Ajmer, Rajasthan, India
2 Department of Biotechnology and Crop Improvement, K. R. C. College of Horticulture, Arabhavi-591218, University of Horticultural Sciences, Bagalkot, Karnataka, India
3 Biotechnology Centre, Jawaharlal Nehru krishi vishva vidhalya Jabalpur, MP, 482004

* Corresponding authors
E-mail : shardaajmer@yahoo.com (SC), mnaikbt@gmail.com (MBNN)

Abstract

Stem gall (Protomyces macrosorus Unger), a serious disease affects on leaves, petioles, stems and fruits, and causes loss in coriander (Coriandrum sativum L.) grown for fresh cut leaves for making different recipes and seeds as a spice purpose globally. Genetic improvement of coriander for stem gall disease is indispensable. Coriander cultivar’s of stem gall resistance (ACr-1) and susceptible (CS-6) leaf samples were utilized and transcriptome sequenced using Illumina NextSeq500 platform. After trimming low quality reads and adapter sequences, total 4,91,63,108 and 4,37,46,120 high-quality reads retained and further assembly resulted validated transcripts of 59,933 and 56,861. We have predicted 52,506 and 48,858 coding sequences, out of which 50,506 and 46,945 were annotated using NCBI nr database. Gene ontology analysis annotated 19,099 and 17,625 terms; pathway analysis obtained 24 different functional pathway categories; signal transduction, transport, catabolism, translation and carbohydrate metabolism pathways etc. were dominated. Differentially expressed genes analysis predicted 13,123 coding sequences commonly expressed, out of which 431 and 400 genes were significantly up and down-regulated in
which R genes, stress inducible transcription factors such as ERF, NAC, bZIP, MYB, DREB and WRKY and antifungal related genes were predicted. The real time PCR analysis of HSP20 gene expression in resistance showed up-regulation by 10 fold over susceptible sample and 18s used as a house keeping gene for normalization. The present results provide an insights into various aspects underlying the development of resistance to stem gall in coriander.

**Key words:** Coriander, Stem gall, transcriptome, annotation, resistance genes

**Introduction**

Coriander (*Coriandrum sativum* L.) commonly known as dhania, cilantro, Mexican coriander, or Chinese parsley, a hardy annual herb with diploid chromosome number (2n=22) and belonging to the family Apiaceae or Umbelliferae. Though it is native to the mediterranean region but, is grown as an important spice crop in India, Morocco, Russia, Turkey and other parts of the world (Aiyar 1958; Diederichsen 1996). Coriander is used as a vegetable and spice, is called dual purpose crop. In coriander, essential oil, fatty acid, antimicrobial, anti-oxidant, anti-diabatic, anti-dyslipidemic etc. were reviewed by Sahib *et al.* 2013. For the first time in India, stem gall disease of coriander casued by (*Protomyces macrosporus* Unger) was reported by Sydow and Butler 1911. The infection of stem gall disease take place at an early stage of seed germination and the symptoms appear in later stage and deteriorating the seed quality (Shrivastava 1955). It affects on leaves, petioles, stem, and fruits of coriander (Nene and Kharbanda 1967). India's popular coriander varieties, such as RCr-436 and CS-6, have been continuing to reduce production and productivity due to the increase in the severity of the stem gall (Malhotra *et al*. 2016a). In Coriander growing regions of India, this disease reduces crop yield and is a limiting component of successful agriculture. Only few studies are available on molecular aspects in coriander; understanding the metabolism of essential oil in coriander, three developmental stages of mericarps
sequenced using illumina technology and identified CsγTRPS and CsLINS terpene synthase candidate genes and were expressed in bacteria and recombinant protein was purified and these two genes account for majority of essential oil constituents in coriander mericarps (Galata et al. 2014), and Choudhary et al. 2017, utilized simple sequence repeats (SSR) markers of carrot and examined its transferability in cultivars of coriander. Hence, there is no genetic information is available on coriander resistance to stem gall disease. The Next generation sequencing (NGS) technique helping us in sequence and understand the genes involved in molecular, cellular and biological functions in model and non-model plants (Wang et al. 2009; Strickler et al. 2012). The cost of NGS has been decreasing and it is interestingly increasing its utilization by researchers of different fields and helping them to address biological questions in a nutshell. Coriander is exceptionally well known as seed flavor and vegetable, yet constrained transcriptome and genomic information hindering the exploration in it. The transcriptome data generated from the study can be utilized in breeding for stem gall disease resistance in coriander.

Materials and methods

Plant material

In the current study for identification of disease resistance genes, cultivar Ajmer Coriander-1 (ACr-1) was having no symptom of stem gall disease and it has been considered as a source for resistance to stem gall due to its high level of quantitative resistance at field level, where as Coriander Selection (CS-6) was having infection of stem gall as a source of susceptible were selected (Malhotra et al. 2016b). Two months old coriander leaves were utilized for RNA sequence.

RNA extraction, cDNA library preparation and sequencing

Total RNA was isolated from the resistance and susceptible leaf samples using ZR plant RNA Miniprep (ZYMO Research) as per the manufacturer’s instruction. The quality and
quantity of the isolated RNA was checked on 1% denaturing RNA agarose gel and Nanodrop. The RNA-seq paired-end sequencing libraries were prepared from the quality check passed RNA samples using Illumina TruSeq stranded mRNA sample preparation kit. The mRNA was enriched from the total RNA using poly-T attached magnetic beads, followed by enzymatic fragmentation, first strand cDNA was converted using SuperscriptII and Act-D mix to facilitate RNA-dependent synthesis. The first strand cDNA was then synthesized to the second strand using second strand mix. The double-stranded cDNA then purified using Ampure XP beads followed by A-tailing, adapter ligation and then enriched by a limited number of PCR cycles. The PCR enriched libraries was analyzed in 4200 Tape Station System (Agilent Technologies) using high-sensitivity D1000 screen tape as per manufacturer instruction. After obtaining the Qubit concentration for the libraries and the mean peak size from Agilent tape station profile, the paired end (PE) Illumina libraries were loaded onto the high-throughput sequencer NextSeq500 for sequencing. A paired-end sequencing allows the template fragments to be sequenced in both the forward and reverse directions on Nextseq 500. The Kit reagents was used in the binding of samples to complementary adapter oligos on the PE flow cell. The adapters were designed to allow selective cleavage of the forward strands after re-synthesis of the reverse strand during sequencing. The copied reverse strand was then used to sequence from the opposite end of the fragment.

**De novo transcriptome assembly**

The sequenced raw data was processed to obtain high-quality clean reads using Trimmomatic V0.35 (Bolger *et al.* 2014) to remove adapter sequences, ambiguose reads (reads with unknown nucleotides N larger than 5%), and low-quality sequence (reads with more than 10% quality threshold (QV) <20 PHRED score). A minimum length of 50 nt (nucleotide) after trimming was applied. After removing the low-quality sequences from the raw data,
high-quality reads (HQRs) was retained for resistance and susceptible samples respectively. This high quality (QV>20), PE reads were used for De novo assembly of both the samples. Filtered HQRs of resistance and susceptible samples were assembled into transcripts using velvet V1.2.10 (Zerbino 2010) and oases V0.2.09 (Schulz et al. 2012) on optimized Kmer 25. All HQRs were mapped back to their respective assembled transcripts using BWA V0.7.12 (Li and Durbin 2009) for validation of assembled transcripts. The raw transcriptome data was submitted to the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the accession number SRP102969.

**Coding sequence prediction and functional annotation**

The TransDecoder (Haas et al. 2013) was used to predict coding sequences (CDS) from validated transcripts. TransDecoder identifies candidate coding regions within transcript sequences, such as those generated by de novo RNA-Seq transcript assembler. Further, the predicted CDS of the resistance and susceptible samples were searched against protein database (nr) at NCBI using Basic local alignment search tool (BlastX) with an E-value of 1e-05 (Altschul et al. 1990).

**Functional annotation of gene ontology analysis**

The gene ontology (GO) annotation of the CDS was determined by using the Blast2GO program (Conesa and Gotz 2008). The GO assignments were used to classify the functions of the predicted CDS. The GO mapping also provides an ontology of defined terms representing gene product properties which are grouped as a Biological Process (BP), Molecular Function (MF) and Cellular Component (CC). The GO mapping was carried out in order to retrieve GO terms for all the BLASTX functionally annotated CDS. Using GO numbers and functional classification graphs were plotted using WEGO (Web Gene Ontology Annotation) an online tool (Ye et al. 2006).
**Functional annotation of KEGG pathway**

To identify the potential involvement of the predicted CDS of resistance and susceptible samples in biological pathways, CDS were mapped to reference canonical pathways in Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2008). The output of KEGG analysis includes KEGG orthology assignments and corresponding enzyme commission numbers and metabolic pathways using KEGG automated annotation server (KASS) (Moriya et al. 2000).

**Differentially expressed genes (DEGs) analysis**

The HQRs for each sample were mapped to their respective set of CDS using Burrows-Wheeler Alignment (BWA) (Li and Durbin 2009) and aligned for reading count calculation. The common hit accessions based on BLAST search against nr database at NCBI were identified for differentially expressed genes (DEGs) analysis and was performed by employing a negative binomial distribution model (DESeq V1.8.1) (Anders and Huber 2010). The P-value threshold of 0.05 was used to filter statistically significant results. A heat map was constructed using the log-transformed and normalized value of genes based on pearson uncented correlation distance as well as based on complete linkage method. A complete linkage hierarchical cluster analysis was performed on top 100 DEGs using Multiple Experiment Viewer (MEV V4.8.1) (Howe et al. 2011). The DEGs were also presented by scater plot and volcano plot.

**Quantitative Real-Time PCR (qRT-PCR) Analysis**

To verify the results of transcriptome analysis, differentially expressed gene HSP 20 (Primers, Forward (F): GCGCACATGGATTACTCGGA and Reverse (R): TGTGTTCCAAAGCAGGGACC) was selected for qRT-PCR analysis. The 18s gene (F: GTGGGCGATTTGTCTGGTT and R: TGTACAAAGGGCAGGGACGT) was used as internal control (Housekeeping gene) (Nicot et al. 2005) in gene expression studies for the
normalization. The RNA and oligo dT were used for the first strand cDNA synthesis by Reverse transcriptase using kit method (Thermo scientific). The Real Time PCR reaction volume of 20 ul containing 2 ul of cDNA and 10 ul of SYBR Green Supermix (Bio Rad, USA) and was carried out for 35 cycles followed by Denaturation 95 ºC for 30 sec, Annealing 52-58 ºC (Gradient) for 30 sec extension for 72 ºC for 15 sec using Bio-Rad CFX96 system. The Ct values of the HSP20 induced gene in resistance and susceptible samples were calculated and the data was expressed in terms of fold change over control sample.

Results

Transcriptome sequencing and De novo assembly
To identify DEGs in the stem gall resistance and susceptible cultivars of coriander, RNA samples and sequencing libraries were prepared and transcriptome sequenced using illumina sequencing platform. The sequence raw data was processed, totally 4,91,63,108 and 4,37,46,120 high-quality reads with 7.4 Gb and 6.5 Gb of clean data were obtained in resistance and susceptible samples respectively. Further, clean reads assembled by following the De novo method using velvet and oases softwares and resulted totally 59,933 and 56,861 validated transcripts with a N50 of 1,894 bp and 1,730 bp with transcript length sizes ranged from minimum of 200 bp to maximum of 13,133 for resistance and susceptible samples respectively (Table 1).

Coding sequence (CDS) prediction and functional annotation
The TransDecoder identified, 52,506 and 48,858 CDS with minimum length of 297 bp to maximum length of 9114 bp for resistance and susceptible samples respectively. Further, the CDS using BLASTX against nr database resulted, 50,506 CDS with Blast hit with GO annotation of 36.73 % (in which CDS with and without GO annotation were 19,099 and 33,407) and 46,945 CDS with BLAST hit with GO annotation of 36.07 % (in which CDS
with and without GO annotation were 17,625 and 31,233) in resistance and susceptible sample respectively. The majority of hits were observed to be against *Vitis vinifera* followed by *Sesamum indicum*, *Coffea canephora*, *Cynara cardunculus* var. *Scolymus*, *Nicotiana sylvestries*, *Theobroma cacao*, *Jatropha curcas*, *citrus sinensis*, *Nelumbu nucifera* etc. for both the samples (Supplementary Table S1).

**Functional annotation of gene ontology**

The Blast2GO program resulted, total 39,750 and 36,428 GO terms and these were further grouped into three different functional catagories for resistance and susceptible samples respectively (Supplementary Table S2_A and S2). In the molecular function; 15,711 and 14,372 GO terms were obtained, in that, catalytic activity, nucleotide binding, transport activity, structural molecule, molecular transducer, transcription regulator, enzyme regulator, electron carrier, antioxidant, translation regulator, nutrient reservoir, metallochaperones were the terms that dominated. In the biological process;13,663 and 12,520 GO terms were obtained, in that, metabolic process, cellular process, biological regulation, establishment of localization, localization, pigmentation, response to stimulus, cellular component organization, cellular component biogenesis, anatomical structure formation, developmental process, multicellular organismal process, reproduction, reproductive process, multiorganism process, reproduction, reproductive process, multi-organism process, growth, death, immuune system process, rhythmic process, biological adhesion etc. In the cellular component; 10,376 and 9,536 GO terms were obtained, in that, cell, cell wall, organelle, macromolecular complex, organelle part, envelope, membrane-enclosed lumen, extracellular region, symplast, virion, extracellular region part were the terms that dominated in resistance and susceptible samples respectively. The GO numbers and functional classification graphs were plotted using WEGO tool (Figs 1A and B).
Functional annotation of KEGG pathway

The KEGG pathways analysis results were classified under five categories: Metabolism (4,534), Genetic information processing (2,319), Environmental processing (1,041), Cellular process (1,309) and Organismal systems (127). Among, total 24 different functional KASS pathway categories were found; in that, signal transduction, translation, carbohydrate metabolism, folding, sorting, degradation, transport and catabolism, amino acid metabolism, energy metabolism, lipid metabolism etc. were dominated in resistance and susceptible samples respectively. KEGG functional pathways categories distribution dipicted (Fig 2).

Differentially expressed genes analysis

The DESeq v1.8.1 with P-value threshold of 0.05 in DEGs of resistance Vs. conrol samples resulted 13,123 genes were commonly expressed, out of which 431 up-regulated and 400 down-regulated in resistance and susceptible samples respectively (Supplementary Table S3). The analyzed top 100 DEGs transcript abundance between resistance Vs. susceptible samples presented as Heat map (cluster) (Fig 3) and also, the DEGs were graphicaly represented as scatter plot (Fig 4) and volcano plot (Fig 5).

In our study transcripts having stress responsive characteristics with fold change values (up and down regulated) were found in DEGs. Downstream signaling components such as Kinases related transcripts i.e., MAP kinase I (4.0), ATP-NAD kinase-like domain (3.1), casein kinase (4.0), Thiamin pyrophosphokinase1 (TPK1) (3.0), cyclin-dependent kinase inhibitor 5 (CDKs) (3.3), somatic embryogenesis receptor kinase 1 (SERK) (3.6), wall-associated receptor kinase-14 (WAKs) (3.5), CBL-interacting serine threonine- kinase 1 (CIPK1) (2.9), uridine kinase isoform X1 (3.1), serine threonine- kinase endoribonuclease IRE1a-like (3.8), ceramide kinase (3.7) and cysteine-rich receptor kinase 2 (CRK2) (-3.2), phosphatidylinositol 4-kinase gamma 5 (PI4Kγ5) (-3.7), receptor kinase 1 (-4.3), ACT
tyrosine kinase (-2.9), cysteine-rich receptor kinase 2 (CRK2) (-5.3), casein kinase I (-4.2) and serine threonine- kinase HT1 (-3.2).

Phosphatase transcripts i.e., Dullard phosphatase (4.0), tyrosine- phosphatase (3.6), type I inositol 1,4,5-trisphosphate 5-phosphatase 2 (2.7), serine threonine phosphatase 2A (3.2), dCTP pyrophosphatase 1 (3.3), and Pyridoxal phosphate phosphatase-related (-3.7), phosphatase 2C (-3.9), phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity phosphatase PTEN (-2.9), serine threonine phosphatase 2A (-2.9), Phosphotyrosyl phosphatase (PTPA) (-2.7), Pyridoxal phosphate phosphatase phosphatase-related (-3.6). Transcription factors (TF) transcripts i.e., WRKY1 (4.1), GATA (3.16), NAC (3.0), ERF (2.7) up-regulated, bZIP (-5.7), DREB (-2.90), ERF (-3.0), myb (-3.30), NAC (-2.9) and WRKY (-4.5) were found. In reactive oxygen scavengers, glutathione transferase (GST) 23 (6.1) found. Pathogenesis-related 2 (-5.4) gene, and some of the R genes were also identified such as Leucine Rich Repeat domain (LRR) (4.2), Leucine-rich repeat receptor-like serine threonine- kinase (LRR-RLK) (3.8) and Leucine-rich repeat (LRR) (-2.8).

Antifungal peptides transcripts such as nascent polypeptide-associated complex subunit alpha 2 (-3.5), Tetratricopeptide repeat (TPR)-like superfamily isoform 1 (-2.9), peptide deformylase chloroplastic (-3.5), neurofilament heavy polypeptide (-2.7), pentatricopeptide repeat-containing mitochondrial (-3.8) were found.

**Validation of selected gene by RT-PCR**

The DEGs, HSP20 gene expression resulted up-regulation by 10 fold in resistance over susceptible sample using RT-PCR analysis.

**Discussion**

Coriander is a commercially vital dual purpose crop, yet stem gall affected on its lower profitability. Understanding the resistance potential of coriander crop to stem gall disease is of indispensible in the context of crop loss. For our study, ACr1 has been considered as a
resistance source and CS-6 for susceptible to stem gall due to its high level of field level response (Malhotra et al. 2016b). These samples utilized and pooled RNA and analyzed using transcriptome sequencing. To our knowledge, this is the first transcriptome resource on stem gall resistance genes in coriander to be made available to the scientific community.

Transcriptome profiling is widely applied in the identification of the genes associated traits in eukaryotes (Wang et al. 2009). The NGS is widely adopted for the qualitative and quantitative analysis of transcriptomes. More number of transcripts can be obtain with high coverage by De novo transcriptome assembly (Zhao et al. 2011). Our study produced high-quality raw reads of 4,91,63,108 and 4,37,46,120, after assembly total 59,933 and 56,861 transcripts obtained in resistance and susceptible samples respectively. Galata et al., 2014 reported total 3,33,30,312 raw reads from three samples i.e., small, medium and large mericarps of coriander and De novo assembly resulted 65,306 transcripts.

The functional annotation among resistance and susceptible samples, gene ontology provided majority of the transcripts involved in molecular function followed by biological process and cellular component. In the GO terms, catalytic activities (Dangl and McDowell 2006), binding activities, transcription (Yang et al., 1997), molecular Transducer (Heidrich et al., 2011) etc. are play an important role at molecular level in plant resistance to disease. The cellular component terms associated with cell followed by cell part, organelle, organelle part, macromolecular complex play an important role in disease tolerance mechanisms in plants (Huckelhoven 2007; Malinovsky et al. 2014).

The pathway analysis revealed the involvement of majority of the transcripts in metabolism process in resistance compared with susceptible sample respectively. The pathways enrichment analysis using stress responsive-genes in Arabidopsis thaliana showed majority of the genes involved in metabolic process under environmental stress conditions (Naika et al. 2013a). In our study, some of the transcripts involved in metabolism pathways expressed
differentially and belonged major to Phenylpropanoid biosynthesis pathway (Dixon and Paiva 1995) a stress induced metabolic pathway and the enzymes associated with this pathway included trans-cinnamate 4-monooxygenase (Bell-Lelong et al. 1997) known to involve in development and environmental stress tolerance. The coniferyl-aldehyde dehydrogenase, cinnamyl-alcohol dehydrogenase and peroxidise involved in the lignin biosynthesis pathways and also involved in disease resistance in plants (Das et al. 2017) were highly expressed in our resistant genotype.

The comparison between resistance Vs. susceptible samples, total 13123 genes were commonly expressed out of which 431 and 400 genes were up and down regulated. Our study found some of the transcripts involved in signalling components like and known to involve in defense response against pathogen i.e. MAP kinase I (Yang et al 2001), E1A/CREB-binding protein, phosphatidylinositol phospholipase C (PLCD) (Vossen et al. 2010), palmitoyltransferase ZDHHC9/14/18 (ZDHHC9_14_18), interleukin-1 receptor-associated kinase 4 (IRAK4), ethylene-insensitive protein 2 (EIN2) (Thomma et al. 1999), casein kinase II subunit alpha (CSNK2A), diacylglycerol kinase (ATP) (dgkA), molecular chaperone HtpG (HSP90A, htpG) (Takahashi et al. 2003), speckle-type POZ protein (SPOP), 5'-AMP-activated protein kinase, catalytic alpha subunit (PRKAA, AMPK), protein transport protein SEC13 (SEC13), serine/threonine-protein phosphatase 2A regulatory subunit B’ (PPP2R5), RuvB-like protein 1 (pontin 52) (RUVBL1, RVB1, INO80H), ubiquinol-cytochrome c reductase iron-sulfur subunit (UQCRFS1, RIP1, petA), two-component response regulator ARR-A family (ARR-A) (To et al., 2007), mitogen-activated protein kinase 3 (MPK3) (Hamel et al. 2012), and 5'-AMP-activated protein kinase (Rutter et al. 2003), catalytic alpha subunit (PRKAA, AMPK) and scavenger enzyme glutathione S-transferase 1 (Marrs 1996). Transcripts involved in down-regulation are phospholipase D1/2
(PLD1_2), auxin-responsive protein (IAA), and serine/threonine-protein phosphatase 2A regulatory subunit B (PPP2R2).

Various transcription factors families such as AREB/ABF, AP2/ERF, bZIP, DREB1/CBF, HB, MYB, HSF, MYC, NAC and WRKY were shown to influence biotic and abiotic stress response in plants (Naika et al. 2013 b). Our study also found transcripts associated with WRKY, NAC, bZIP, HSF, DREB, ERF, and MYB which could be further utilized for breeding of disease resistance in coriander. Some of the TFs such as WRKY33 (Zheng et al. 2006), ethylene-responsive transcription factor 1 (ERF1) (McGrath et al. 2005) and ERF6 (Gutterson and Reuber 2004) which is very well characterized in response to pathogen and similar transcripts were found in our study.

The RT-PCR output showed up-regulation of HSP20 gene expression by 10 fold increased in resistance over susceptible sample. The HSP 20 gene family showed its role in abiotic and biotic stress tolerance in plants (Lopes-Caitar et al. 2013) and this validated gene may be play an important role in coriander stem gall resistance.

The R genes belonging to the NBS-LRR family known to detect pathogen effectors, and lead to pathogen resistance (DeYoung and Innes, 2006). In our study, two transcripts associated with R genes were specifically up-regulated and one is down regulated in resistance over susceptible samples and these could be a candidate genes for disease resistance in coriander. It is very well known that plants produce many antifungal peptides (De Lucca 2000). Our study identified five antifungal peptides, and these could be further validated for disease control in coriander. The resource generated from this study could be useful for the biotechnologist, molecular breeder, pathologist, plant biochemist and computational biologist.
Acknowledgments

We thank Dr. Gopal Lal, Director, NRCSS for timely support and encouragement during the research work. The research is supported by the ICAR-NRC on Seed Spices, Ajmer, Rajasthan, India. SC, RS, RDM, GL thank ICAR-NRC on Seed Spices, Ajmer for providing all the facilities. MBNN thank UHS, Bagalkot for infrastructure facilities.

Author Contributions

Conceived theme of the study: SC

Performed the experiments and collected samples: RS, RDM and RS

Analysed the data: MBNN and SC

Drafted the manuscript: MBNN and SC

Edited the manuscript: GL, RS, RDM and RS

All authors read and approved the final manuscript

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References


Anders S. and Huber W. 2010 Differential expression analysis for sequence count data. Genome Biol. 11(10), R106.


Diederichsen A. 1996 Coriander, Coriandrum Sativum L. International Plant Genetic Resources Institute ed, Rome, Italy.


Galata M., Sarker L. S. and Mahmoud S. S. 2014 Transcriptome profiling, and cloning and characterization of the main monoterpene synthases of Coriandrum sativum L. Phytochemistry. 102, 64-73.


Takahashi A., Casais C., Ichimura K. and Shirasu K. 2003 HSP90 interacts with RAR1 and SGT1 and is essential for RPS2-mediated disease resistance in Arabidopsis. Proc Natl Acad Sci U S A. 100 (20), 11777-82.


Received 20 July 2018; revised 3 October 2018; accepted 18 October 2018
Tables

Table 1: Summary of the sequencing data and *de novo* sequence assembly of *Coriandrum sativum* transcriptome in response to *Protomyces macrosorpus* Unger infection.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Cultivars</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Resistance (ACr1)</td>
</tr>
<tr>
<td>Number of reads</td>
<td>49,163,108</td>
</tr>
<tr>
<td>Number of bases</td>
<td>7,404,005,575</td>
</tr>
<tr>
<td>Number of transcripts</td>
<td>59,933</td>
</tr>
<tr>
<td>Total transcript length (bases)</td>
<td>82,309,113</td>
</tr>
<tr>
<td>N50</td>
<td>1894</td>
</tr>
<tr>
<td>Maximum transcript length</td>
<td>12930</td>
</tr>
<tr>
<td>Minimum transcript length</td>
<td>200</td>
</tr>
<tr>
<td>Mean transcript length</td>
<td>1373</td>
</tr>
</tbody>
</table>

**CDS statistics**

<p>| Number of CDS length (bases)      | 52506               | 48858               |
| Total CDS length (bases)          | 54256515            | 45786651            |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum CDS length</td>
<td>9114</td>
<td>6597</td>
</tr>
<tr>
<td>Minimum CDS length</td>
<td>297</td>
<td>297</td>
</tr>
<tr>
<td>Mean CDS length</td>
<td>1033</td>
<td>937</td>
</tr>
<tr>
<td><strong>Range of CDS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200&lt; CDS &lt;500</td>
<td>13533</td>
<td>13892</td>
</tr>
<tr>
<td>500&lt; CDS &lt;1000</td>
<td>18429</td>
<td>17914</td>
</tr>
<tr>
<td>1000&lt; CDS&lt;2000</td>
<td>15538</td>
<td>13831</td>
</tr>
<tr>
<td>2000&lt; CDS &lt;3000</td>
<td>3617</td>
<td>2610</td>
</tr>
<tr>
<td>3000&lt; CDS &lt;4000</td>
<td>1005</td>
<td>485</td>
</tr>
<tr>
<td>4000&lt; CDS &lt;5000</td>
<td>225</td>
<td>100</td>
</tr>
<tr>
<td>CDS &gt;500</td>
<td>159</td>
<td>26</td>
</tr>
</tbody>
</table>

**Supporting information**

S1. Based on sequence homology, transcripts in resistance and susceptible were assigned to GO term annotations.

S2_A_B. CDS were mapped to reference canonical pathways in KEGG and were assigned to KEGG pathways.

S3. Differential expression of genes in resistance and susceptible samples.
Figure legends

1. A and B. Histogram of GO (gene ontology) classifications of assembled transcripts of coriander. The results are summarized in the following three main categories: (1) biological process, (2) cellular component and (3) molecular function. The left y-axis indicates the percentage of the specific category of genes in the main category. The right y-axis indicates the number of genes in a specific category.
2. KEGG pathway distribution of resistance and susceptible samples.
3. Heatmap of differentially expressed genes in control and treated sample. Up-regulation and down-regulation is represented by red shading and green shading, respectively (NRCSS1: Susceptible; NRCSS2: Resistance).
4. Scatter plot of differentially expressed genes in susceptible and resistance sample.

The scatter plot view is useful for examining the expression level of genes in two distinct conditions. It helps to identify genes that are differentially expressed in one sample versus another and also allows comparing two values associated genes. In given scatter plot each dot represents a gene. The vertical position of each gene represents its expression level in the treatment condition and horizontal position represents its control strength. Thus, genes that fall above the diagonal are over-expressed and genes that fall below the diagonal are down-regulated as compared to their median expression level in experimental grouping of the experiment.
5. Volcano plot of differentially expressed genes in susceptible and resistance sample.

The volcano plot arranges expressed genes along dimensions of biological as well as statistical significance. The x-axis represents the log fold change treated sample to control sample on which red block on the right side zero represents the up-regulated genes whereas green block on the left side of zero represents significant down regulated genes. While Y-axis represents the negative log of p-value (p value <=0.05) of the performed statistical test where data points with low p-values (highly significant) appearing towards the top of the plot. Grey block shows the non-differentially expressed genes.