J Plant Biotechnol (2016) 43:189-203
DOI:http://dx.doi.org/10.5010/JPB.2016.43.2.189

ISSN 1229-2818 (Print)
ISSN 2384-1397 (Online)

Research Article

In silico analysis of MeJA-induced comparative transcriptomes in *Brassica oleraceae* L. *var. capitata*

Ok Ran Lee · Dae-Soo Kim

Received: 23 May 2016 / Revised: 27 May 2016 / Accepted: 7 June 2016
ⓒ Korean Society for Plant Biotechnology

Abstract *Brassica oleraceae var capitata* is a member of the Brassicaceae family and is widely used as an horticultural crop. In the present study, transcriptome analysis of *B. oleraceae L. var capitata* was done for the first time using eight-week old seedlings treated with 50 μM MeJA, versus mock-treated samples. The complete transcripts for both samples were obtained using the GS-FLX sequencer. Overall, we obtained 275,570 and 266,457 reads from seedlings treated with or without 50 μM MeJA, respectively. All the obtained reads were annotated using biological databases and functionally classified using gene ontology (GO), the Kyoto Encyclopedia of Genes and Genomics (KEGG). By using GO analyses, putative transcripts were examined in terms of biotic and abiotic stresses, cellular component organization, biogenesis, and secondary metabolic processes. The KEGG pathways for most of the transcripts were involved in carbohydrate metabolism, energy metabolism, and secondary metabolite synthesis. In order to double the sequenced data, we randomly chose two putative genes involved in terpene biosynthetic pathways and studied their transcript patterns under MeJA treatment. This study will provide us a platform to further characterize the genes in *B. oleracea var capitata*.

Keywords *Brassica oleracea*, EST, Gene ontology, KEGG pathway, Methyl jasmonate

O. R. Lee (✉)
Department of Plant Biotechnology, College of Agriculture and Life Science, Chonnam National University, Gwangju 61186, Korea
e-mail: mpizlee@jnu.ac.kr

D.-S Kim
Rare Disease Research Center and Korea Research Institute of Bioscience and Biotechnology (KRIBB), 125 Gwahangno, Yuseong-gu, Daejeon 34141, Korea

Introduction

B. oleraceae L. var. capitata (cabbbage) is one of the most important vegetable crops belong to the species *B. oleraceae* L., and grouped under the family Brassicaceae. It is an herbaceous, biennial and dicotyledonous flowering plant with leaves forming a characteristic compact head. Cabbage has a positive impact on human health and is, in addition to being a source of vitamins and fiber, connected with secondary metabolites called glucosinolates, which are known to possess anti-carcinogenic properties (Sarikami et al. 2009). *Brassicaceae* vegetables are a good source of antioxidants because of their high phenolics and glucosinolate content. Throughout the growth and developmental stages, plants are ordinarily exposed to various environmental biotic and abiotic factors. Biotic and abiotic stresses are major concern for the sustainable production of these crops. Each type of biotic and abiotic stresses function through different types of molecular mechanisms that affect plants, and eventually cause damage (Dixon and Lamb 1990). In particular, some biotic and abiotic elicitors can activate specific secondary metabolite production. Under these conditions, a number of signaling pathways can be pre-activated by salicylic acid (SA), jasmonic acid (JA), ethylene or abscisic acid pathways which are generally, involved in the defense responses. Efforts to develop plants resistant to biotic and abiotic stresses is an important task in biotechnology and functional genomics study of which is one of the most important tool for identifying potential genes related to stress resistance.

Genome sequence of *B. olearaceae* was not available till now. An alternative and efficient method for analyzing transcriptome can be done using ESTs. ESTs are short, unedited, randomly selected single pass sequence read derived from cDNA libraries, also called as 'poor' man's genome and proven to be a valuable tool in molecular biology (Nagaraj et al. 2007). Recently, EST sequencing is a widely used application tool to study gene expression pattern in response to a given

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

environmental stimulus and a trancriptomics study of a plant at various stages of development under different experimental conditions with various plant tissues (Kim et al. 2003). EST technology was primarily introduced in the human genome project, and used widely to clone new genes, determine tissue-specific gene expression profiles, annotate functional genome sequences and so on (Ohlrogge and Benning 2000; Brandle et al. 2003). Raw sequences obtained from high-through put screening are quite impossible to annotate manually. In these cases, computational methods are used to process EST sequences, such as sequence cleaning, vector masking, clustering, assembly and annotation to yield biological information to putative sequences. The use of a bioinformatics tool for homology-based functional annotation and statistical information is simple for ESTs from a range of organisms in the public non redundant databases (Falgueras et al. 2006). Gene ontology (GO), are liable biological annotation schema that is globally accepted by biological communities, was developed with three structured vocabularies (i.e., ontologies) to describe genes and proteins in terms of cellular components, biological processes, and molecular function (Pal 2006).

In this study, *in silico* approach was used to classify *B. oleraceae L.var. capitata* ESTs based on GO vocabularies, quantification of transcript abundance, classification of biochemical pathways based on KEGG pathway, and alternatively spliced transcripts. Putative functional protein domains were also analyzed using biological databases. Using similarity search, unique sequences was assigned as putative sequences and further analysis of those sequences aid to understand a complex gene network expression.

Materials and methods

Plant materials and MeJA treatments

B. oleraceae seeds were germinated in soil under *in vitro* conditions at 25°C with continuous light and dark conditions (16h/8h). Each pot had five cabbage seedlings. After eight weeks, seedlings were treated with 50 μM MeJA and without (mock). For treatment, sterilized filter paper (Whatman No. 6, ø185 mm) was dipped in 5 mL of 40 % ethanol (control) or 50 μM MeJA placed in pots on the top of cabbage seedlings without touching its leaf surfaces. After treatment, pots were covered by polyethylene bags and wrapped well as shown in Figure 1. All the plants were grown at 25°C, 16h light/8h dark condition for 3 days. Cabbage leaves harvested and frozen immediately with liquid nitrogen and stored at -70°C which was used for sequencing. For expression analysis, eight week old seedlings were treated with and without 50 μM MeJA for up to 9 days. Seedlings were collected every alternative day (3rd, 5th, 7th, and 9th day) and immediately frozen with liquid nitrogen and stored at -70°C until required.

RNA isolation and cDNA library construction

The total RNA was isolated via the aqueous phenol extraction procedures using TRIzol® Reagent (Thermo Fisher Scientific, Catalog No. 15596-026). Messenger RNA was isolated from 50 μg of the total RNA sample using the FastTrack™ MAG Micro mRNA Isolation Kit (Life Technologies, cat# K1580-01). The kit was used following the protocol as provided by the manufacturer. The isolated mRNA was quantified using Agilent RNA 6000 NANO CHIP (Agilent, cat.# 5067-1511) and 200 ng of this material was used in the cDNA Rapid Library Preparation protocol. The first step of cDNA Rapid library preparation is fragmentation of mRNA using $ZnCl_2$ and heat treatment. The cleaved RNA fragments primed with random hexamers were reverse transcribed into first strand cDNA, and double-stranded (ds) cDNA was generated after removing the RNA template and synthesizing a replacement strands. The double-stranded cDNA ends were polished (blunted), and short adapters were ligated onto both ends. The adapters provide priming sequences for both amplification and sequencing of the sample library fragments, as well as the “sequencing key”, a short sequence of 4 nucleotides used by the system software for base calling and, following purification and size selection using AMpure beads. Finally the quality of the library of cDNA fragments was assessed using 2100 BioAnalyzer (Agilent), and the library was quantitated to determine the optimal amount of the library to use as input for emulsion-based clonal amplification.

Emulsion PCR and sequencing run

Single “effective” copies of template species from the DNA library to be sequenced were hybridized to DNA Capture Beads. The immobilized library was then resuspended in the amplification solution, and the mixture is emulsified, followed by PCR amplification. After amplification, the DNA-carrying beads were recovered from the emulsion and enriched. The second strands of the amplification products were melted away as part of the enrichment process, leaving the amplified single-stranded DNA library bound to the beads. The sequencing primer is then annealed to the immobilized amplified DNA templates. After amplificatioin, the DNA-carrying beads were set into the wells of a PicoTiterPlate device (PTP) such that wells contain single DNA beads. The loaded PTP was

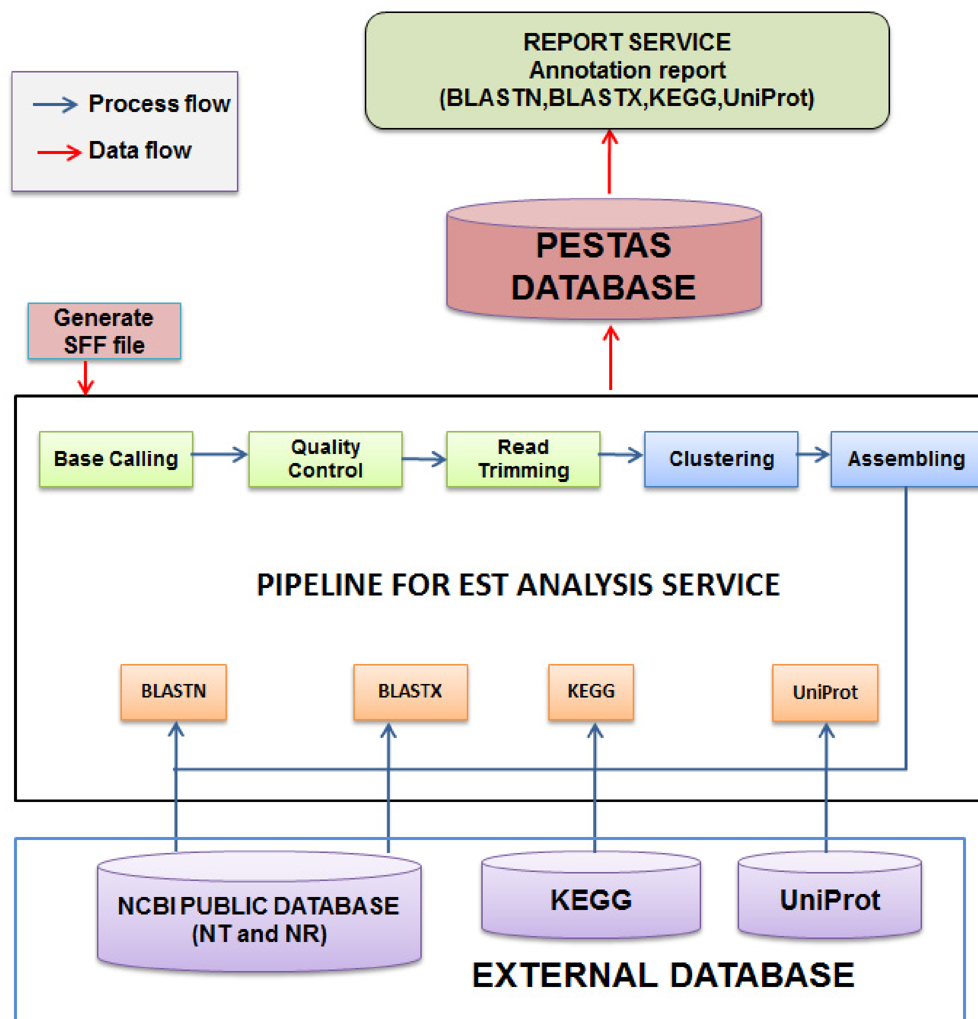

Fig. 1 Pipelines used for EST analysis of *Brassica oleraceae*

then inserted into the Genome Sequencer FLX Instrument, and sequencing reagents were sequentially flowed over the plate. Information from all the wells of the PTP is captured simultaneously by the camera, and can be processed in real time by the onboard computer. Samples were sequenced by Macrogen Inc, Korea (http://www.macrogen.com)

RNA Isolation and semi quantitative PCR analysis

RNA was extracted from *B. oleraceae* that were subjected to the experimental treatment using the RNeasy kit (Qiagen, USA) according to the manufacturer's instructions. The quality and concentration of RNA was measured using a spectrophotometer (GE nanovalue, USA). To obtain the first strand cDNA, 1.5 µg of total RNA was used, and cDNA was synthesized using a Power cDNA kit (Invitrogen, USA) following the manufacturer's instructions. We performed RT-PCR using optimum PCR conditions: 92°C for 2 min, 92°C for 40 sec, 54.5 ~ 60°C for 30 sec, 72°C for 1 min, 72°C for 10 min for 28 cycles, gene specific primers along with optimum annealing temperatures were listed in Table S1. The house-keeping gene encoding *actin* was used as a standard for all samples. For the analysis of all transcripts, untreated samples were used as negative control.

Unigene dataset generation

The unigene data set contains a set of non-redundant sequences composed of singlets and contigs. Sequence files were produced with the SFF (Standard Flowgram Format) file, and then SFF files were processed using the GS assembler FLX software tool kit (v.2.6) provided by Roche. During the assembly process, the software identified pairwise overlaps between reads, constructed multiple alignments of overlapping reads and divided or introduced breaks into the multiple alignments in regions where consistent differences are found between different sets of reads. The software attempted to resolve branching structures between contigs, and generated consensus base calls of the

contigs by using quality and flow signal information for each nucleotide flow included in the contigs multiple alignments. The contig consensus sequences and corresponding quality scores were generated along with an ACE file of the multiple alignments and assembly metrics files. When paired end data is available, the assembler performed extra steps; organized the contigs into scaffolds using Paired End information to order and orient the contigs and to approximate the distance between contigs, and output scaffolded consensus sequences and corresponding quality scores, along with an AGP file of the scaffolds and specific metrics tables. A *de novo* assembler project was created for the short cDNA sequence reads with default parameters. Using SeqClean and Lucy tools, we obtained singleton and contig sequences. Low quality and low-complexity sequences were removed using SeqClean (http://sourceforge.net/projects/seqclean/). The Lucy as a sequence cleanup program used for the processes of quality assessment, confidence reassurance, vector trimming and vector removal. The remaining sequences were used in the functional analysis (Fig. 1).

Gene ontology and KEGG pathway assignment

Gene ontology (GO) term annotation and function-based analysis of unique sequences were performed using Blast2go (B2G), a sequence-based high-through put sequence analysis tool. Using B2G, unigenes were subjected to a BLASTX query against the national center for biotechnology information (NCBI) public non-redundant (NR) database. Based on the BLASTX results, the sequences were putatively named using the BLAST description annotator (BDA) tool embedded in B2G. More collective logic models were embedded in B2G to retrieve GOs, EC numbers, and KEGG maps. GO terms for each of the three main categories (biological process, molecular function and cellular component) were obtained from sequence similarity and BLAST scores (E\=10-3) with default parameters. Those annotations were simplified into plant functional categories using the plant GO slim. Interproscan embedded with B2G was used to obtain the protein domain information for the putative sequences (Conesa and Gotz 2008).

Comparative analysis of alternative splicing in MeJA treated samples and control samples

The Arabidopsis genome sequence, annotation and annotated sequence features were downloaded from TAIR (TAIR 10 database release). All transcripts sequences were mapped against the Arabidopsis genome using the C/S version of the BLAT program. We used the BLAT version 35 downloaded from Dr. W.J. Kent's homepage (http://users.soe.ucsc.edu/~kent/src/). BLAT output contains many suboptimal alignments. Only the best alignment with the highest BLAT score was kept unless we had multiple hits of the same quality. We first grouped transcripts sequences and reference genes into clusters on the genome if they mapped onto the same genomic region, were orientated on the same strand, and had overlapping sequences. Alternative splicing events and differentially expressed gene features were identified using database queries. Only initial and terminal nucleotides were allowed to remain unmapped, >90% of the transcripts sequence had to be involved in perfect matches with the genome, and every exon had to be ≥85% identical (or contain at most five errors) to the genome. These transcripts were clustered as described above. Single-exon and unspliced transcripts were discarded as they do not reveal any information about alternative splicing.

Results and Discussion

Sequence and EST assembly

In order to find optimum concentration of MeJA, growth test was done on 10-, 50-, 100 μM of MeJA containing medium using 9-day grown seedling for 3, 5, 7, 9 days respectively (Fig. 2A). Shoot dry weight (Fig. 2B) was measured, and found 50 μM is in biological range and 100 μM is slightly over that range. For the construction of cDNA library, eight-week old *B. oleraceae* seedlings were treated with 50 μM MeJA and without (control) (Fig. 2C), and previously reported MeJA-inducible glucosinolate biosynthesis-related gene's expression (Mikkelsen et al. 2003; Pozo et al. 2008) was confirmed that the concentration was still working in different aged plant (Fig. 2D). The quality of cDNA library was assessed using 2100 Bio analyser (Agilent). The cDNA library was constructed using GS-FLX sequencer (Roche v2.6) resulted in the total of 275,570 and 266,457 reads respectively. All the obtained reads were subjected to seq clean (http://sourceforge.net/projects/seqclean/) and lucy (http://lucy.sourceforge.net/) with default parameters for removing low quality and vector sequences, and assembled into 458,961 ESTs. Control and MeJA treated ESTs had an average length of 442 bp and 448 bp respectively. The maximum number of ESTs was higher in control compared to the MeJA treated samples. All the assembled sequences were functionally annotated using blast2go (Table 1).

EST similarity search against public non-redundant databases

Homology-based functional annotation of putative sequences was obtained through BLAST X queries against public non-

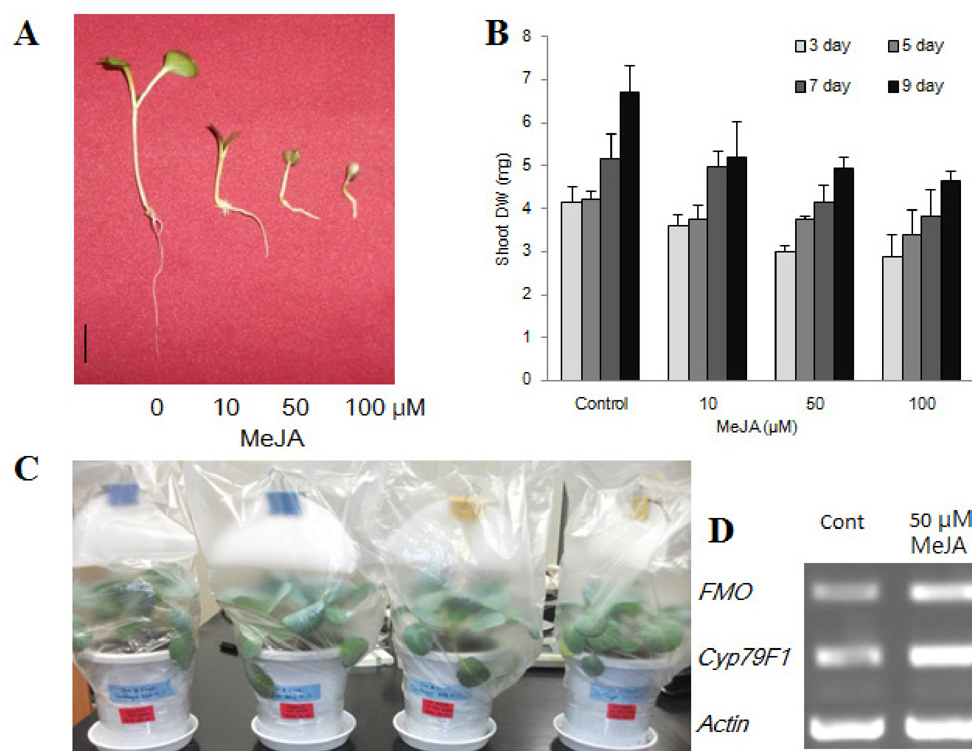

Fig. 2 Physiological status of *B. oleracea L. var capitata* following MeJA treatment (A) *B. oleracea L. var capitata* grown for nine days were mock treated or treated with 10, 50, or 100 μM of MeJA. (B) Shoot dry weight (DW) gradually decreased as MeJA concentration and number of treatment days increased. (C) Eight-week old *B. oleracea L. var capitata* were treated with 50 μM MeJA for 3 days. Relative expression patterns of MeJA-inducible flavin-containing monooxygenase (FMO) and N-hydroxylase for short chain methionine derivative (Cyp79F1) show that the MeJA treatment was done properly

Table 1 EST assembly and annotation reports

Description	No of contigs (Control)	No of contigs (MeJA)
Number of contigs	11,248	6,410
Number of Contigs with GO (redundancy)	96,371	64,952
Number of Contigs with GO (non redundancy)	9,272	5,414
Number of Contigs without GO	1,976	996
Number of sequence with EC	4,879	3,851
Number of Contigs enzyme codes	1,098	997

redundant databases with an average E-value of 10^{-3} or below, an HSP cut off of 33 and a maximum of 20 blast hits per sequence (Gotz et al. 2008; Mao et al. 2005). The BLAST X search for the control showed 9563 (85.2%) meaningful matches and 1685 (14.9%) didn't show any significant hits. Similarly, MeJA treated samples showed 5636 (84.3%) matches with known sequences and 1049 (15.6%) didn't show any similarity matches. The total blast hits for the control samples of *B. oleraceae var capitata* showed 74.89% from the following plants: *Arabidopsis lyrata* (27.7%), *Arabidopsis thaliana* (20.25%), *Brassica rapa* (13.45%) *Thellungiella halophila* (5.1%), *Zea mays* (3.25%), *Brassica napus* (2.8%), *Brassica oleracea* (1.5%), *Brassica juncea* (0.9%) and other plants (25.11%). However, the MeJA treated samples showed 92.3% matching sequences through BLAST X program. This program showed 27.1 %, 19.8 %, 13.02%, 8.78%, 3.67%, 5.6%, 2.9%, and 1.6% similarities with the following plants: *Arabidopsis lyrata, Arabidopsis thaliana, Brassica rapa, Thellungiella halophila, Zea mays, Brassica napus, Brassica oleracea,* and *Brassica juncea*. The results of both control and MeJA treated samples are displayed separately in Table 2. The above-mentioned plants have been relatively well-studied, so there are more experimental data rather than in *B. oleraceae L. var capitata*. Annotations were obtained using BLAST description

Table 2 Classification based on species from BLAST hit sequences for control and MeJA-treated samples

Species	Control	MeJA
Arabidopsis lyrata	2,656	1,528
Arabidopsis thaliana	1,937	1,114
Brassica rapa	1,287	734
Thellungiella halophila	495	326
Zea mays	311	207
Brassica napus	268	318
Brassica oleracea	147	165
Brassica juncea	81	91

annotator (BDA) tool, and all the putative sequences were used for further experimental analysis.

Functional analysis based on gene ontology

Functional gene annotation is a difficult task for newly sequenced non-model plants than human since the plant genome contains numerous genes reflecting adaptations to environmental factors. To simplify this process, gene ontology (GO) has evolved in the field of functional genomics. GO describes the function of gene based on control vocabularies and for elucidating hierarchical relationships between gene groups. Control vocabularies are grouped into three major categories, namely molecular function, biological process, and cellular component (The Gene Ontology 2010). The data obtained through this study were organized into one, a combination of two and belonging to all three categories. The data were then grouped in a Venn diagram (Fig. 3). The total EST sequences were 3,617 (2,322), 943 (498), 31,418 (21,236) for the control and MeJA (in parenthesis) respectively, and grouped into cellular component (CC), molecular function (MF), and biological process (BP). A total of 381 ESTs were annotated into CC and MF groups, 2294 ESTs were annotated into MF and BP, and 5143 ESTs were annotated into BP and CC groups, and a total of 8420 ESTs were mapped into all three GO categories in case of MeJA-treated samples. According to the plant-GOslim, plant-specific GO vocabularies were screened. In our result, a large number of unique sequences were grouped under the first category of biological processes with subcategories such as the responses to cadmium ion, salt stress, oxidation-reduction process, cold, regulation of plant-type hypersensitive response, serine family amino acid metabolic process, protein targeting to membrane, chitin, DNA-dependent regulation of transcription, glycolysis, bacterium, gluconeogenesis, protein phosphorylation and wounding (Table S2). It is noteworthy that the number of wound responsive and Golgi organization-related contigs are more highly ranked by MeJA treatment (Table S2). The second category included cellular component which includes unique sequences associated with plasma membrane, nucleus, cytosol, chloroplast, plasmodesmata, Golgi apparatus, mitochondria, integral to membrane, chloroplast stroma, chloroplast envelope, vacuolar membrane, cytoplasm, nucleolus and so on (Table S3). The third representative category was molecular function with ATP binding, protein binding, zinc ion binding, sequence-specific DNA binding transcription factor activity, DNA binding, nucleotide binding, protein serine/threonine kinase activity and so forth based on their hit numbers of contigs (Table S4). The GO results rely on previously annotated GO information of other plants.

Generally, various transcription factors and kinases play a crucial role in stress related signaling pathways (Horan et al. 2008). For example MeJA is the important signaling cascade linked to various biotic and abiotic stress mechanisms that modulates various physiological processes in plants, including root senescence, and the defense responses against insect and pathogen attack (Wasternack and Parthier 1997; War et al., 2012). Also MeJA induces or increases the biosynthesis of plant secondary metabolites that play an important role in various environmental conditions (Yukimine et al. 1996; Ramakrishna and Ravishankar, 2011). For instance, MeJA increases the production of soyasaponin in glycyrrhiza glabra cells (Hayashi et al. 2003), saikosaponin in the adventitious roots of *Bupleurum falcatum* (Aoyagi et al., 2001), and ginsenoside production in *P. ginseng* (Shim et al., 2010). For example a putative gene transcript of squalene monooxygenase located in integral to membrane (cellular component), involved in biosynthetic process (biological process), and resulted in nucleotide binding activity (molecular function) (Table S1)

KEGG biochemical pathway analysis

As an alternative method of grouping unique sequences on the basis of biochemical functions, were assigned to metabolic pathways via Kyoto Encyclopedia of Genes and Genomes

Table 3 Secondary metabolite-related genes in *Brassica oleraceae L. var. capitata* based on KEGG biochemical analysis

Pathways	Enzymes	ECNo.	No. of EST	Pathways	Enzymes	ECNo.	No. of EST
Phenyl propanoid biosynthesis	Peroxidase	EC:1.11.1.7	21		Prenylcysteine oxidase	EC:2.3.1.92	3
	Cinnamoyl-CoA reductase	EC:1.2.1.44	13		Hydroxymethylglutaryl-CoA reductase	EC:1.2.1.68	2
	Cinnamyl-alcohol dehydrogenase	EC:1.1.1.195	6		Hydroxymethylglutaryl-CoA reductase (NADPH)	EC:2.1.1.104	1
	Trans-cinnamate 4-monooxygenase	EC:1.14.13.11	6		Flavonol synthase	EC:1.14.11.23	9
	4-Coumarate---CoA ligase	EC:6.2.1.12	6		Trans-cinnamate 4-monooxygenase	EC:1.14.13.11	6
	Beta-glucosidase	EC:3.2.1.21	5		Dihydrokaempferol 4-reductase	EC:1.1.1.219	5
	Sinapoylglucose---malate O-sinapoyltransferase	EC:2.3.1.92	3		Flavonoid 3'-monooxygenase	EC:1.14.13.21	3
	Coniferyl-aldehyde dehydrogenase	EC:1.2.1.68	2	Flavanoid Biosynthesis	Caffeoyl-CoA O-methyltransferase	EC:2.1.1.104	1
	Caffeoyl-CoA O-methyltransferase	EC:2.1.1.104	1		Shikimate O-hydroxycinnamoyltransferase	EC:2.3.1.133	1
	Caffeate O-methyltransferase	EC:2.1.1.68	1		Leucocyanidin oxygenase	EC:1.14.11.19	1
	Sinapate 1-glucosyltransferase	EC:2.4.1.120	1		Myrcene synthase	EC:4.2.3.15	7
	Quinate O-hydroxycinnamoyltransferase	EC:2.3.1.99	1	Monoterpene and	(+)-neomenthol dehydrogenase	EC:1.1.1.208	4
	Shikimate O-hydroxycinnamoyltransferase	EC:2.3.1.133	1	Diterpene	(-)-menthol dehydrogenase	EC:1.1.1.207	4
Terpene backbone	1-deoxt-D-xylulose-5-phosphate reductoisomerase	EC:1.11.1.7	2	Biosynthesis	(R)-limonene synthase	EC:4.2.3.20	2
Biosynthesis	Acetyl-CoA C-acceyltransferase	EC:1.2.1.44	13		(4S)-limonene synthase	EC:4.2.3.16	2
	(E)-4-hydroxy-3-methylbut-2-enly-diphosphate synthase	EC:1.1.1.195	6		taxane 13alpha-hydroxylase	EC:1.14.13.77	2
	1-deoxy-D-xylulose-5-phosphate synthase	EC:1.14.13.11	6	Glucosinolate Biosynthesis	branched-chain-amino-acid transaminase	EC:2.6.1.42	7
	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	EC:6.2.1.12	6		N-hydroxythioamide S-beta-glucosyltransferase	EC:2.4.1.195	1
	Dimethylallyltranstransferase	EC:3.2.1.21	5		9-cis-epoxycarotenoid dioxygenase	EC:1.13.11.51	8
					abscisic-aldehyde oxidase	EC:1.2.3.14	2
					zeaxanthin epoxidase	EC:1.14.13.90	1
					phytoene synthase	EC:2.5.1.32	1
					(+)-abscisic acid 8' hydroxylase	EC:1.14.13.93	1

(KEGG) (Kaneshisa M, 2000). Using enzyme commission (EC) numbers as the basis for assignment all the putative ESTs were subjected to KEGG database query with BLAST score to retrieve KEGG enzyme codes and pathway maps. In this study we identified and categorized KEGG pathways into carbohydrate metabolism, amino acid metabolism, nitrogen metabolism, lipid metabolism, and secondary metabolism (Table S5). Among annotated KEGG pathways, several secondary metabolite-related pathway genes were listed up (Table 3). Phenyl propanoid contributes to all aspects of plant responses including biotic and abiotic stresses (La camera et al. 2004). Phenolic compounds are the biologically active compounds used in traditional medicine as anti-viral, anti-cancer, anti-inflammatory agents (Korkina 2007). Terpenoids pathways are classified into groups based on the number of carbon atoms attached to isoprene units. The major groups of terpenes are monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), and triterpenes (C30). We also identified ESTs involved in other pathways such as steroid metabolism, glucosinolate pathways, and flavonoid pathways.

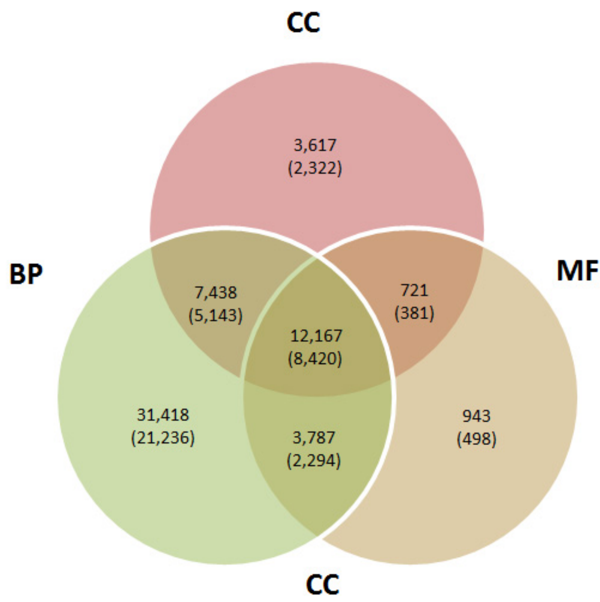

Fig. 3 Venn diagram of *Brassica oleraceae* data set showing numbers annotated to one, a combination of two, and/or all three GO vocabularies. Numbers in parentheses indicate the results from MeJA-treated samples compared to those of the control (MF, molecular function; BP, biological process; CC, cellular compound)

Secondary metabolite-related gene expressions and alternatively spliced genes by MeJA

MeJA is a small signaling molecule in the plant kingdom. Environmental stresses such as wounding and pathogen attack can induce MeJA production. Thus, screening MeJA-responsive genes can be a useful pool for the utilization of genetic engineering of crop plants. Several previously reported genes were also screened from MeJA-treated *B. oleraceae* (Fig. 4A). Several known genes including lipoxygenase, myrosinase-associated protein, and lipase, were read abundantly as reported in other plants (Ren et al. 2013). Other uncharacterized genes might be cabbage-specific suitable candidates. In plant, MeJA is well-known to induce monoterpenoid indole alkaloids and isoprenoid biosynthesis. Two selected isoprenoid biosynthesis-related genes, squalene monooxygenase and squalene synthase 1, were up-regulated by MeJA within 3 ~ 5 days respectively using semi-quantitative PCR. The genes selected in the present study are attributed to play an important role in direct or indirect defense mechanisms.

Alternative splicing has recently caught the attention of many plant researchers, as it can be spatially and temporally regulated, and is frequently associated with tissue types and environmental conditions to enhance transcriptome plasticity and proteome diversity. When all the obtained cabbage transcripts were mapped against the Arabidopsis genome using the BLAT program, total 33 transcripts which covered 80% showed different alternative splicing among the control and MeJA-treated samples (Fig. 5A). For example, the exact alternative splicing zone of Acyl-CoA: diacylglycerol acyltransferase (DGAT) and CBL-interacting protein kinase 1 are displayed

A

blastx - meja (top 10 reads#)

accession no.	description	contigs#	reads#
AAO03559	**lipoxygenase 2** [Brassica napus]	12	6860
ABO32545	**LOX** [Brassica oleracea var. gemmifera]	11	6083
BAJ33862	unnamed protein product [Thellungiella halophila]	7	5937
XP_003544036	PREDICTED: uncharacterized protein LOC100806341 [Glycine max]	2	5517
XP_003614387	RRNA intron-encoded homing endonuclease [Medicago truncatula]	3	4814
XP_003637074	Cell wall-associated hydrolase, partial [Medicago truncatula]	1	3533
CAA71237	**myrosinase-associated protein** [Brassica napus]	3	3450
XP_002888626	hypothetical protein ARALYDRAFT_894537 [Arabidopsis lyrata subsp. lyrata]	1	3402
XP_002488947	hypothetical protein SORBIDRAFT_1368s002010 [Sorghum bicolor] SORBIDRAFT_1368s002010 [Sorghum bicolor]	1	2837
AAQ01575	putative **lipase** [Brassica rapa subsp. pekinensis]	4	2802

B

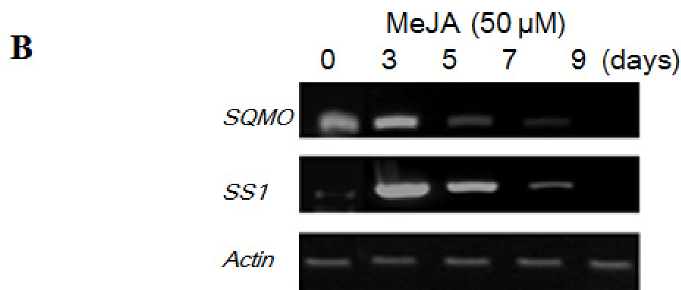

Fig. 4 MeJA-inducible genes from *Brassica oleraceae L. var. capitata.* (A) Top 10 reads from control and MeJA-treated EST. (B) Expression patterns of selected genes assessed using semi-quantitative PCR analysis. (*SQMO*, squalene monooxygenase; *SS1,* squalene synthase 1)

A

Arabidopsis	Gene Discription
AT1G09270.1	importin alpha isoform 4
AT1G09640.1	Translation elongation factor EF1B, gamma chain
AT1G61240.1	Protein of unknown function (DUF707)
AT1G71820.1	Encodes a member of the exocyst complex gene family (SEC6)
AT2G03120.1	signal peptide peptidase
AT2G19450.1	**Acyl-CoA: diacylglycerol acyltransferase (DGAT)**
AT2G30600.1	BTB/POZ domain-containing protein
AT2G35260.1	unknown protein
AT2G40840.1	disproportionating enzyme 2
AT2G43070.1	SIGNAL PEPTIDE PEPTIDASE-LIKE 3
AT3G12780.1	phosphoglycerate kinase 1
AT3G17510.1	**CBL-interacting protein kinase 1**
AT3G27240.1	Cytochrome C1 family
AT3G55620.1	Translation initiation factor IF6
AT3G62550.1	Adenine nucleotide alpha hydrolases-like superfamily protein
AT4G01050.1	thylakoid rhodanese-like
AT4G13590.1	Uncharacterized protein family (UPF0016)
AT4G14160.1	Sec23/Sec24 protein transport family protein
AT4G14210.1	phytoene desaturase 3
AT4G21660.1	proline-rich spliceosome-associated (PSP) family protein
AT4G24550.1	Clathrin adaptor complexes medium subunit family protein
AT4G25970.1	phosphatidylserine decarboxylase 3
AT4G33210.1	F-box family protein
AT4G33420.1	Peroxidase superfamily protein
AT4G34000.1	abscisic acid responsive elements-binding factor 3
AT4G38790.1	ER lumen protein retaining receptor family protein
AT5G06290.1	2-cysteine peroxiredoxin B
AT5G08080.1	syntaxin of plants 132
AT5G11700.1	BEST Arabidopsis thaliana protein match is: glycine-rich protein
AT5G27850.1	Ribosomal protein L18e/L15 superfamily protein
AT5G45130.1	RAB homolog 1
AT5G63890.1	histidinol dehydrogenase
AT5G66680.1	dolichyl-diphosphooligosaccharide-protein glycosyltransferase 48kDa subunit family protein

B

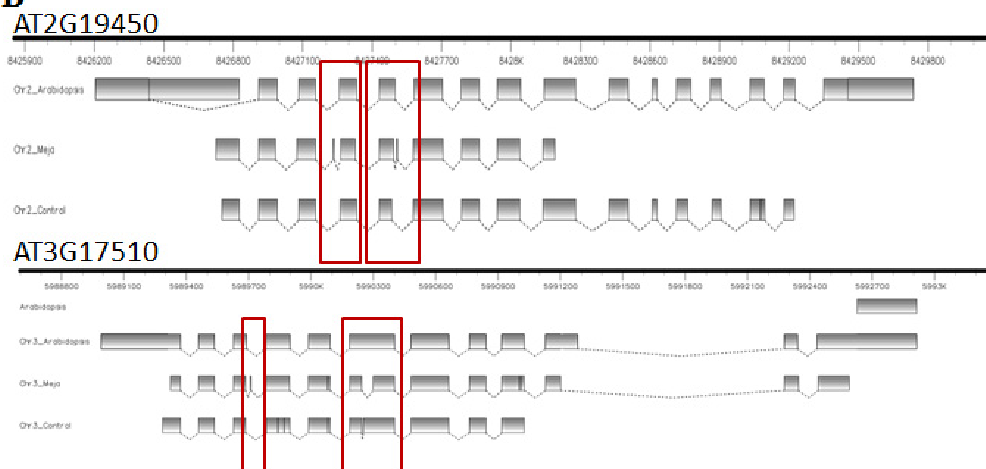

Fig. 5 Alternatively spliced genes as compared with the Arabidopsis genome. (A) List of a total of 33 alternatively spliced genes. (B) Two of the selected genes show descriptive alternative splicing between the control and MeJA-treated samples

(Fig. 5B). Taken together, all the transcripts and differentially transcribed genes in different conditions together with alternatively spliced genes will provide the basis for the functional study of *B. oleraceae L. var capitata.*

Acknowledgements

This study was financially supported by grants from the Next-Generation BioGreen 21 Program (SSAC, grant: PJ01112901), Rural Development Administration, Republic of Korea, and the Cabbage Genomics assisted breeding supporting Center (CGsC) research programs funded by Ministry for Food, Agriculture, Forestry and Fisheries of the Korean Government.

References

Brandle JE, Richman A, Swanson AK, Chapman BP (2000) Leaf ESTs from *Stevia rebaudiana*: a resource for gene discovery

in diterpene synthesis. Plant Mol Biol 50:613-622

Conesa A, Gotz S (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008:619832

Dixon RA, Lamb CJ (1990) Molecular communication in interactions between plants and microbial pathogens. Annu Rev Plant Physiol Plant Mol Biol 41:339-367

Falgueras J, Antonio JL, Noe FP, Francisco RC, Guillermo PT, Claros MG (2010) SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. BMC Bioinformatics 1471-2105

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36:3420-3435

Hayashi H, Huang P, Inoue K (2003) Up-regulation of soyasaponin biosynthesis by methyl jasmonate in cultured cells of *Glycyrrhiza glabra*. Plant Cell Physiol 44:401-411

Horan K, Jang C, Bailey-Serres J, Mittler R, Shelton C, Harper JF, Zhu JK, Cushman JC, Gollery M, Girke T (2008) Annotating genes of known and unknown function by large-scale coexpression analysis. Plant Physiol 147:41-57

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32(Database issue):D277-D280

Kim SI, Kim JY, Kim EA, Kwon KH, Kim KW, Cho K, Lee JH, Nam MH, Yang DC, Yoo JS, Park YM (2003) Proteome analysis of hairy root from *Panax ginseng* C.A. Meyer using peptide finger printing, internal sequencing and expressed sequence tag data. Proteomics 3:2379-2392

Korkina L.G (2007) Phenylpropanoids as naturally occurring antioxidants: from plant defense to human health. Cell Mol Biol 53:15-25

La Camera S, Gouzerh G, Dhondt S, Hoffmann L, Fritig B, Legrand M, Heitz T (2004) Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunol Rev. 198:267-284

Mao X, Cai T, Olyarchuk JG, Wei L (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 21:3787-3793

Mikkelsen MD, Petersen BL, Glawisching E, Jensen AB, Andreasson E, Halkier BA (2003) Modulation of CYP79 genes and glucosinolate profiles in Arabidopsis by defense signaling pathways. Plant Physiology 131:298-308

Nagaraj SH, Gasser RB, Ranganathan S (2007) Ahitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinform 8:6-21

Ohlrogge J, Benning C (2000) Unraveling plant metabolism by EST analysis. Curr Opin Plant Biol. 3:224-228

Pal D (2006) On gene ontology and function annotation. Bioinformation 1:97-98

Pozo MJ, Van Der Ent S, Van Loon C, Pieterse CMJ (2008) Transcription factor MYC2 is involved in priming for enhanced defense during rhizobacteria-induced systemic resistance in Arabidopsis thaliana. New Phytologist 180(2):511-523

Ramakrishna A, Ravishankar GA (2011) Influence of abiotic stress signals on secondary metabolites in plants. Plant Signal Behav. 6(11):1720-1731

Ren A, Li MJ, Shi L, Mu DS, Jiang AL, Han Q, Zhao MW (2013) Profiling and quantifying differential gene transcription provide insights into ganoderic acid biosynthesis in Ganoderma lucidum in response to methyl jasmonate. PloS One 8:e65027

Sarikami SG, Balkaya A, Yanmaz R (2009) Glucosinolates within a collection of white head cabbages (*Brassica oleracea* var. *capitata* sub. var. *alba*) from Turkey. African Journal of Biotechnology. 8(19):5046-5052

Shim JS, Lee OR, Kim YJ, Lee JH, Jung DY, In JK, Lee BS, Yang DC (2010). Increased production of ginsenosides by overexpressing PgSQS1 affects on ginseng growth rate. J. Ginseng Res 34(2):98-103

The Gene Ontology Consortium 2010. The Gene Ontology in 2010: extensions and refinements. Nucl Acids Res 38(suppl 1): D331-D335

War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, Sharma HC (2012) Mechanisms of plant defense against insect herbivores. Plant Signal Behav. 7(10): 1306-1320

Wasternack C, Parthier B (1997) Jasmonate-singalled plant gene expresseion. Trends in Plant Science 2:302-307

Aoyagi H, Kobayashi Y, Yamada K, Yokoyama M, Kusakari K, Tanaka H (2001) Efficient production of saikosaponins in *Bupleurum falcatum* root fragments combined with signal transducers. Applied Microbiology and Biotechnology. 57(4): 482-488

Yukimine Y, Tabata H, Higashi Y, Hara Y (1996) Methyl jasmonateinduced overproduction of paclitaxel and baccatin III in Taxus cell suspension cultures. Nat Biotechnol. 14: 1129-1132

Supplementary Table 1 List of *Brassica oleraceae*-specific primers used in this study

S. No	Gene Name	Accession No./EST No.	Specific primer sequence (5'-3')	Tm (°C)
1	*Squalene monooxygenase (SQMO)*	isotig04516	**F-** GATTCGCCCTTCTTCACGA **R-** AGGGAAGTTGTTGCCAACAA	57.8
2	*Squalene Synthase 1 (SS1)*	isotig03319	**F-** CGTGATCCTTCTATATTTCGG **R-** TGAGATATGCAAAGACTATAGCCAG	54.5
3	*Flavin-containing monooxygenase (FMO)*	FJ376070	**F-** ATGGGTTTCAGGGATTTTCCG **R-** CTTCGGAGAGATCATCGGAGTT	56
4	*N-hydroxylase for short chain methionine derivative (CYP79F1)*	FJ376044	**F-** TGTCCATGGCATCAATCAGTTTG **R-** GTGTGTTCCGGCGAAGTTGAAA	58
3	*Actin*	DC03003B12	**F-** CGTGATCTTACAGATAGCTTGATGA **R-** AGAGAAGCTAAGATTGATCCTCC	55

Supplementary Table 2 Comparative GO-BP between control and MeJA-treated samples

	GO Description	Control		Meja	
		contigs#	contigs%	contigs#	contigs%
1	response to cadmium ion	873	1.59	693	1.87
2	response to salt stress	861	1.57	750	2.02
3	oxidation-reduction process	588	1.07	478	1.29
4	response to cold	505	0.92	346	0.93
5	regulation of plant-type hypersensitive response	503	0.92	286	0.77
6	serine family amino acid metabolic process	500	0.91	243	0.66
7	protein targeting to membrane	497	0.91	281	0.76
8	response to chitin	496	0.90	220	0.59
9	regulation of transcription, DNA-dependent	490	0.89	247	0.67
10	glycolysis	488	0.89	374	1.01
11	defense response to bacterium	444	0.81	290	0.78
12	gluconeogenesis	440	0.80	326	0.88
13	protein phosphorylation	382	0.70	173	0.47
14	response to wounding	376	0.69	340	0.92
15	negative regulation of defense response	357	0.65	164	0.44
16	defense response to fungus	351	0.64	200	0.54
17	positive regulation of transcription, DNA-dependent	344	0.63	176	0.47
18	Golgi organization	340	0.62	336	0.91
19	jasmonic acid mediated signaling pathway	336	0.61	197	0.53
20	pentose-phosphate shunt	327	0.60	180	0.49
21	translation	327	0.60	150	0.40
22	systemic acquired resistance, salicylic acid mediated signaling pathway	326	0.59	153	0.41
23	response to abscisic acid stimulus	324	0.59	252	0.68
24	ribosome biogenesis	322	0.59	155	0.42
25	electron transport	319	0.58	244	0.66
26	MAPK cascade	317	0.58	155	0.42
27	response to water deprivation	313	0.57	280	0.75
28	isopentenyl diphosphate biosynthetic process, mevalonate-independent pathway	300	0.55	168	0.45
29	cysteine biosynthetic process	298	0.54	219	0.59
30	cellular membrane fusion	294	0.54	177	0.48
31	embryo development ending in seed dormancy	293	0.53	150	0.40
32	rRNA processing	276	0.50	124	0.33
33	water transport	273	0.50	229	0.62
34	photorespiration	268	0.49	162	0.44

Supplementary Table 3 Comparative GO-CC analysis between control and MeJA-treated samples

	GO Description	Control		Meja	
		contigs#	contigs%	contigs#	contigs%
1	plasma membrane	2450	10.23	1650	10.14
2	nucleus	2422	10.12	1426	8.77
3	cytosol	1948	8.14	1316	8.09
4	chloroplast	1533	6.40	983	6.04
5	plasmodesma	1073	4.48	796	4.89
6	Golgi apparatus	991	4.14	691	4.25
7	mitochondrion	976	4.08	764	4.70
8	integral to membrane	789	3.30	544	3.34
9	chloroplast stroma	719	3.00	460	2.83
10	chloroplast envelope	664	2.77	410	2.52
11	vacuolar membrane	637	2.66	487	2.99
12	cytoplasm	627	2.62	368	2.26
13	transcription factor complex	584	2.44	270	1.66
14	nucleolus	521	2.18	334	2.05
15	apoplast	501	2.09	360	2.21
16	extracellular region	499	2.08	360	2.21
17	membrane	436	1.82	275	1.69
18	endoplasmic reticulum	392	1.64	267	1.64
19	cell wall	385	1.61	286	1.76
20	vacuole	349	1.46	342	2.10
21	chloroplast thylakoid membrane	333	1.39	219	1.35
22	ribosome	219	0.91	133	0.82
23	trans-Golgi network	208	0.87	130	0.80
24	endosome	201	0.84	127	0.78
25	plant-type cell wall	184	0.77	130	0.80
26	peroxisome	183	0.76	140	0.86
27	cytosolic large ribosomal subunit	161	0.67	65	0.40
28	cytosolic ribosome	138	0.58	96	0.59
29	thylakoid	102	0.43	65	0.40
30	cytosolic small ribosomal subunit	102	0.43	61	0.38
31	plastoglobule	92	0.38	71	0.44
32	protein serine/threonine phosphatase complex	83	0.35	51	0.31
33	stromule	83	0.35	50	0.31
34	plant-type vacuole membrane	81	0.34	69	0.42

Supplementary Table 4 Comparative GO-MF analysis between control and MeJA-treated samples

	Pathway	Control		Meja	
		contigs #	contigs %	contigs #	contigs %
1	Purine metabolism	243	4.98	148	3.84
2	Starch and sucrose metabolism	239	4.90	169	4.39
3	Glycolysis / Gluconeogenesis	145	2.97	100	2.60
4	Amino sugar and nucleotide sugar metabolism	140	2.87	106	2.75
5	Carbon fixation in photosynthetic organisms	129	2.64	90	2.34
6	Methane metabolism	128	2.62	88	2.29
7	Glutathione metabolism	115	2.36	73	1.90
8	Pyruvate metabolism	109	2.23	103	2.67
9	Cysteine and methionine metabolism	104	2.13	98	2.54
10	Nitrogen metabolism	103	2.11	98	2.54
11	Glyoxylate and dicarboxylate metabolism	103	2.11	60	1.56
12	Citrate cycle (TCA cycle)	101	2.07	64	1.66
13	Glycine, serine and threonine metabolism	88	1.80	72	1.87
14	Oxidative phosphorylation	87	1.78	85	2.21
15	Phenylalanine metabolism	83	1.70	74	1.92
16	Glycerolipid metabolism	79	1.62	65	1.69
17	Carbon fixation pathways in prokaryotes	79	1.62	59	1.53
18	Pentose phosphate pathway	79	1.62	42	1.09
19	Phenylpropanoid biosynthesis	78	1.60	67	1.74
20	Fatty acid metabolism	75	1.54	70	1.82
21	Arginine and proline metabolism	73	1.50	61	1.58
22	Tryptophan metabolism	69	1.41	56	1.45
23	Glycerophospholipid metabolism	69	1.41	36	0.93
24	Galactose metabolism	66	1.35	52	1.35
25	Drug metabolism - other enzymes	65	1.33	56	1.45
26	Alanine, aspartate and glutamate metabolism	64	1.31	48	1.25
27	Fructose and mannose metabolism	63	1.29	42	1.09
28	Aminoacyl-tRNA biosynthesis	63	1.29	16	0.42
29	Pentose and glucuronate interconversions	61	1.25	44	1.14
30	Pyrimidine metabolism	61	1.25	37	0.96
31	Drug metabolism - cytochrome P450	60	1.23	43	1.12
32	Valine, leucine and isoleucine degradation	57	1.17	52	1.35
33	Metabolism of xenobiotics by cytochrome P450	56	1.15	42	1.09
34	Porphyrin and chlorophyll metabolism	55	1.13	52	1.35

Supplementary Table 5 Comparative KEGG pathway analysis between control and MeJA-treated samples

	GO Description	Control		Meja	
		contigs#	contigs%	contigs#	contigs%
1	ATP binding	1271	7.21	694	5.99
2	protein binding	948	5.38	656	5.66
3	zinc ion binding	674	3.83	410	3.54
4	sequence-specific DNA binding transcription factor activity	560	3.18	262	2.26
5	DNA binding	508	2.88	277	2.39
6	nucleotide binding	419	2.38	249	2.15
7	protein serine/threonine kinase activity	366	2.08	162	1.40
8	structural constituent of ribosome	347	1.97	165	1.42
9	RNA binding	263	1.49	136	1.17
10	GTP binding	243	1.38	134	1.16
11	copper ion binding	240	1.36	182	1.57
12	electron carrier activity	237	1.35	177	1.53
13	metal ion binding	221	1.25	165	1.42
14	protein tyrosine kinase activity	202	1.15	78	0.67
15	iron ion binding	189	1.07	197	1.70
16	calcium ion binding	187	1.06	122	1.05
17	nucleic acid binding	170	0.96	78	0.67
18	heme binding	168	0.95	139	1.20
19	ubiquitin-protein ligase activity	144	0.82	87	0.75
20	sequence-specific DNA binding	144	0.82	55	0.47
21	magnesium ion binding	141	0.80	94	0.81
22	calmodulin binding	138	0.78	71	0.61
23	GTPase activity	137	0.78	81	0.70
24	unfolded protein binding	129	0.73	54	0.47
25	protein homodimerization activity	109	0.62	87	0.75
26	translation initiation factor activity	103	0.58	71	0.61
27	2-alkenal reductase [NAD(P)] activity	97	0.55	90	0.78
28	protein dimerization activity	96	0.54	48	0.41
29	flavin adenine dinucleotide binding	93	0.53	57	0.49
30	pyridoxal phosphate binding	85	0.48	65	0.56
31	protein serine/threonine phosphatase activity	85	0.48	57	0.49
32	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	81	0.46	53	0.46
33	ATPase activity	79	0.45	34	0.29
34	oxygen binding	77	0.44	70	0.60