Full Length Article

Discovery of Pathogenesis Related and Effector Genes of *Erwinia mallotivora* in *Carica papaya* (Eksotika I) Seedlings *via* Transcriptomic Analysis

Nor Mustaiqazah Juri[1], Aimera Farhana Samsuddin[1,2], Abdul Munir Abdul Murad[2], Amin Asyraf Tamizi[1], Noor Azmi Shaharuddin[3] and Norliza Abu Bakar[1*]

[1]*Biotechnology & Nanotechnology Research Centre, Malaysian Agricultural Research & Development Institute, 43400 MARDI Head Quarter Serdang, Selangor, Malaysia*
[2]*School of Bioscience and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia*
[3]*Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor Malaysia*
[*]For correspondence: lizaab@mardi.gov.my; mustaiqazah@yahoo.com

Abstract

A comparative transcriptome of *Erwinia mallotivora* library across early infection time points (6, 24 and 48 h) on papaya seedling *Carica papaya* (Eksotika I) was performed. A total of 5,680 genes were identified as differentially expressed genes (DEGs). The highest numbers of DEGs in all three *E. mallotivora* infection time points were accounted in the biosynthesis of secondary metabolites, microbial metabolism in diverse environments and ATP binding cassette transporters based on KEGG-based analysis. The functional annotation of the DEGs *via* Gene Ontology analysis has revealed a highly complex (more than 2,000 functional terms) yet a specific virulence strategy adapted by *E. mallotivora* across the infection time points. Our findings have uncovered the key factors and pathogenicity mechanism adopted by *E. mallotivora* as the infection progresses. © 2020 Friends Science Publishers

Introduction

Carica papaya or papaya is a perennial plant that belongs to the Caricaceae family, under the genus *Carica*. Papaya fruit is highly nutritious and is normally consumed as fresh fruit, vegetable or used as a processed product (Milind and Gurditta 2011). The leaves, roots, fruit and stems of papaya are well known for their medicinal properties (Macalood *et al*. 2013). Papaya is the first flowering plant subjected to genome sequencing and is the subject of many biotechnology studies worldwide due to its importance (VanBuren and Ming 2014). Papaya world production is dominated by India and Mexico with total annual production amounting to 10.4 million tonnes (Evans and Ballen 2012). During year 2004, papaya was used to be widely cultivated in Malaysia due to popular demand with an estimated export worth of about RM100–120 million annually, amounting to a total volume of 58,149 MT per year (Anon 2006). This, however, was the scenario before the occurrence of papaya dieback disease; a devastating disease inflicted by a Gram-negative bacterium, *Erwinia mallotivora*. The threat of papaya dieback disease has

caused the papaya industry in Malaysia to collapse and has since become the major causal agent for the rapid decline of quality papaya production in Malaysia (Bakar *et al.* 2017).

E. mallotivora is a part of the family of Enterobacteriaceae and is a Gram-negative bacterium which was first isolated in 1976 from a bacterial leaf spot disease in *Mallotus japonicus* plant (Gardan *et al*. 2004). Similar to other plant bacterial pathogens, *E. mallotivora* enters its papaya host through stomata and wound openings. Subsequently, the entire parts of the papaya plant, including shoot, leaf, frond, bar and also the fruit will be infected, which leads to early slimy, greasy, water-soaked patches as well as spot signs and symptoms on the foliage and petioles. Eventually, this results in necrosis, blemished or premature fruit drop followed by the death of the papaya plant (Fullerton *et al*. 2011; Bakar *et al*. 2017).

Papaya is the source of environmental niche and nutrient reservoir for several microbial pathogens, including *E. mallotivora* (Redzuan*et al.* 2014; Bakar *et al.* 2017). Production of highly specific sets of effectors enables plant pathogenic bacteria to manipulate and exploit the host cellular pathways for their own benefit and survival

(Melotto and Kunkel 2013). These effectors are essential for providing nutrients, cell-to-cell communication, detoxification of the environment and killing of potential competitors (Kunkel and Chen 2006; Deng *et al.* 2010). Upon gaining entry into their specific hosts, they proceed to multiply, move within and lastly exit the hosts for any new infection cycle to colonize their host plants (Djamei *et al.* 2011). In the past few years, myriads of plant pathogens have been sequenced and with the availability of bioinformatic tools, it is now easy to predict genes that potentially encode virulence factors, pathogenesis related according to sequence resemblance of a known pathogenic protein already contained in the database (Deng *et al.* 2003; Anders *et al.* 2015). However, characterization and validation of the information obtained from the *in silico* effectors researches need to be carried out. The emergence of the field of pathogen effector biology has valuable implications for breeding and deployment of disease-resistance strategy for plant protection against pathogens. The aim is to control plant pathogen and improved the plant disease resistance for increased food safety and security (Fanny *et al.* 2018).

E. *mallativora* produces and secretes virulence factors that are essential to its pathogenesis, colonization and causes disease in papaya (Redzuan*et al.* 2014; Bakar *et al.* 2017). The *E. mallotivora* genome was sequenced by our group and bioinformatic tools were used to predict genes putative virulence factors based on known effectors protein in the database (Redzuan*et al.* 2014). The majority of *E. mallotivora* genes were shown to be related to iron homeostasis, indicating the importance of iron to the bacteria's survival and pathogenesis. Proteomics analysis of the pathogen grown in effectors-inducing media was also carried out revealing overexpression of putative virulent factors, which include proteases, hydrolase, achromobactin, Type III secretion system and chorismate mutase (Bakar *et al.* 2017). However, the pathogenic mechanism of *E. mallotivora* has not been fully explored yet even though knowledge of the underlying mechanism of the papaya dieback disease is crucial, especially for the management and future studies of the disease. Knowledge related to the pathogen effectors especially the *E. mallotivora* effectors protein is crucial to design an efficient screening system to identify the resistance (R) gene candidates in papaya. Furthermore, characterization of the bacterial virulence factors target in the host can also provide unique insights into the basic plant cellular processes such as vesicle trafficking, hormone signalling and innate immunity that can be helpful in designing strategies to manage the plant pathogen for long lasting and sustainable disease resistance in plants (Prasannath 2013; Piqué *et al.* 2015).

A set of powerful and comprehensive approaches is required to elucidate important plant pathogen gene functions (Ferreira *et al.* 2016). Sequencing of either plant genome or plant pathogen has offered ample sequence information particularly for plants or phytopathogen, which

carries a huge impact on the economy (Fanny *et al.* 2018). However, annotation of the sequences obtained to reveal their functions, requires laborious effort. Despite various analyses in well-studied plant–pathogen interaction model plants such as rice and *Arabidopsis*, the majority of the genes are still uncovered especially involving the phytopathogen during their direct interactions with their host (Kunkel and Chen 2006). The latest advance technologies in biotechnology which includes system biology approaches, genomics, transcriptomics, proteomics and metabolomics platform are needed to understand the regulation of important genes during *E. mallotivora* infection in papaya for the improvement of the disease management and ultimately the improvement of crop yield and quality.

In this study, *E. mallotivora* were infected into papaya seedlings and samples were taken at different time points. Identification of potential proteins associated with hypersensitive response and pathogenicity (hrp) from *E. mallotivora* was carried out *via* transcriptomic and bioinformatic technology. Semi-quantitative polymerase chain reaction (PCR) was carried out to further validate the expression of these putative virulent factors and hypersensitive response proteins. Our aim is to understand the active molecular mechanisms in the pathogenicity of *E. mallotivora* which can be crucial in determining its pathogenicity mechanism through transcriptomic study.

Materials and Methods

Plant materials, bacterial inoculation and sampling

The 4-months old seedlings of *C. papaya* (Eksotika I) were supplied by the Malaysian Agricultural Research and Development Institute (MARDI) Pontian, Johor. Twelve healthy uniform-sized seedlings were grown in a greenhouse, where they received 13 h of light a day (30°C). *E. mallotivora* strains were cultured in LB broth and were let to grown at 28°C incubator shaker until reaching OD_{600} of 1. About 5 mL of *E. mallotivora* (EM) culture was injected into the stem of all 12 seedlings at around 15 cm from the shoot area. For the control sample (EM0), the plant samples were collected directly after the injection was made (0 h), followed by the next sampling time points at 6 h (EM6), 24 h (EM24) and 48 h (EM48) after *E. mallotivora* inoculation. All plant samples were immediately immersed in liquid nitrogen and stored at –80°C. The experimental design adopted complete randomized design (CRD) with three replicates for each sampling time point.

Total RNA extraction, cDNA library construction and sequencing

The extraction of *E. mallotivora* total RNA from papaya plant samples was conducted by following the protocol reported by Schenk *et al.* (2008). The extracted total RNAs

were subjected to integrity and purity assessments via gel electrophoresis, Nanodrop spectrophotometer, Qubit and digital electrophoresis using Bio-analyzer (Chan *et al.* 2017). The triplicate good quality RNAs in each time point (EM0, EM6, EM24 and EM48) were pooled and were subsequently used for cDNA library preparation and sequenced by Novogene (U.S.A.). Roughly, the process started with mRNA fragmentation and double-stranded (ds) cDNA synthesis. The ds cDNAs were end-repaired, polyadenylated, ligated with adapter sequences and size-selected using AMPure XP beads. The remaining strands were amplified using PCR before introduction to an Illumina HiSeq^{TM} 2000 instrument for paired-end sequencing.

Reads mapping and transcripts assembly

Prior to mapping, raw sequenced reads were subjected to data filtration by removal of adapters, low-quality reads with >10% of uncertain nucleotides and reads with >50% of low-quality nucleotides (Q_{phred} score $\leq$ 5). The remaining high-quality reads were mapped to the *E. mallotivora* BT-MARDI reference genome (Redzuan*et al.* 2014) by Bowtie2. The mismatch parameter was set to two, and other parameters were set to default. Reads were assembled according to the reference genomes using Rockhopper (McClure *et al.* 2013). Functional annotation was conducted by subjecting the transcripts to NCBI non-redundant (Nr) database through Blastx (cut-off: evalue < 1e–5). Meanwhile, Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/) pathways of transcripts were identified by the GOSeq and KOBAS programs, respectively.

Identification of differentially expressed genes (DEGs)

Fragments per kilobase of exon per million fragments mapped (FPKM) was used as a unit for the quantification, whereas FDR cut-off < 0.05, absolute log2 ratio (Log_2FC) > 1 and q-value < 0.005 were set as the thresholds to identify the significant genes.

Validation of RNA-seq expression

The reproducibility of the DEGs identified from the RNA-seq experiment was further validated by semi-quantitative RT-PCR. Fifteen representative DEGs and primers were selected (Table 1 and 2). The image of ethidium bromide-stained PCR products in agarose gels were quantified using ImageJ software (Schneider *et al.* 2012; Antiabong *et al.* 2016). An identical rectangular selection was used to measure the intensity of each band of PCR product. Analysis on the quantified band intensity of each candidate gene in different cDNA samples was performed according to Luke Miller's method at lukemiller.org/index.php/2010/11/analyzing-gels-and-

western-blots-with-image-j. The relative expression values were calculated using the Log_2(Sample1/Sample2) formula with *gltA* as the internal control.

Results

Total RNA extraction and library preparation

Total RNA was extracted from the 4-months old papaya seedlings infected with *E. mallotivora* at different time points, namely, 0 h (EM0), 6 h (EM6), 24 h (EM24) and 48 h (EM48). Results from the Bio-analyzer indicated that the integrity of the extracted total RNA is intact with RNA Integrity Number (RIN) of 7.4, 7.8, 8 and 7.4 for EM0, EM6, EM24 and EM48, respectively.

RNA-sequencing and mapping

From the RNA-sequencing, there are a total of 19,658,966, 18,718,754, 14,381,708 and 14,557,640 raw reads generated for samples EM0, EM6, EM24 and EM48, respectively. The numbers of clean reads after the removal of low-quality reads, which is around 3.4% of total raw reads, are 18,977,814, 18,075,418, 13,896,170 and 14,062,280 for EM0, EM6, EM24 and EM48, respectively. When base quality filtering was performed using Phred values of more than 20 and 30, the results showed that about 97 and 92% of the total nucleotides had fulfilled the requirement. The low percentage of low-quality reads and high percentage of nucleotides that fulfilled Q_{phred} score $\geq$ 30 suggested that the raw data generated from the RNA-sequencing process were of high quality and the clean reads were suitable to be used in the mapping. The results from the data quality control and raw data pre-processing are summarized in Table 3.

It is important to use only high-quality and clean reads for mapping because the low-quality reads will affect the mapping process and will result in a low total mapped rate and high percentage of reads with multiple mapping sites against the reference genome. In this study, the total clean reads were mapped against *E. mallotivora*'s whole transcriptome as the reference as we are interested in looking at the transcriptomic changes of *E. mallotivora* post infection in papaya seedlings. The total mapped reads from the total clean reads are 102,421 (0.54%), 97,641 (0.54%), 497,285 (3.58%) and 449,644 (3.2%) for sample EM0, EM6, EM24 and EM48, respectively. In general, the amount of mapped reads from the total clean reads is low because the total RNAs used in RNA-sequencing were comprised of total RNAs that originated from both the papaya seedlings tissue (majority) and *E. mallotivora*. Considering the mapping was performed against the *E. mallotivora*'s whole transcriptome as a reference, only reads specific to *E. mallotivora* were mapped while the reads sequenced from the papaya seedlings were not mapped. Hence, the percentages of the mapped reads were low as compared with the total clean reads. Then, it can be

Table 1: List of candidate genes validated by semi-quantitative RT-PCR

Gene ID	Annotation	Gene expression of illumina (Log2$^{(Sample1/Sample2)}$)		
		6 h	24 h	48 h
6666666.159625.peg.1092	Ammonium transporter	7.6374	8.7431	9.4059
6666666.159625.peg.1253	DspF	6.6374	6.5207	5.8698
6666666.159625.peg.1735	Extracellular metalloprotease precursor (EC 3.4.24.-)	2.5537	3.2431	3.4019
6666666.159625.peg.1302	Ferric iron ABC transporter2C iron-binding protein	9.4448	8.3508	8.0073
6666666.159625.peg.1236	HrpF	5.6374		3.5479
6666666.159625.peg.1221	HrpP	7.6374		3.5479
6666666.159625.peg.1240	HrpV	7.6374	4.9357	4.5479
6666666.159625.peg.1251	HrpW	12.296	10.079	9.923
6666666.159625.peg.1245	Hypothetical protein	10.959	7.9946	7.3553
6666666.159625.peg.1227	RNA polymerase sigma factor RpoE	12.012	9.0232	8.1918
6666666.159625.peg.1232	Type III secretion protein HrpB(Pto)	10.03	7.1581	6.5479
6666666.159625.peg.1234	Type III secretion protein HrpD	9.8074	6.3508	6.8698
6666666.159625.peg.1236	Type III secretion protein HrpG	7.9594	6.1581	5.5479
6666666.159625.peg.1224	Type III secretion protein HrpQ	8.9594	6.3508	3.5479
6666666.159625.peg.145	Virulence factor VirK		5.3508	6.7178

Table 2: Description of designated primer of candidate and housekeeping genes

Gene name	Sequence (5'->3')	Tm	GC%	Product length
Ammonium transporter (*AMT*)	TACCTGGTGGGTAAACGTGC	59.96	55	145
	CCGCAATTTCATTAGCCGCA	59.9	50	
HrpP	TTTCACTGCGTTTTTCGCCG	60.59	50	124
	TATGCGCTTCCCTGCTCAAT	59.82	50	
Type III secretion protein HrpQ	GCGCACCTTCTATCACCACT	60.11	55	121
	GACGATATTGGCGTGTGTGC	59.97	55	
RNA polymerase sigma factor RpoE	CCCAGCCAGATTACCGAACA	59.75	55	121
	CCTGATAGCTGCCGTCCTTC	60.25	60	
Type III secretion protein HrpB(Pto)	TTAGCCATCAGTACCGGGGA	60.03	55	118
	GGTTATCCTGAGTGTCCGCC	60.18	60	
Hypothetical protein	GTTATGGTATGTCACGCGCC	59.42	55	144
	GGCTGGGCGAGTCTTTGATA	59.82	55	
HrpW	CGAGGATGCGATTACCGTGA	59.97	55	119
	GGTCGGTATCCGCATTCAGT	59.9	55	
Ferric iron ABC transporter2C iron-binding protein	TGGTGCAGTCATGGGTTGAT	59.59	50	134
	GGTCAGAAAAACGTCGGCAG	59.77	55	
Virulence factor VirK	GACCTTTGCCGTTGTGTGTC	59.97	55	125
	GGAACAGGCCATAACAGGCT	60.03	55	
Extracellular metalloprotease precursor (EC 3.4.24.-)	GACGGTGACGGGGAGATTTT	60.04	55	132
	CGATTCATTCAGTGCGCCAG	59.97	55	
Type III citrate synthase (glta)	AACACATTGCGCTGAACGAC	60.04	50.00	157
	CAGTGTGCAATCCAGCCAAC	60.04	55.00	

Table 3: Summary of raw reads quality control and preprocessing

Sample name	Raw reads	Clean reads	Q20(%)	Q30(%)
EM0	19,658,966	18,977,814	97.27	92.79
EM6	18,718,754	18,075,418	96.96	92.11
EM24	14,381,708	13,896,170	97.19	92.59
EM48	14,557,640	14,062,280	97.25	92.71

observed that the amount of total mapped reads was increased up to fivefold for EM24 and EM48 as compared with EM0 and EM6. This could indicate that transcriptomic changes including overexpression and production of different transcripts responsible for pathogenesis such as virulent genes are present, whereby these transcripts are generally not expressed under normal conditions. As for the analysis on the specificity of the mapped reads, the numbers of multiple mapped reads are 3176 (3.1%), 2180 (2.2%), 10108 (2.0%) and 9273 (2.1%) for the sample EM0, EM6, EM24 and EM48, respectively. The percentages of the multiple mapped reads for all samples are lower than 10% (2–3%) of the total mapped reads and this indicates that a good mapping coverage of the reads was obtained as more uniquely mapped reads were obtained. A summary of the mapping of the clean reads is shown in Table 4.

Quantification of gene expression

HTSeq v0.6.1 software (Anders *et al.* 2015) was used to measure the genes expression level in each library, while FPKM was used as the unit of measurement (Mortazavi *et al.* 2008). In our RNA-seq analysis, the gene expression level is estimated by counting the reads mapped to genes or exons. A total of 5,680 genes were detected in all four libraries. The largest percentage (46 to 48.82%) of genes in

Table 4: Summary of the clean reads mapped to *E. mallotivora*'s transcriptome as reference

Sample name	EM0	EM6	EM24	EM48
Total reads	18,977,814	18,075,418	13,896,170	14,062,280
Total mapped	102421 (0.54%)	97641 (0.54%)	497,285 (3.58%)	449,644 (3.2%)
Multiple mapped	3,176 (3.1%)	2,180 (2.2%)	10,108 (2.0%)	9,273 (2.1%)
Uniquely mapped	99,245 (96.9%)	95,461 (97.8%)	487,177 (98.0%)	440,371 (97.9%)

Table 5: The number of genes with different expression levels

| FPKM Interval	EM0	EM6	EM24	EM48
0~1	2,638 (46.44%)	2,386 (42.01%)	1,330 (23.42%)	1,404 (24.72%)
1~3	0 (0.00%)	0 (0.00%)	11 (0.19%)	17 (0.30%)
3~15	51 (0.90%)	55 (0.97%)	363 (6.39%)	390 (6.87%)
15~60	698 (12.29%)	626 (11.02%)	1,203 (21.18%)	1,184 (20.85%)
>60	2,293 (40.37%)	2613 (46.00%)	2,773 (48.82%)	2,685 (47.27%)
Total	5,680	5,680	5,680	5,680

the EM6, EM24 and EM48 library were expressed at FPKM interval >60. In contrast, for the EM0 library, the highest percentage (46.44%) of their total genes was expressed at FPKM interval of 0–1 (Table 5). This dominance of genes with high expression level (FPKM > 60) in the EM6, EM24 and EM48 as compared with the EM0 in which the majority of the genes were expressed at a very low expression level (0–1) serves as an early indicator in the presence/activation of a high number of pathogenicity-related genes in the generated libraries. The expression level of all 5,680 genes in each cDNA library (EM0, EM6, EM48 and EM24) were represented in the cluster analysis (Fig. 1).

Identification of differentially expressed genes (DEGs)

Comparative differential expression of genes in each *E. mallotivora*'s post-inoculation time point library (EM6 *vs.* EM0, EM24 *vs.* EM0, EM48 *vs.* EM0, EM24 *vs.* EM6, EM48 *vs.* EM6 and EM48 *vs.* EM24) was conducted using the DEGSeq R package. At FDR cut-off < 0.05, absolute log2 ratio (Log$_2$FC) > 1 and q-value < 0.005, a total of 5,680 genes were identified as DEGs in at least one of the *E. mallotivora* libraries (Table 6).

To identify the pathogenic-related genes activated by *E. mallotivora* at the 6, 24 and 48 h of the infection period, comparative expression was made between 0 h *E. mallotivora*'s post-inoculation library (EM0) with other *E. mallotivora*'s infection time points (EM6 *vs.* EM0, EM24 *vs.* EM0 and EM48 *vs.* EM0). Specifically, 2,303, 2,504 and 2,888 DEGs were identified at 6, 24 and 48 h of *E. mallotivora* post inoculation in which 1,029, 719 and 820 genes were down-regulated and 1,274, 1,785 and 2,068 genes were up-regulated at 6, 24 and 48 h of *E. mallotivora* post inoculation, respectively (Fig. 2A). Based on this trend of DEGs accumulation, it can be seen that the total number of pathogenicity-related DEGs was increased as the *E. mallotivora* infection progresses, in which most of them work by up-regulating their expression.

Meanwhile, to identify the DEGs that were specifically expressed between each infection time point, comparative differential expression was made between EM6 and EM24
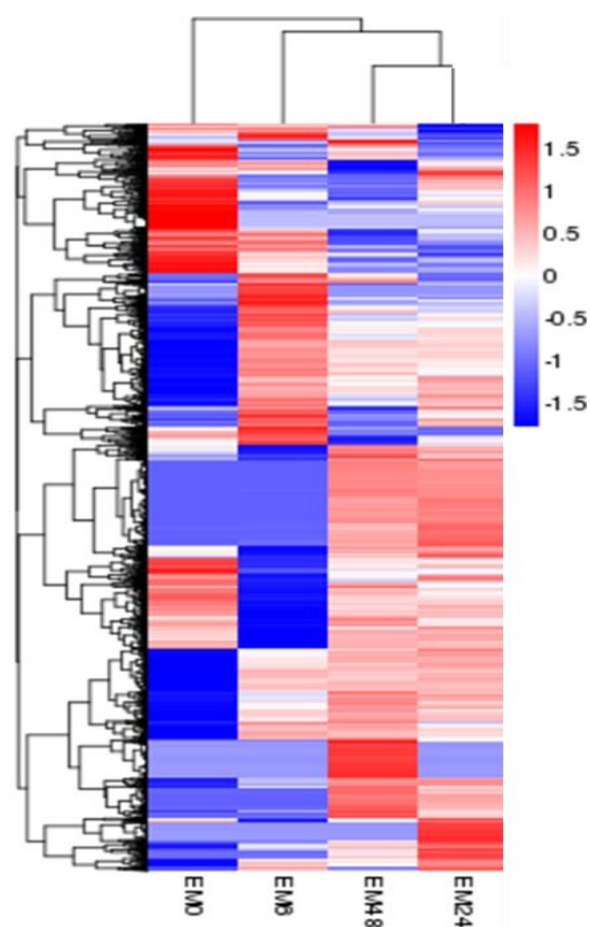
Fig. 1: The overall results of FPKM cluster analysis, clustered using the $\log_{10}$(FPKM+1) value. Each horizontal line refers to one gene. Red denotes genes with high expression levels, and blue denotes genes with low expression levels. The color range from red to blue represents the $\log_{10}$(FPKM+1) value from large to small

(EM24 *vs.* EM6) as well as EM24 and EM48 (EM48 *vs.* EM24).

Table 6: The top 20 hits of DEGs generated via comparative differential expression within libraries

Gene description	$\log_2$ change	Fold q-value
EM6 *vs*. EM0 (6 h)		
FIG00635037: hypothetical protein	12.52	0
HrpW	12.296	1.03E-286
RNA polymerase sigma factor RpoE	12.012	2.33E-248
hypothetical protein	11.66	2.17E-207
hypothetical protein	10.959	2.56E-144
L-arabinose transport ATP-binding protein AraG (TC 3.A.1.2.2)	10.959	2.56E-144
L-arabinose-binding periplasmic protein precursor AraF (TC 3.A.1.2.2)	10.807	2.82E-133
Ribose ABC transport system2C ATP-binding protein RbsA (TC 3.A.1.2.1)	10.637	6.56E-122
FIG00634607: hypothetical protein	10.592	5.36E-119
Type III secretion injected virulence protein (YopP2CYopJ2C induces apoptosis2C prevents cytokine induction2C inhibits NFkb activation)	10.495	4.21E-113
3-isopropylmalate dehydrogenase (EC 1.1.1.85)	10.161	1.04E-94
Type III secretion inner membrane channel protein (LcrD2CHrcV2CEscV2CSsaV)	10.097	1.59E-91
3-dehydro-L-gulonate 2-dehydrogenase (EC 1.1.1.130)	10.097	1.59E-91
type III secretion protein HrpB(Pto)	10.03	2.68E-88
Type III secretion bridge between inner and outermembrane lipoprotein (YscJ2CHrcJ2CEscJ2C PscJ)	10.03	2.68E-88
Gluconolactonase (EC 3.1.1.17)	10.03	2.68E-88
Manganese transport protein MntH	9.9594	4.99E-85
Hydroxymethylpyrimidine phosphate synthase ThiC (EC 4.1.99.17)	9.9594	4.99E-85
Alkanesulfonate monooxygenase (EC 1.14.14.5)	9.8854	1.05E-81
type III secretion protein HrpD	9.8074	2.33E-78
EM24 *vs*. EM0 (24 h)		
Candidate zinc-binding lipoprotein ZinT	11.334	1.10E-171
22C3-butanediol dehydrogenase2C S-alcohol forming2C (R)-acetoin-specific (EC 1.1.1.4) / Acetoin (diacetyl) reductase (EC 1.1.1.5)	11.126	4.31E-154
Acetolactate synthase2C catabolic (EC 2.2.1.6)	10.691	9.76E-123
FIG00635037: hypothetical protein	10.627	1.10E-118
Manganese transport protein MntH	10.449	5.02E-108
Hydroxymethylpyrimidine phosphate synthase ThiC (EC 4.1.99.17)	10.438	2.01E-107
hypothetical protein	10.258	1.31E-97
Alpha-acetolactate decarboxylase (EC 4.1.1.5)	10.246	5.48E-97
Transcriptional regulator2C TetR family	10.145	5.97E-92
3-polyprenyl-4-hydroxybenzoate carboxy-lyase (EC 4.1.1.-)	10.119	1.12E-90
Non-ribosomal peptide synthetase modules2C pyoverdine??	10.119	1.12E-90
HrpW	10.079	9.52E-89
Non-hemolytic enterotoxin lytic component L1	10.051	1.88E-87
Manganese ABC transporter2C ATP-binding protein SitB	9.8899	3.25E-80
5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (EC 2.1.1.14)	9.8102	7.41E-77
iron aquisition yersiniabactin synthesis enzyme (Irp2)	9.7083	9.48E-73
Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)	9.6545	1.12E-70
Carbonic anhydrase (EC 4.2.1.1)	9.6545	9.48E-73
FIG00634911: hypothetical protein	9.5005	1.12E-70
LSU ribosomal protein L31p	9.4593	1.35E-63
EM48 *vs*. EM0 (48 h)		
Candidate zinc-binding lipoprotein ZinT	11.559	3.81E-189
Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)	10.757	1.46E-124
22C3-butanediol dehydrogenase2C S-alcohol forming2C (R)-acetoin-specific (EC 1.1.1.4) / Acetoin (diacetyl) reductase (EC 1.1.1.5)	10.603	7.39E-115
Manganese transport protein MntH	10.559	3.27E-112
FIG00635037: hypothetical protein	10.548	1.50E-111
hypothetical protein	10.479	1.56E-107
iron acquisition yersiniabactin synthesis enzyme (Irp2)	10.248	3.53E-95
Putative membrane protein	10.248	3.53E-95
Transcriptional regulator2C TetR family	10.163	6.25E-91
Acetolactate synthase2C catabolic (EC 2.2.1.6)	10.007	1.06E-83

Table 6: Continued

Table 6: Continued

| Protein				
HrpW			9.923	6.25E-91
FIG00634911: hypothetical protein			9.923	1.06E-83
Manganese ABC transporter2C ATP-binding protein SitB			9.8517	5.31E-80
AmpG protein			9.7574	5.31E-80
Hydroxymethylpyrimidine phosphate synthase ThiC (EC 4.1.99.17)			9.7574	5.39E-77
Ribose ABC transport system2C ATP-binding protein RbsA (TC 3.A.1.2.1)			9.7377	3.46E-73
3-polyprenyl-4-hydroxybenzoate carboxy-lyase (EC 4.1.1.-)			9.7178	3.46E-73
Ferrichrome-iron receptor			9.6564	2.04E-72
Alpha-acetolactate decarboxylase (EC 4.1.1.5)			9.5479	1.19E-71
Non-ribosomal peptide synthetase modules2C pyoverdine??			9.5479	2.52E-69
EM24 vs. EM6				
Transcriptional regulator2C Cro/CI family			11.096	2.25E-65
Acetolactate synthase2C catabolic (EC 2.2.1.6)	10.756	7.22E-124		
Alpha-acetolactate decarboxylase (EC 4.1.1.5)			10.311	4.04E-148
Non-hemolytic enterotoxin lytic component L1			10.117	7.22E-124
Colanic acid biosysnthesis protein WcaK			9.4166	9.04E-98
hypothetical protein			9.117	3.99E-88
hypothetical protein			9.0016	3.00E-60
Dethiobiotin synthetase (EC 6.3.3.3)			8.702	4.01E-51
probable exported protein YPO3518			8.6645	5.64E-48
Mobile element protein			8.626	1.21E-40
hypothetical protein			8.5865	8.30E-40
Putative inner membrane protein			8.461	5.72E-39
Putative arylsulfatase regulatory protein			8.4166	4.05E-38
Putative membrane protein			8.4166	1.52E-35
Cysteine desulfurase CsdA-CsdE (EC 2.8.1.7)2C main protein CsdA			8.3708	1.12E-34
Phage protein			8.3708	1.12E-34
O-antigen acetylase			8.3235	8.48E-34
hypothetical protein			8.2746	8.48E-34
N-acetylmuramic acid 6-phosphate etherase			8.2746	6.51E-33
hypothetical protein			8.2746	4.84E-32
EM48 vs. EM6				
Putative membrane protein			10.319	3.89E-96
Transcriptional regulator2C Cro/CI family	10.127	1.07E-86		
hydrolase			10.127	3.89E-96
Acetolactate synthase2C catabolic (EC 2.2.1.6)			10.078	1.07E-86
Alpha-acetolactate decarboxylase (EC 4.1.1.5)			9.619	1.07E-86
probable exported protein YPO3518			9.5732	1.82E-84
hypothetical protein			9.3194	7.50E-66
Non-hemolytic enterotoxin lytic component L1			9.2628	3.10E-64
hypothetical protein			9.0113	6.40E-56
Colanic acid biosysnthesis protein WcaK			8.9409	3.29E-54
Putative inner membrane protein			8.9044	3.25E-47
Homoserine O-succinyltransferase (EC 2.3.1.46)			8.7889	2.04E-45
Phage protein			8.7483	1.63E-44
Dethiobiotin synthetase (EC 6.3.3.3)			8.7064	9.15E-42
ABC transporter ATP-binding protein			8.5732	7.75E-41
FIG139552: Putative protease			8.5732	6.76E-40
hypothetical protein			8.477	4.72E-37
Putative arylsulfatase regulatory protein			8.4263	4.72E-37
N-acetylmuramic acid 6-phosphate etherase			8.3739	3.95E-35

A total of 2,092 genes were specifically expressed between 6 h to 24 h of infection time points (EM24 *vs*. EM6), of which 1,383 were up-regulated and 709 were down-regulated. In the meantime, 825 genes were specifically expressed between 24 h to 48 h of infection time points (EM48 *vs*. EM24), of which 96 genes were up-regulated and 429 genes were down-regulated. By including the genes generated between 0 h to 6 h infection time points (Fig. 2B), it can be seen that the number of DEGs that specifically expressed within time points was decreased as the infection

Table 7: Validation of DEGs expression *via* sqRT-PCR

Annotation	Gene expression of illumina (Log2$^{(Sample1/Sample2)}$)			Relative gene expression by sqRT-PCR (Log2$^{(Sample1/Sample2)}$)		
	6 h	24 h	48 h	6 h	24 h	48 h
Ammonium transporter	7.6374	8.7431	9.4059		1.29	5.43
DspF	6.6374	6.5207	5.8698	1.47	4.49	4.62
Extracellular metalloprotease precursor (EC 3.4.24.-)	2.5537	3.2431	3.4019		2.56	5.33
Ferric iron ABC transporter2C iron-binding protein	9.4448	8.3508	8.0073		3.15	5.25
HrpF	5.6374		3.5479	1.86	2.78	3.81
HrpP	7.6374		3.5479		4.45	4.76
HrpV	7.6374	4.9357	4.5479	3.95	4.25	4.23
HrpW	12.296	10.079	9.923	2.91	4.56	4.34
Hypothetical protein	10.959	7.9946	7.3553	0.91	2.86	5.24
RNA polymerase sigma factor RpoE	12.012	9.0232	8.1918	2.52	4.53	4.4
Type III secretion protein HrpB(Pto)	10.03	7.1581	6.5479		4.03	5
Type III secretion protein HrpD	9.8074	6.3508	6.8698	-0.6	0.21	2.12
Type III secretion protein HrpG	7.9594	6.1581	5.5479	1.74	4.5	4.6
Type III secretion protein HrpQ	8.9594	6.3508	3.5479	6.02		
Virulence factor VirK		5.3508	6.7178		3.09	5.25
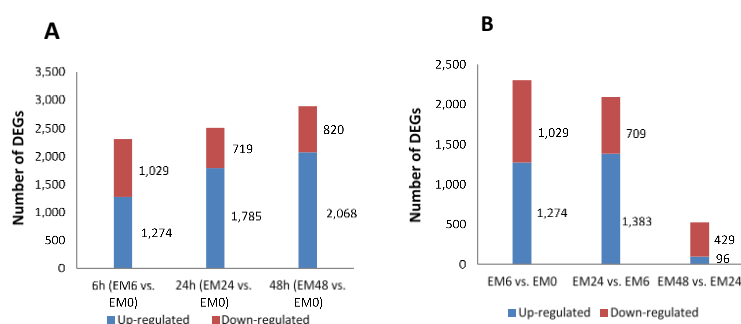
Fig. 2: Statistical chart of DEGs during *E. mallotivora* infection in papaya at seedling stage (**A**) Number of DEGs at each time point (**B**) Number of DEGs between time points

time increased. This suggests that the largest percentage number of genes involved in *E. mallotivora* pathogenesis were activated during the early phase of infection (6 h post inoculation), while a lesser number of genes were activated during the later stage of *E. mallotivora* infection.

Functional annotation of DEGs

Gene ontology (GO) enrichment analysis: Functional categorization of DEGs across each infection time point (6, 24 and 48 h) were identified by subjecting the DEGs of each of *E. mallotivora*'s infection time points for GO annotation analysis. Overall, the DEGs of *E. mallotivora* at 6, 24 and 48 h post inoculation were classified into 2,434, 2,521 and 2,455 functional terms, respectively. The summary of 30 most significant enriched terms were shown in Fig. 3.

During the early phase of infection, 6 h of post inoculation DEGs, the most significant biological process enrichments were in 'response to stimulus', 'transmembrane transport' and 'signal transduction'. For the cellular component, the most significant enrichments occurred in 'cytoplasmic part', 'ribosome', 'intracellular ribonucleoprotein' and 'ribonucleoprotein complex', while

molecular function enrichments were in 'lyase activity', 'molecular transducer activity' and 'structural molecule activity' (Fig. 3A). Meanwhile, during the 24 h of post, the most significant biological process enrichments were in 'regulation of cellular metabolism'. For the cellular component, the most significant enrichments occurred in 'cytoplasm' and 'cytoplasmic part', while molecular function enrichments were in 'structural molecule activity' and 'transferase activity' (Fig. 2B). During 48 h of post inoculation DEGs, the most significant biological process enrichments were in 'localization', 'establishment of localization' and 'transport'. For the cellular component, the most significant enrichments occurred in 'membrane', while molecular function enrichments were in 'transporter activity' (Fig. 3C).

KEGG enrichment analysis: The KEGG pathway analysis showed a total of 80, 78 and 78 pathways were identified at the three *E. mallotivora*'s infection time points, namely 6, 24 and 48 h, respectively. The summary of 20 most pathways indicates the diverse pathogenicity mechanism carried out by *E. mallotivora* to cause the dieback disease in papaya. Interestingly, all three time points showed a similar component of pathways enrichment, in which all their up-regulated DEGs showed the most enriched pathways in the
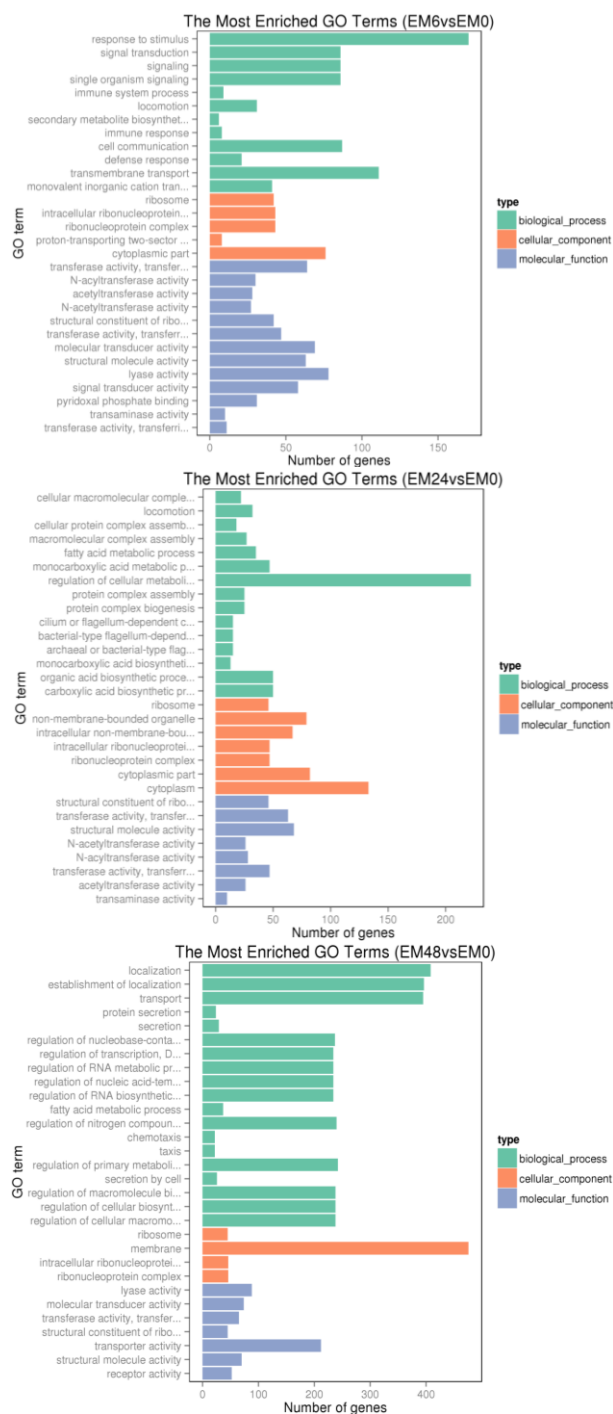
Fig. 3: GO functional enrichment analyzes of DEGs of (A) 6 h, (B) 24 h and (C) 48 h of *E. mallotivora's* post inoculation. The GO enrichment bar chart of DEGs presents the number of DEGs enriched in biological process, cellular component and molecular function. The 30 most significant enriched terms are selected

biosynthesis of secondary metabolites, microbial metabolism in diverse environments and ATP binding cassette (ABC) transporters (Fig. 4). In contrast, their down-

regulated DEGs showed enrichment in the two-component system and ribosome.

Validation of RNA-Seq expression levels by semi-quantitative PCR

The reliability of the DEGs expression generated by RNA-seq analysis was validated by semi-quantitative PCR. Fifteen genes were selected for this purpose. As shown in Table 7, the resulting semi-quantitative expression agreed well with the RNA-seq patterns in which all genes showed up-regulation patterns in all time points relative to 0 h of *E. mallotivora* post inoculation. Thus, the RNA-seq data are reliable. However, we can see that several genes (ammonium transporter, extracellular metalloprotease precursor, ferric iron ABC transporter 2C iron-binding protein, type III secretion protein HrpB(Pto) and HrpP which were supposed to activate their expression within 6 h of infection were only detected after 24 h. This might be due to the low sensitivity of gene profiling *via* sqRT-PCR compared with the RNA-seq, which is able to detect even at a very low expression of the gene. The gel image of the candidate genes expressions can be referred in Fig. 5.

Discussion

From our study, it has been identified that the largest percentage of genes involved in *E. mallotivora* pathogenesis were activated during the early phase of infection, specifically at the 6 h of post inoculation. Based on our data, during the early phase of infection, the strongest up-regulation of DEGs (~12 fold) was expressed by HrpW and RNA polymerase sigma factor RpoE (Table 6). In previous studies, HrpW was classified as a 'helper protein' that assists the translocation of type III secretion system (T3SS) secreted proteins (effectors) into the host (Kunkel and Chen 2006). Harbouring harpin and pectate lyases domains, HrpW binds and cleaves plant cell wall pectic polymers, thus facilitating the assembly of type III secretion complexes at the interface of bacteria and the host cell wall (Kunkel and Chen 2006) and initial penetration of the T3SS pilus (Büttner and He 2009). HrpW was shown to be one of *Pseudomonas cichorii* virulence factor during early stages of infection in lettuce leaves. In addition, mutation in HrpW resulted in the loss of *Pseudomonas cichorii* virulence on aubergine (Kajihara *et al.* 2012). Therefore, it is not surprising *E. mallotivora* also uses the same mechanism as its virulence strategy. Meanwhile, RNA polymerase sigma factor is a regulator gene that controls transcription initiation of hundreds of prokaryotic genes including the virulence and virulence-related genes in bacterial pathogens (Kazmierczak, 2005). This resulted in the enhancement of the capacity of the bacteria to spread their colonization to new individuals or to survive passage through a host organism. Belonging to the extracytoplasmic function (ECF) factors subfamily, this gene contributes to oxidative
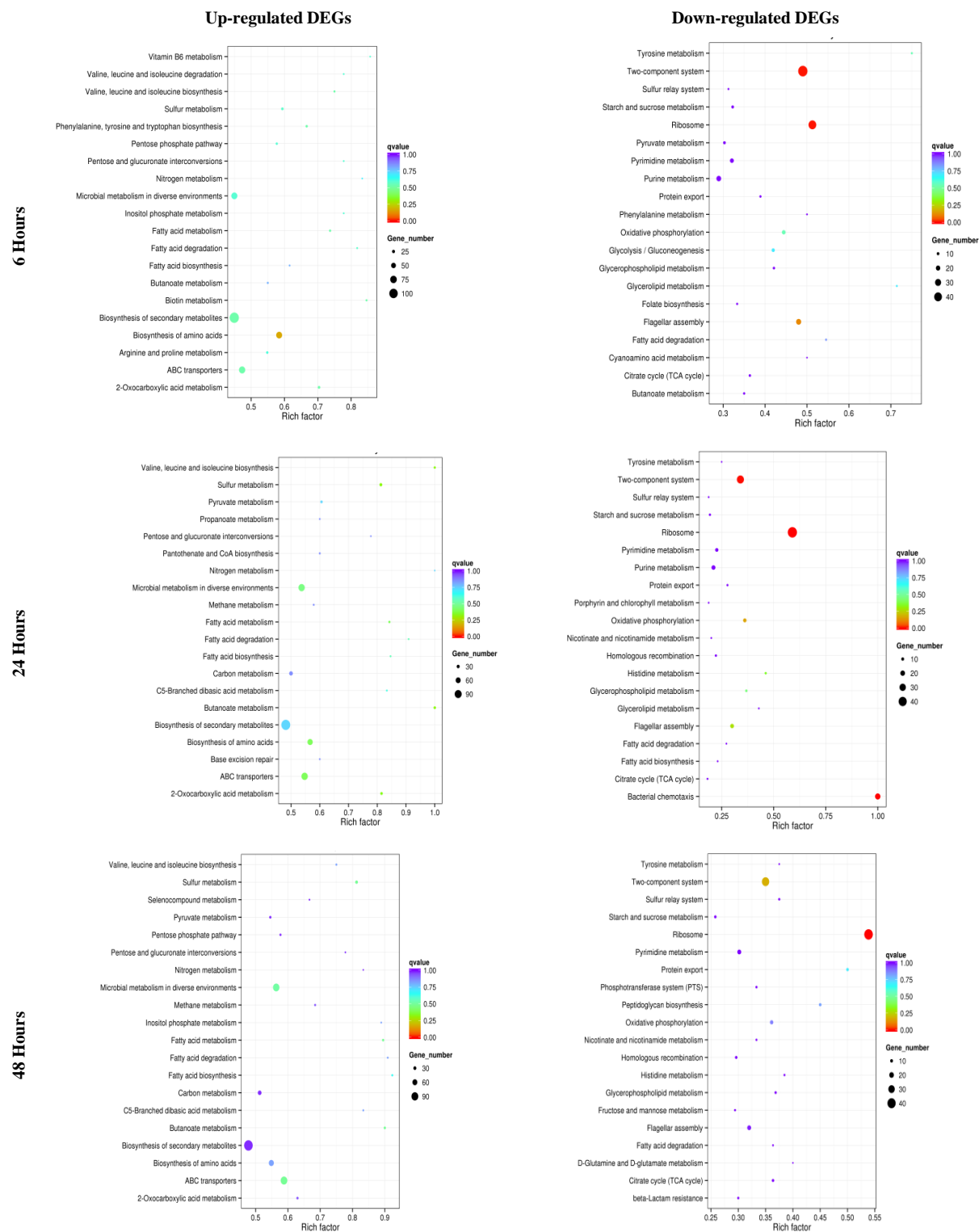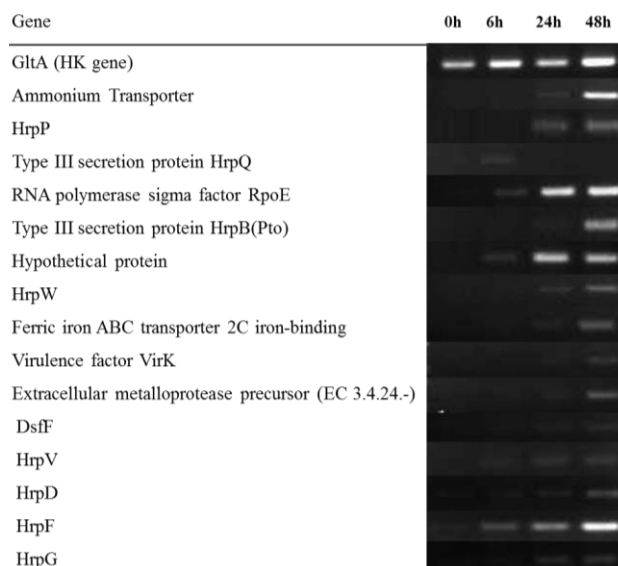
Fig. 4: The 20 most significant enriched KEGG pathways of DEGs

stress resistance in several Gram-negative pathogens, in which initiation of oxidative stress is one of the tolerance mechanisms in a plant in response to pathogen attack (Helmann 2002).

The Gene Ontology (GO) enrichment analysis has demonstrated that GO terms were enriched differently in

Fig. 5: Validation of DEGs expression. The PCR products were run in 2.0% gel electrophoresis

each of the three time points, thus representing the complex, yet specific, virulence strategy of *E. mallotivora* during the colonization period. In terms of molecular function, for example, the enrichment of 78 DEGs involved in lyase activity indicated that degradation of host cell wall is the earliest strategy (6 h) taken by *E. mallotivora* for papaya plant invasion. Pectin degradation protein KdgF and pectate lyase precursor (EC 4.2.2.2) are examples of two genes that are involved in this activity where they were up-regulated seven- to eightfold during this time point. As the infection progressed (24 h), transferase activity was activated via production of proteins and enzymes to counter oxidative stress initiated by the host defence mechanism. Later (48 h), transporter activity is activated for nutrients and water uptake.

In response to the tolerance mechanism mediated by the host in response to pathogen attack, such as the releasing of antimicrobial compounds (Piasecka *et al.* 2015) and initiation of oxidative stress (Sagi and Fluhr 2006), the activation of secondary metabolites and microbial metabolism in diverse environments metabolism pathways was seen as an effective virulence strategy for *E. mallotivora* to facilitate their colonization. For secondary metabolites, a total of 159, 142 and 153 DEGs related to this pathway were identified during 6, 24 and 48 h of *E. mallotivora* infection, respectively. For microbial metabolism in diverse environments metabolism, a total of 102, 94 and 109 DEGs related to this pathway were identified during 6, 24 and 48 h of *E. mallotivora* infection, respectively.

The transport system is an important component to facilitate colonization and modulation of host cell physiology (Melotto and Kunkel 2013). In our data, a total of 73, 82 and 94 ABC transporter-related DEGs were

identified during 6, 24 and 48 h of *E. mallotivora* infection time point, respectively. This large number of ABC transporter-related genes showed the significant involvement of this transport system in transporting the virulence factor/effector to the host cell (papaya). As has been reported in *E. amylovora* and *E. chrysanthemi*, this type I secretion system (T1SS) transport shows great importance for virulence proteins, including proteases, lipases, toxins and scavenging molecules. Governed by active transport through ATP hydrolysis, the unfolded virulence proteins were translocated across the inner and outer bacterial membrane to the external environment (Toth *et al.* 2006; Liu *et al.* 2008).

Fifteen genes that showed a wide range of expression patterns along the 6, 24 and 48 h of *E. mallotivora*'s infection were selected for validation of RNA-Seq data. Those 15 genes also were selected based on their involvement in the pathogenesis-associated mechanism including those that encoded for type III secretion system (T3SS) virulence factors and hypersensitive response proteins. The resulting semi-quantitative expression agreed well with the RNA-seq patterns in which all genes showed up-regulation patterns in all time points relative to 0 h of *E. mallotivora* post inoculation. The result also highlighted a very strong up-regulation of T3SS protein HrpQ as early as 6 h after *E. mallotivora* infection (sixfold detected *via* sqRT-PCR, ninefold detected *via* RNA-seq), thus suggesting the significant role of this T3SS protein family member to initiate the pathogenesis mechanism in *E. mallotivora*. Classified under Hrp1 subfamilies, HrpQ has mostly been studied in *Pseudomonas syringae* (Büttner and He 2009), which encoded for one of the injectisome proteins to deliver the effector proteins.

Conclusion

In this study high number of identified DEGs suggested that three key pathways are responsible for *E. mallotivora* pathogenesis. They are biosynthesis of secondary metabolites, microbial metabolism in diverse environments and ABC transporters. Meanwhile, functional annotation of DEGs showed a response to a stimulus, transmembrane transport, lyase activity and structural molecule activity were identified as among the first-line biological processes that occur during the early phase of *E. mallotivora* infection. The key effectors proteins secreted by *E. mallotivora* and the pathogenicity mechanism adopted by *E. mallotivora* as the infection progresses especially during early hours of infection were revealed from our study.

Acknowledgements

References

Anders S, PT Pyl, W Huber (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169

Anon (2006). *FAOSTAT. Agricultural production*. Crops primary (papaya). Available at: http://apps.fao.org/faostat/ (Accessed: 23 August 2018)

Antiabong JF, MG Ngoepe, AS Abechi (2016). Semi-quantitative digital analysis of polymerase chain reaction electrophoresis gel: Potential applications in low-income veterinary laboratories. *Vet World* 9:935–939

Bakar NA, B Barun, L Rozano, L Ahmad, MFRM Raih, AA Tarmizi (2017). Identification and validation of putative *Erwinia mallotivora* effectors *via* quantitative proteomics and real time analysis. *J Agric Food Technol* 7:10–21

Büttner D, SY He (2009). Type III protein secretion in plant pathogenic bacteria. *Plant Physiol* 150:1656–1664

Chan SN, NA Bakar, M Mahmood, CL Ho, NM Dzaki, NA Shaharuddin (2017). Identification and expression profiling of a novel Kunitz trypsin inhibitor (KTI) gene from turmeric, Curcuma longa, by real-time quantitative PCR (RT-qPCR). *Acta Physiol Plantarum* 39:1–12

Deng W, CLD Hoog, HB Yu, Y Li, MA Croxen, NA Thomas, JL Puente, LJ Foster, BB Finlay (2010). Comprehensive proteomic analysis of the type III secretome of *Citrobacter rodentium*. *J Biol Chem* 285:6790–6800

Deng WL, AH Rehm, AO Charkowski, CM Rojas, A Collmer (2003). *Pseudomonas syringae* exchangeable effector loci: sequence diversity in representative pathovars and virulence function in *P. syringae* pv. *Syringae* B728a. *J Bacteriol* 185:2592–2602

Djamei A, K Schipper, F Rabe, A Ghosh, V Vincon, J Kahnt, S Osorio, T Tohge, AR Fernie, I Feussner, K Feussner, P Meinicke, YD Stierhof, H Schwarz, B Macek, M Mann, R Kahmann (2011). Metabolic priming by a secreted fungal effector. *Nature* 478:395–398

Evans EA, FH Ballen (2012). *An Overview of Global Papaya Production, Trade, and Consumption. Food and Resource Economics Department, Institute of Food and Agricultural Sciences, University of Florida,* pp:1–7. Gainesville, Florida, USA

Fanny EH, AM Bruce, C Daniel (2018). Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Mol Ecol* 27:2725–2741

Ferreira RM, LM Moreira, JA Ferro, MRR Soares, ML Laia, AM Varani, JCFD Oliveira, MIT Ferro (2016). Unravelling potential virulence factor candidates in *Xanthomonas citri*. subspp. *citri* by secretome analysis. *Peer J* 4:1-29

Fullerton RA, L Taufa, JL Vanneste, J Yu, DA Cornish, D Park (2011). First record of bacterial crown rot of papaya (*Carica papaya*) caused by an *Erwinia papaya* like bacterium in the Kingdom of Tonga. *Plant Dis* 95:70

Gardan L, R Christen, W Achouk, P Prior (2004). *Erwinia papaya* sp. nov., a pathogen of papaya (*Carica papaya*). *Intl J Syst Evol Microbiol* 54:107–113

Helmann JD (2002). The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* 46:47–110

Kajihara S, H Hojo, M Koyanagi, M Tanaka, H Mizumoto, K Ohnishi, A Kiba, Y Hikichi (2012). Implication of hrpW in virulence of *Pseudomonas cichori*. *Plant Pathol* 61:355–363

Kazmierczak MJ, M Wiedmann, KJ Boor (2005). Alternative sigma factors and their roles in bacterial virulence. *Microbiol Mol Biol Rev* 69:527–543

Kunkel BN, Z Chen (2006). Virulence strategies of plant pathogenic bacteria. *In: The Prokaryotes*, *The Prokaryotes*, pp: 421–440. Dworkin M, S Falkow, E Rosenberg, KH Schleifer, E Stackberrandt (Eds.). Springer, New York, USA

Liu H, SJ Coulthurst, L Pritchard, PE Hedley, M Ravensdale, S Humphris, T Burr, G Takle, MB Brurberg, PRJ Birch, GPC Salmond, IK Toth (2008). Quorum sensing coordinates brute force ans stealth modes of infection in the plant pathogen *Pectobacterium atrosepticum*. *PLoS Pathol* 4; Article 1000093

Macalood JS, HJ Vicente, RD Boniao, JG Gorospe, EC Roa (2013). Chemical Analysis of *Carica Papaya* L Crude Latex. *Amer J Plant Sci* 4:1941–1948

McClure R, D Balasubramanian, Y Sun, M Bobrovskyy, P Sumby, CA Genco, CK Vanderpool, B Tjaden (2013). Computational analysis of bacterial RNA-Seq data. *Nucl Acids Res* 41: 140

Melotto M, BN Kunkel (2013). Virulence strategies of plant pathogenic bacteria. *In: The Prokaryotic,* pp:61–82. Rosenberg E, EF Delong, S Lory, S Stackebrandt, F Thompson (eds). Springer, Heidelberg, Germany

Milind P, G Gurditta (2011). Basketful benefits of papaya. *Intl Res J Pharm* 2:6–12

Mortazavi, A, BA Williams, K McCue, L Schaeffer, B Wold (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* 5:621–628

Piasecka A, N Jedrzejczak-Rey, P Bednarek (2015). Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytol* 206:948–964

Piqué N, D Miñana-Galbis, S Merino, JM Tomás (2015). Virulence Factors of *Erwinia amylovora*: A Review. *Intl J Mol Sci* 16:12836–12854

Prasannath K (2013). Pathogenicity and Virulence Factors of Phytobacteria. *Sch Acad J Biosci* 1:24–33

Redzuan RA, NA Bakar, L Rozano, R Badrun, NM Amin, MFM Raih (2014). Draft genome sequence of *Erwinia mallotivora* BT-MARDI, causative agent of papaya dieback disease. *Genome Announ* 2:1-1

Sagi M, R Fluhr (2006). Production of reactive oxygen species by plant NADPH oxidases. *Plant Physiol* 141:336–340

Schenk A, H Weingart, MS Ullrich (2008). Extraction of high-quality bacterial RNA from infected leaf tissue for bacterial in planta gene expression analysis by multiplexed fluorescent Northern hybridization. *Mol Plant Pathol* 9:227–235

Schneider CA, WS Rasband, KW Eliceiri (2012). NIH Image to Image J: 25 years of image analysis. *Nat Meth* 9:671–675

Toth IK, L Pritchard, PR Birch (2006). Comparative genomics reveals what makes an enterobacterial plant pathogen. *Annu Rev Phytopathol* 44:305–36

VanBuren R, R Ming (2014). Sequencing and assembly of the transgenic papaya genome. *In: Genetics and Genomics of Papaya,* pp: 187–203. Ming R, P Moore (Eds.). Springer Verlag, New York, USA