Identification of genes differentially expressed in avocado fruits (cv. fuerte) infected by C*olletotrichum gloeosporioides* using Roche 454 GS FLX Titanium Platform.

A. T. Djami-Tchatchou and C. J. Straker, University of the Witwatersrand, Johannesburg, South Africa.

Abstract

Plants respond to pathogen infection by activating a complex defence mechanism that acts both locally and systemically. Identification of genes differentially expressed in avocado during infection with *Colletotrichum gloeosporioides,* which is the aim of this study, represents an important step towards understanding disease mechanism in 'fuerte' avocado fruit. In this study 454 sequencing and analysis of the transcriptome of infected avocado fruits were performed using Roche 454 GS FLX Titanium platform to determine the effect of the infection on gene expression. cDNA libraries enriched for differentially expressed genes in avocados were constructed from unharvested and harvested avocado fruit tissues collected after 1, 4 and 24 hours post-infection then pooled together for the early response and after 3, 4, 5 and 7 days post-infection then pooled together for the late response. The pooled samples were sequenced with the uninfected samples. Over two hundred thousand 454 reads were produced and subjected to BLAST searches from which we generated many genes and their functions. To gain insight into this interaction we have identified a large number of avocado genes encoding proteins predicted to function in metabolism, signal transduction, transcriptional control, defence, stress, transportation processes and some genes with unknown functions. This study represents the first transcriptome analysis of avocado fruit following infection with *Colletotrichum gloeosporioides.*

Key words: *Persea americana*, Fuerte, RNA extraction, *Colletotrichum gloeosporioides*, gene expression, defence responses, 454 sequencing

Identificación de la genes expresados diferancialmente en frutos de aguacate fruits (cv. fuerte) infectados por Colletotrichum gloeosporioides utilizando **Roche 454 GS FLX Titanio Plata forma.**

A. T. Djami-Tchatchou and C. J. Straker, Universidad de Witwatersrand, Johannesburgo, Sudáfrica

Las plantas responden a la infección por patógenos mediante la activación de un mecanismo de defensa complejo que actúa tanto a nivel local y sistémica. Identificación de genes expresados diferencialmente en el aguacate durante la la infección con *Colletotrichum gloeosporioides*, que es el objectivo de este estudio reprensenta un paso importante para entender el el mecanismo de la enfermedad en frutos de aguacate Fuerte. En este estudio 454 secuenciación y análisis del transcriptoma de frutos de aguacate infectados se realizaron utilizando Roche 454 GS FLX Titanio Plataforma para determinar el efecto de la infección en la expresión génica. Bibliotecas de cDNA enriquecida de genes expresados diferencialmente en los aguacates se construye a partir de tejidos sin cosechar y cosechar frutos de aguacate recopiladas después del 1, 4 y 24 horas después de la infección se agruparon para la pronta respuesta y después de los días 3, 4,5 y 7 después de la infección después agruparon para la respuesta tardía. Las muestras fueron secuenciados con la agrupación de las muestras no infectadas Más de 200 mil 454 lecturas fueron producidas y sometidas a BLAST busca de la que hemos generado muchos genes funciones y sus funciones. Para comprender mejor esta interacción se han identificado un gran número de genes que codifican proteínas aguacates prevé que función en el metabolismo, la transducción de señales, control de la transcripción, la defensa, el estrés,transporte y procesado de algunos genes con función desconocida en respuesta a *Colletotrichum gloeosporioides.* Este estudio representa el primer análisis del transcriptoma de frutos de aguacate después de la infección con *Colletotrichum gloeosporioides*.

Palabras clave: *Persea americana*, 'fuerte', la extracción de RNA, *Colletotrichum gloeosporioides*, la expresión de genes, las respuestas de defensa, 454 de secuenciación

1. Introduction

Avocados (*Persea americana* Mill.), are one of the most important sources of human nutrition, providing food and income to million of people all over the world. Avocado production in South Africa is concentrated mainly in the warm subtropical areas of the Limpopo and Mpumalanga provinces in the north east of the

country (Donkin, 2007). Avocado like others crops is susceptible to attack by filamentous fungi of the genus *Colletotrichum*, pathogens which cause anthracnose disease. Anthracnose is the most severe post harvest disease of avocado fruit (Pernezny et al, 2000). It reduces avocado fruit shelf life and negatively affects fruit quality, taste, and marketability (Freeman et al, 1998; Pernezny et al, 2000).

The elucidation of gene expression profiles in the infected tissue may bring more understanding on how avocado fruits react at the molecular level to *C. gloeosporioides* and therefore contributes to the development of strategies to improve its production and storage away of this pathogen attack. Over the last few years, next-generation sequencing technologies have been used as powerful approaches for discovering new genes and analyzing gene expression profiles in plant tissues (Mardis, 2008).The sequencing of cDNA instead of genomic DNA focuses analysis on the transcribed portion of the genome which reduces the size of the sequencing target space. Many applications have been elucidated through transcriptome sequencing such as: gene expression profiling, genome annotation, and rearrangement detection to non coding RNA discovery and quantification (Morozova and Marra, 2008). Based on the high number of reads generated per run together with the low sequencing error rate in the contigs obtained, 454 sequencing is well adapted to sequence the transcriptome of both model and non-model plants (Barakat et al, 2009).

In the current study, we focused on comparing the transcriptomes generated from healthy avocado fruit and avocado fruits infected with *C. gloeosporioides* during pre and post harvest using Roche 454 GS FLX titanium platform. This comparison will enable us to identify a large number of differentially expressed candidate genes. The strategy undertaken here does not require prior sequence knowledge or genome reference, and relies exclusively on publicly available software and basic scripting tools. To the best of our knowledge, this study is the first exploration to discover differentially expressed candidate genes in avocado fruit following infection with *C. gloeosporioides.*

2. Materials and Methods

2.1 Preparation of fungal Inoculum

The strain of *C. gloeosporioides* used was isolated from an anthracnose lesion on a 'Fuerte' avocado fruit picked in an organic orchard in Mpmulanga province (Giovanelli, 2008). The fungus was cultured on 5% (w/v) Malt Extract Agar (Merck). Spore suspensions (conidia) of the fungus were prepared as described previously by Prusky et al (1990).

2.2 Plant materials and inoculation

The experiment was done on unharvested (inoculated and uninoculated) and harvested (inoculated and uninoculated) 'Fuerte' avocado fruits from Roodewal farm, Nelspruit, South Africa. For in *planta* inoculations, the fruits were wounded and inoculated with 100 µl of *C. gloeosporioides* spore suspension (1×10^7 spores per ml^{-1}) using a sterilized needle of 2 mm length and 1 mm thick. The inoculated fruits were covered with a clean plastic bag in which distilled water was sprayed to maintain humidity. For the post harvest inoculation, after the sterilization steps, the fruits were wounded and inoculated with 100 µl of *C. gloeosporioides* spore suspension (1×10^7 spores per ml^{-1}). Fruits were incubated at 25 °C in aluminium trays covered with aluminium foil and humidity was maintained by placing a Petri dish filled with sterile distilled water at the centre of each tray (Kwang-Hyung et al, 2004). The control fruits for both unharvested and harvested treatments were treated similarly to infected fruits and were inoculated with 100 µl of sterile distilled water. The harvested control fruit was placed in a separate tray to prevent the spread of infection.

2.3 The infection efficiency

The infection efficiency by *C. gloeosporioides* was observed by Scanning Electron Microscopy (SEM) of harvested and unharvested fruits using a method described by Palhano et al (2004), with some modifications.

2.4 Total RNA extraction

Total RNA was extracted from the mesocarp of avocado fruit (1 g) using the method described by Djami-Tchatchou and Straker (2011).

2.5. Synthesis of double-stranded cDNA from total RNA

Doubled strand cDNA was synthesized starting with the total RNA, using a cDNA Synthesis System Kit (Roche), following the manufacturer's instructions.

2.6. 454 library construction and sequencing

cDNA of the control and infected samples of each time point were used to construct a 454 library following the supplier's instructions (Roche Diagnostics). All the samples; Control, Early Unharvested, Late Unharvested, Early Harvested and Late Harvested (C, EU, LU, EH and LH respectively) were pooled for sequencing and, therefore, the individual samples had to be separated based on their individual tags.

2.7. Transcript assembly and analysis

The data from the 454 read sequences of each sample were assembled into contigs using the proprietary Roche 454 Newbler Assembler software. Reads from each library were assembled separately following a combined assembly of all data from all time points to yield contigs which correspond to transcripts.
The cDNA sequences were annotated using BLAST [Basic Local Alignment Search Tool (Altschul et al, 1990)]. Similarities at the nucleotide level were identified using BLASTN and Protein homologies were identified using the non-redundant protein databases BLASTX (Altschul et al, 1990). Each gene was classified into a functional category based on the putative function played by the gene product.

2.8. Reverse Transcription PCR

Reverse Transcription PCR was done to validate the results of the sequencing. Seven primers were designed using the Integrated DNA Technologies's PrimerQuest incorporates Primer3 software to test the expression of seven targets genes. The PCR reactions were carried out as follow: One cycle of 94 °C for 3 min (initial denaturation), 35 cycles of 94 °C for 30 s (denaturation), Tm of each primer for 1 min (annealing) and 72 °C for 2 min (elongation). Final elongation was achieved at 72 °C for 10 min.

3. Results

3.1 Inoculation of avocado fruits with *Colletotrichum gloeosporioides*

Symptoms of anthracnose developed 3 d after inoculation. These symptoms were characterized by black fruit rot and spots, as well as white mycelial growth developed on the wounded inoculated and adjacent uninoculated areas. Scanning electron micrographs of avocado fruit showed an increased number of spores which resulted from a successful hyphal colonization of the fruit surfaces in both harvested and unharvested fruits (Fig 1 B). At 4 and 7 day post infection in harvested fruit, severe tissue destruction and damage were observed with some hyphae protruding out of the lenticels, which resulted in cell wall and cuticle destruction (Fig 1 C and D). Fig 1 A shows an uninfected sample

3.2. Total RNA isolation and cDNA synthesis

High-quality RNA was obtained through the protocol described here. The relative yields of total RNA per gram of tissue ranged from 87.76 to 174.94 μg g^{-1} of fresh weight. The A260/280 ratios for isolated RNA varied from 2.09 to 2.15 and the A260/230 ratios from 2.06 to 2.18.

3.3 sequencing and *de novo* assembly

The single sequencing run produced 215 781 reads from avocado fruit transcriptome, with an average sequence length of 252-300 nucleotides (range = 41–562). A total of 70.6 megabases (MB) of sequence data were generated; more specifically, 11.4 MB of healthy transcriptome sequence, 11.5 MB of EU transcriptome sequence, 8.3 MB of LU transcriptome sequence, 23.9 MB of EH transcriptome sequence and 15.5 MB of LH of transcriptome sequence.

3.4. Sequences analysis

The comparison between the healthy and infected transcriptomes enabled us to identify a large number of candidate pathogen response genes. We first determined how many times a gene was represented in each of the libraries based on the number of reads for each unigene count. Putative functions of each of the genes were determined by comparing their sequence with other sequences present in public GenBank databases using BLASTN and BLASTX programme. Of the 709 sequences of the genes analysed 639 showed significant homology to previously known plant gene sequences and 70 had no significant homology to plant genes in the database. Of the 639 sequences exhibiting homology to plant genes, 358 had homology to senescence-associated proteins, 114 had homology to cytochrome (cytochrome P450 like_TBP protein and, cytochrome P450 monooxygenase, cytochrome c oxidase and cytochrome f), 102 had homology to a hypothetical protein, 18 showed similarities to proteins involved in metabolism, 14 showed homology to plant defence and stress-related proteins, 9 showed homology to transcription factor and cellular communication, 9 had homology to expressed protein, 8 had homology to proteins involved in photosynthesis and cell structure and 7 showed homology to proteins involved in electron transport.
Of the 639 cDNA sequences exhibiting differential expression in response to *C*. *gloeosporioides* infection, we selected some sequences to show their expression per time point (Table 1).

During the assembly process, there were reads that did not map from the uninfected and infected samples. These unmapped reads also represent genes and/or parts of the genes. These unassembled reads likely correspond to transcripts as well, but in very low copy number. The unmapped reads were also quantified in order to determine the number of copies expressed per time-point. The function assignment of these genes was done based on homology after comparing their sequences to the non-redundant protein databases BLASTX program. Then the unmapped reads from all the uninfected samples (EU, LU, EH and LH) were mapped to the unmapped /unassembled reads from the uninfected sample in order to quantify their expression. Table 2 shows the putative function and how many of the unmapped reads from EU, LU, EH and LH map to the unassembled reads from the control. During the assembly process there are other group of reads from the infected samples that did not map with the reads from the uninfected samples (data not shown). Genes obtained from these reads are considered to be induced after *C*. *gloeosporioides* infection

3.4. Reverse Transcription PCR

PCR products from avocado flesh cDNA amplified with defence-related gene primers produced single bands of between 100-200 bp, depending on the primer sets used confirming that the sequences obtained from the 454 sequencing are those of avocado fruit (Fig 2).

4. Discussion

The infection process by *C*. *gloeosporioides* was successful as seen on the SEM micrographs (Fig 1). In this study, we produced about 215 781 reads from avocado fruit transcriptome in a one-quarter run with the Roche 454 GS FLX Titanium platform. From the anthracnose symptoms exhibited by the fruits and the transcriptome sequencing, it is clear that numerous genes and the products of many of these genes are directly or indirectly involved in the interaction between avocado fruit and *C*. *gloeosporioides.* In addition the BLASTN revealed that many sequences obtained showed similarities at the nucleotide level (Altschul et al, 1990) with others species belonging to the order Laurales such as: *Cinnamomum camphora, Peumus boldus, Gyrocarpus americanus, Calycanthus occidentalis, Gomortega keule, e*tc. cDNA sequences generated from the uninfected and infected fruits cover various biological activities and molecular functions indicating that 454 sequencing constitutes a powerful tool for sequencing the transcriptome and gene discovery of non model species such as avocado. Many categories of genes, such as those involved in signal transduction, were differentially expressed in avocado fruits following *C*. *gloeosporioides* infection. MAPK are serine/threonine-specific protein kinases that participate in tranducing extracellular stimuli to the host genome and would be activated after C. *gloeosporioides* infection to enable pathogen recognition and to stimulate plant responses. It has been shown that MAPKs in several plant species are activated during plant responses to elicitors or pathogens (Zhang and Klessig, 2001). Since the defence signalling pathways that lead to the plant-pathogen response are activated after elicitor binding to receptors, it is more probably that LRR receptor-like protein kinase and signal recognition particle receptor protein were expressed in order to mediate the fungus recognition by the plant (Dangl

and Jones, 2001). In addition the expression of salicylic acid-binding protein may suggest that salicylic acid, which is involved in the activation of various plant defence responses following pathogen attack, plays an important signalling role following C. *gloeosporioides* infection in avocado (Hammond-Kosack and Parker, 2003). Calcium ion binding protein and calcium-dependent protein kinase were also expressed, indicating Ca^{2+} signalling activities in avocado following *C*. *gloeosporioides* infection. Previous studies revealed that the fluctuations in cytosolic Ca^{2+} levels that are mediated by Ca^{2+} permeable channels located at the plasma membrane of the plant cell can serve as a regulation of the plant response to pathogen invasion (White and Broadley, 2003). The other category of genes that were found to be differentially expressed following *C*. *gloeosporioides* infection encode proteins involved in lignin biosynthesis, such as cinnamoyl-CoA reductase, a key enzyme in lignin biosynthesis (Kawasaki et al *,* 2006) and cytochrome P450 monoxygenases (Barakat et al, 2009). It is well established that lignification is a mechanism for disease resistance in plants, which lead to an ultra-structurally modified reinforced cell wall (Bhuiyan et al, 2009). In many plants species, the transcription factor WRKY and *Myb* genes expressed in avocado are strongly and quickly up-regulated in response to pathogen attack, wounding or abiotic stresses. Studies demonstrated that in tobacco, multiple WRKY genes are induced after infection with bacteria or tobacco mosaic virus, or treatment with fungal elicitors SA or H_2O_2 (Takemoto et al, 2003). Another category of genes differentially expressed in response to *C*. *gloeosporioides* infection are genes involved in defence response such as β-glucanases, endochitinases and endopeptidase known to have antifungal activity (van Loon and van Strien, 1999) and plant aspartic proteinase which exhibit antimicrobial activity. Other defence genes expressed are known to be involved in various processes of plant defence against pathogens, such as cell death related to hypersensitivity response, construction of a physical barrier to block the pathogen progression, as well as systemic resistance. We found elongation factor-1 which is involved in controlling the extent of the cell death in the defence response and acetyl co-enzyme A carboxyltransferase involved in the regulation of resistance gene expression (Barakat et al, 2009). Another functional category observed in the avocado transcriptomes are genes encoding proteins predicted to function in photosynthesis and oxidation processes. Such genes may play a crucial role in energy production during response to the pathogen infection. For instance ATP synthase expressed in avocado fruit is an important enzyme that creates energy for the cell to use through the synthesis of adenosine triphosphate (Thilmony et al, 2006).

In conclusion, this study allowed us to characterize the genomic response of avocado fruit to anthracnose disease. The overall goal was to sequence the whole uninfected and infected avocado transcriptome, then to identify several candidate genes which are differentially expressed in avocado as a result of *C*. *gloeosporioides* infection*.* These findings yielded a first insight into some of the genes expressed in this plant-pathogen interaction at the molecular level and could contribute to the design of effective disease management strategies to improve the resistance of avocado varieties to anthracnose disease.

5. Acknowledgements

We acknowledge financial support from the National Research Foundation, South Africa (ICD2007052500003: 2008-2011), the University of the Witwatersrand Postgraduate Merit Award, the Bradlow Award and the Oppenheimer Memorial Trust**.**

6. References

Altschul S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215 (3): 403-410.

Barakat A., DiLoreto D. S., Zhang Y., Smith C., Baier K., Powell W. A., Wheeler N., Sederoff R., Carlson J. 2009. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BiodMed Central Plant Biology*.9:51

Bhuiyan N. H., Selvaraj G., Wei Y. & King J. 2009. Role of lignification in plant defense. *Plant Signaling & Behavior* 4(2): 158-159.

Dangl J.L & Jones J.D.G. 2001. Plant pathogens and integrated defence responses to infection. *Nature*. 411:826-833.

Djami-Tchatchou, A T Straker, CJ. 2011. The isolation of high quality RNA from avocado (*Persea americana* Mill.) fruit. South African Journal of Botany (in press).

Donkin, D, J 2007, An overview of the South African Avocado Industry, *South African Subtropical Growers' Association, SAAGA.*

Freeman, S, Katan, T & Shabi, E 1998, 'Characterization of *Colletotrichum* species responsible for anthracnose diseases of various fruits', *Plant disease*, 82, pp. 596-605.

Giovanelli, L,C 2008, 'The use of Ethanol Extract of Propolis in the control of *Colletotrichum gloeosporioides, Pestalotia guipinii* and *Pseudocercospora purpurea* infections of avocado (*Persea americana*)', MSc-Dissertation, University of the Witwatersrand, Johannesburg-South Africa.

Hammond-Kosack, K, E & Parker, J E 2003, 'Deciphering plant–pathogen communication: fresh perspectives for molecular resistance breeding', *current opinion in biotechnology*, 14, pp. 177-193.

Kawasaki, T, Koita, H, Nakatsubo, T, Hasegawa, K, Wakabayashi, K, Takahashi, H, Umemura, K, Umezawa, T & Shimamoto, K 2006, 'Connamoyl-CoA reductase, a key enzyme in lignin biosynthesis, is an effector of small GTPase Rac in defense signaling in rice', *Proceedings of the national academy of sciences*,*USA*, 103, pp. 230-235.

Kwang-Hyung, K, Jae-Bok, Y, Hyo-Guen, P, Eun, Woo, P & Youmg Ho, K 2004, 'Phytopathology', *The american phytopathological society*, Vol 94 (12) , pp. 1295-1304.

Mardis, E, R 2008, The impact of next-generation sequencing technology on genetics, *Trends in genetics*. 24(3), pp. 133-141.

Morozova, O& Marra, M, A 2008, Applications of next-generation sequencing technologies in functional genomics, *Genomics*, 92(5) , pp. 255-264.

Palhano, F,L, Vilches, T,T,B, Santos, R,B, Orlando, M,T,D, Ventura, J,A & Fernandes, P,M,B 2004, 'Inactivation of *Colletotrichum gloeosporioides* spores by high hydrostatic pressure combined with citral or lemongrass essential oil', *International Journal of food microbiology*, 95, pp. 61– 66.

Pernezny, K, Glade, B & Marlatt, R, B 2000, 'Diseases of Avocado in Florida', *Plant Pathology*, *Fact sheet* PP-21.

Prusky, D, Karni, L, Kobiler I & Plumbley, R,A 1990, 'Induction of the antifungal diene in unripe avocado fruits: Effect of inoculation with *Colletotrichum gloeosporioides'*, *Physiological molecular plant patholology* 37, pp. 425-435.

Takemoto, D, Yoshioka, H, Doke, N & Kawakita, K 2003, 'Disease stress inducible genes of tobacco: expression profile of elicitor responsive genes isolated by subtractive hybridization', *Physiologia plantarum*, 118, pp. 545-553.
Thilmony, R, Underwood, W & He, S,Y 2006, 'Genome-wide transcriptional analysis of the Arabidopsis thaliana interaction with the plant pathogen Pseudomonas syringae pv. tomato DC3000 and the human pathogen Escherichia coli O157:H7',The *plant journal*, 46, pp. 34-53.

Van Loon, L,C & Van Strien, E,A 1999, 'The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins', *Physiological molecular plant patholology*,55 , pp. 85-97.

Zhang, S & Klessig, D, F 2001, MAPK cascades in plant defense signaling,*Trends plant science*, 6 (11) , pp. 520-527.

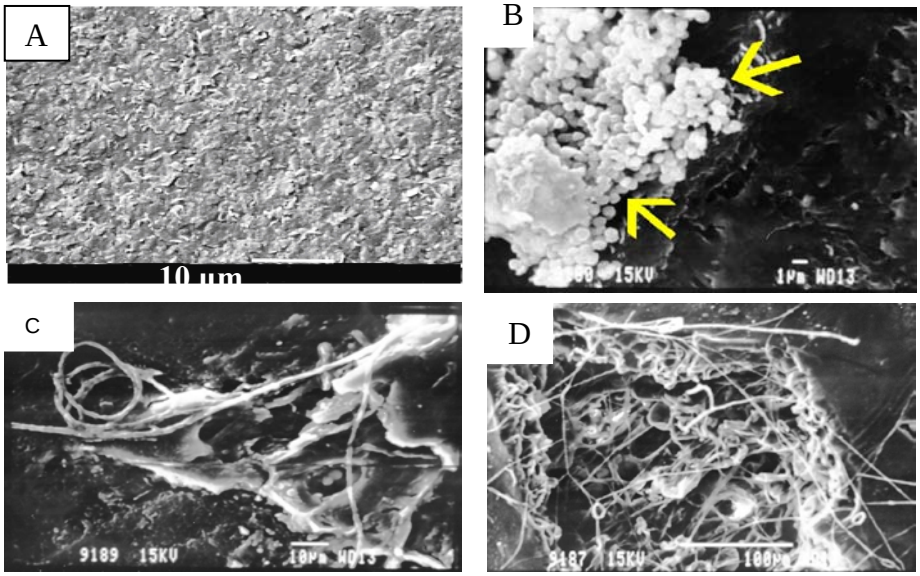

Fig1. Scanning electron micrographs of avocado fruit infected with *C.gloeosporioides*. (A) An uninfected sample. An increased number of *C. gloeosporioides* spores showed by the arrows on the surface of inoculated fruit (B) at 4 dpi. (C) and (D) showing severe tissue destruction and damage with some hyphae protruding out of the lenticels at 4 and 7 dpi.

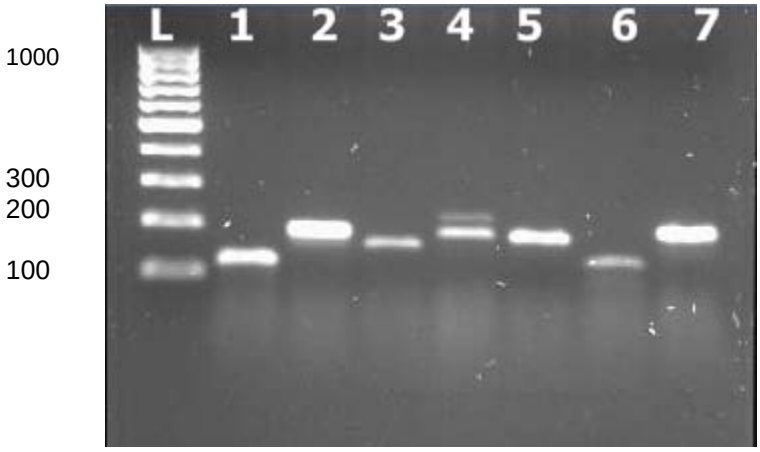

Fig 2. PCR products amplified from avocado flesh cDNA using specifics primers separated on 2% non-denaturing agarose gel containing EtBr and photographed under ultraviolet light. L, GeneRuler™ 100 bp DNA Ladder. Lane 1 catalase (162bp), lane 2 endochitinase (186), lane 3 pathogenesis related protein 6 (158), lane 4 CC-NBS resistance protein (185), lane 5 pathogenesis related protein 5 (171 bp), lane 6 Endo-1,4-D-glucanase (121) and lane 7 actin gene (166 bp).

Table 1. Summary of some selected candidate genes differentially expressed in avocado fruits in response to *C*. *gloeosporioides* infection with their different expression per time point (reads mapped to the individual transcripts for every time point). Control, Early Unharvested, Late Unharvested, Early Harvested and Late Harvested (C, EU, LU, EH and LH respectively).

Name	**BLASTn** (Accession num)	**Max Identity** %	**BLASTx** (Accession num)	**E-value**	**Putative function**	**Copies per time-point expressed**				
						Cont	**EU**	**LU**	**EH**	**LH**
gene00001	Cinnamomum camphora large subunit ribosomal gene (DQ008772.1|)	99	leucine rich protein [Arachis hypogaea] (ABH09320.1)	4e-04	Signal transduction	107	145	41	88	83
gene00082	Peumus boldus 26S ribosomal RNA gene (AY095466.1)	96	hypothetical protein [Arabidopsis thaliana] (BAF01964.1)	2e-48		126	207	178	262	167
gene00086	Gyrocarpus americanus 26S ribosomal RNA gene, (DQ008624.1)	100	unknown [Zea mays] (ACR36970.1)	1e-45		68	165	127	139	91
gene00091	Calycanthus occidentalis 26S ribosomal RNA gene (AY095454.1)	95	putative senescence-associated protein [Trichosanthes dioica] (ABN50032.1)	6e-30		135	187	184	349	236
gene00132	Gomortega keule 26S ribosomal RNA gene (AY095460.1)	98	cytochrome P450 monooxygenase [Pyrus communis] (AAR25996.1)	2e-06	Lignification	12	40	20	49	30
gene00144	Cryptocarya meissneriana 26S ribosomal RNA gene (DQ008627.1)	98	Cytochrome c oxidase subunit 5B, mitochondrial precursor (ACN10266.1)	7e-05	Oxidation	8	19	9	22	6
gene00665	Liriodendron tulipifera chloroplast, complete genome (DQ899947.1)	95	photosystem I assembly protein Ycf4 [Zea mays] (NP_043035.1)	1e-18	photosynthesis	1	2	2	0	0
gene00609	Daphnandra micrantha 26S ribosomal RNA gene |DQ008629.1)	96	transcription factor WRKY36 [Physcomitrella patens subsp. Patens (XP_001775684.1)	8.7	Transcription factor	7	7	4	9	3
gene00237	Arabidopsis thaliana clone 34690 mRNA, (AY087376.1)	72	fructose-bisphosphate aldolase [Persea americana] (CAB77243.2)	2e-81	Metabolism	7	4	1	13	9
gene00307	Persea americana stearoyl-acyl-carrier-protein desaturase mRNA, (|AF116861.1)	99	stearoyl-acyl-carrier-protein desaturase [Persea americana]	9e-90	Transportation	4	1	2	2	1
gene00308	Laurus nobilis 26S ribosomal RNA gene (DQ008626.1)	100	heat shock protein [Cucumis sativus] (ADF30255.1)	7e-79	Stress related protein	2	3	1	2	3

Table 2 The number of unmapped reads from EU, EH, LU and LH samples mapping to the unmapped reads from the control and their putative function. Control, Early Unharvested, Late Unharvested, Early Harvested and Late Harvested (C, EU, LU, EH and LH respectively).

Name	**BLASTx** (Accession num)	**Similar Sequence From database (Putative function)**	**E-value**	**Max Identity** %	number of unmapped reads			
					EU	**LU**	**EH**	**LH**
DEFENCE								
GSKRSVY03F2JKP	ACG44564.1	endopeptidase Clp [Zea mays]	4 e-28	52	0	0	0	1
GSKRSVY03GNNXL	CAB01591.1	endochitinase [Persea americana	2 e-58	77	3	0	0	2
GSKRSVY03G5UE1	AAD30292.1	catalase 3 [Raphanus sativus]	5 e-57	76	0	2	2	3
GSKRSVY03GCUDV	NP_172655.1	aspartic proteinase A1 [Arabidopsis thaliana]	4 e-28	77	0	0	0	1
GSKRSVY03GX18B	ABX79341.1	cysteine protease [Vitis vinifera]	4 e-46	81	1	0	0	0
GSKRSVY03FYVXX	ABY58189.1	endo-1,4-D-glucanase [Persea americana]	6 e-48	100	0	0	0	3
GSKRSVY03HB7OO	ADQ39593.1	class II chitinase [Malus x domestica]	4 e-06	82	3	0	0	4
GSKRSVY03F3ZMX	AAK15049.1	asparaginyl endopeptidase [Vigna radiata]	2 e-20	66	0	1	0	0
GSKRSVY03GNVEE	ABK78689.1	cysteine proteinase inhibitor [Brassica rapa]	2 e-15	75	0	0	2	2
GSKRSVY03FTET8	XP_002527223.1	oligopeptidase A, putative [Ricinus communis]	5 e-09	86	0	0	1	0
TRANSCRIPTION FACTOR								
GSKRSVY03FZACC	AEF30544.1	ethylene transcription factor [Castanea sativa]	6 e-06	53	1	0	0	0
GSKRSVY03F03Q1	NP_850583.1	ethylene-responsive transcription factor RAP2-2 [Arabidopsis thaliana]	3 e-16	69	0	0	1	0
GSKRSVY03FVBLZ	AAS10005.1	MYB transcription factor [Arabidopsis thaliana]	2 e-07	83	0	0	1	0
GSKRSVY03GSUTC	CAD56217.1	transcription factor EREBP-like protein [Cicer arietinum]	0.069	100	1	1	1	0
METABOLISM / PHOTOSYNTHESIS								
GSKRSVY03F9QFM	YP_004021302.1	ATP synthase CF1 alpha subunit [Theobroma cacao]	6 e-90	99	0	0	2	0
GSKRSVY03F5V9Q	XP_002532986.1	Flavonol synthase/flavanone 3-hydroxylase, putative [Ricinus communis]	1. e-04	65	1	0	0	0
GSKRSVY03GZDGL	ABI18045.1	acetyl co-A carboxylase [Strombosia grandifolia]	4 e-10	81	3	2	0	0
GSKRSVY03GMBAO	CAP12013.1	photosystem II protein Z [Coffea myrtifolia]	2 e-06	94	2	0	0	0
PROTEIN SYNTHESIS								
GSKRSVY03GUM9Q	XP_002528028.1	elongation factor 1-alpha, putative [Ricinus communis]	9 e-51	100	1	1	2	1
GSKRSVY03GX76M	AAZ75913.1	ribosomal protein L16 [Coffea humilis]	0.014	100	1	1	0	0
GSKRSVY03GFXL2	BAD83474.2	ribosomal protein S3 [Nicotiana tabacum]	0.001	100	0	0	1	0

Table 2 (continue)

Name	BLASTx (Accession num)	Similar Sequence From database (Putative function)	E-value	Max Identity %	number of unmapped reads			
					EU	LU	EH	LH
SIGNAL TRANSDUCTION								
GSKRSVY03F8UTL	AAR87711.1	salicylic acid-binding protein 2 [Nicotiana tabacum]	1 e-10	63	0	0	0	1
GSKRSVY03G6CLQ	NP_196670.1	serine-rich protein-like protein [Arabidopsis thaliana]	7 e-11	60	0	0	0	1
GSKRSVY03G40FC	ADN96595.1	thioredoxin h [Vitis vinifera]	3 e-08	64	1	0	1	2
GSKRSVY03GBQJL	NP_201509.1	mitogen-activated protein kinase kinase kinase 19 [Arabidopsis thaliana]	7 e-21	60	1	0	0	0
GSKRSVY03HAFMN	XP_002880383.1	kinase family protein [Arabidopsis lyrata subsp. lyrata]	0.094	73	0	0	1	1
OXIDATIVE BURST								
GSKRSVY03GTJHX	ACG39782.1	NADH-cytochrome b5 reductase [Zea mays]	9 e-35	90	0	0	0	1
GSKRSVY03F44AG	YP_784442.1	NADH-plastoquinone oxidoreductase subunit 1 [Drimys granadensis]	6 e-23	100	0	0	1	0
GSKRSVY03GYLF9	ACO37154.1	ACC oxidase [Stenocereus stellatus]	4 e-21	88	0	0	1	0
GSKRSVY03FT4FZ	BAD83480.2	NADH dehydrogenase subunit 4 [Nicotiana tabacum]	2 e-16	100	1	0	0	0
GSKRSVY03G8YH8	AAF61392.1	glutathione S-transferase [Persea americana]	8 e-43	73	2	0	2	0
STRESS RESPONSE								
GSKRSVY03GOQET	NP_191404.2	universal stress protein (USP) family protein [Arabidopsis thaliana]	8 e-29	86	1	0	3	6
GSKRSVY03GEF32	ADF30255.1	heat shock protein [Cucumis sativus]	1 e-20	58	1	0	1	0
GSKRSVY03FK7ZJ	AAL49788.1	putative heat shock protein 90 [Arabidopsis thaliana]	4 e-13	83	0	0	2	1
TRANSPORTATION								
GSKRSVY03HHL3Q	AAF15308.1	stearoyl-acyl-carrier-protein desaturase [Persea americana]	9 e-20	100	0	0	1	1
GSKRSVY03FW9PQ	XP_002526521.1	Peptide transporter, putative [Ricinus communis]	2 e-15	62	0	0	0	1
GSKRSVY03GIDYK	CAB41144.1	H+-transporting ATPase-like protein [Arabidopsis thaliana]	3 e-23	81	0	0	0	1
FUNCTION: UNCLASSIFIED								
GSKRSVY03GYEPY	ACU24411.1	Unknown protein product [Glycine max]	3 e-06	92	0	5	0	8
GSKRSVY03FX7MT	XP_002888340.1	predicted protein [Arabidopsis lyrata subsp. lyrata]	8.6	73	0	0	3	0
GSKRSVY03F4G1J	NP_001077933.1	ubiquitin fusion degradation 1 [Arabidopsis thaliana]	9.1	93	1	0	0	1