

---

**TRANSCRIPTOME CHARACTERIZATION AND  
TRANSCRIPTIONAL REPROGRAMMING OF ZUCCHINI  
PLANT IN RESPONSE TO *APHIS GOSSYPPII* FEEDING**

---

**Alessia Vitiello**

Dottorato in Biotecnologie – XXX ciclo

Università di Napoli Federico II





Dottorato in Biotecnologie – XXX ciclo

Università di Napoli Federico II



---

**TRANSCRIPTOME CHARACTERIZATION AND  
TRANSCRIPTIONAL REPROGRAMMING OF ZUCCHINI  
PLANT IN RESPONSE TO *APHIS GOSSYPYII* FEEDING**

---

**Alessia Vitiello**

Dottorando: Alessia Vitiello

Relatore: Prof. Rosa Rao

Correlatore: Dott. Nunzio D'Agostino

Coordinatore: Prof. Giovanni Sannia



*Alla mia famiglia*



## INDEX

<b>RIASSUNTO</b>	1
<b>SUMMARY</b>	6
<b>CHAPTER 1-General Introduction</b>	8
<b>CHAPTER 2-De novo transcriptome assembly of the zucchini variety “San Pasquale”</b>	
<b>Abstract</b>	20
<b>2.1 Introduction</b>	20
<b>2.2 Results</b>	22
<b>2.3 Discussion</b>	28
<b>2.4 Conclusions</b>	30
<b>2.5 Materials and Methods</b>	30
<b>CHAPTER 3-Identification of zucchini (<i>Cucurbita pepo</i> L.) differentially expressed genes in response to <i>Aphis gossypii</i> feeding</b>	
<b>Abstract</b>	34
<b>3.1 Introduction</b>	34
<b>3.2 Results</b>	36
<b>3.3 Discussion</b>	51
<b>3.4 Conclusions</b>	56
<b>3.5 Materials and Methods</b>	56
<b>CHAPTER 4-Identification of semiochemicals released by “San Pasquale” leaves upon infestation by <i>Aphis gossypii</i></b>	
<b>Abstract</b>	60
<b>4.1 Introduction</b>	60
<b>4.2 Results</b>	62
<b>4.3 Discussion</b>	66
<b>4.4 Conclusions</b>	68
<b>4.5 Materials and Methods</b>	69
<b>CHAPTER 5-General Conclusions</b>	73
<b>REFERENCES</b>	77
<b>APPENDIX</b>	91
<b>STAGE</b>	117
<b>PUBLICATIONS</b>	118





## RIASSUNTO

Gli insetti dotati di apparato boccale puntente-succhiatore, tra i quali gli afidi, sono tra i principali agenti di danno per le colture. Essi, infatti, grazie alla loro modalità di alimentazione, sottraggono linfa elaborata e fotosintati alla pianta ospite, modificandone in maniera profonda la fisiologia. Gli afidi producono un ridotto danno meccanico in confronto agli erbivori con apparato boccale masticatore. Inoltre, la prolungata interazione che gli afidi stabiliscono con i tessuti della pianta si traduce nell'attivazione di processi di difesa in gran parte diversi da quelli attivati in risposta agli insetti masticatori. Attualmente, il controllo degli afidi si basa principalmente sull'uso di pesticidi che si traduce in un forte impatto negativo sull'ambiente, incluse le popolazioni microbiche del suolo e gli insetti benefici. Inoltre, biopesticidi di origine naturale ampiamente diffusi, come quelli costituiti da spore e molecole prodotte da *Bacillus thuringiensis*, sono disponibili per il controllo di insetti masticatori come lepidotteri e coleotteri, ma non per il controllo degli afidi. Per questi motivi è importante potenziare le difese endogene delle piante (identificando geni in grado di conferire resistenza o tolleranza) e sviluppare strategie di controllo integrato per la difesa delle colture (Le Mire *et al.*, 2016). Lo zucchini (*Cucurbita pepo* L.) appartiene alla famiglia delle *Cucurbitaceae*, alla quale appartengono anche cetriolo, melone e anguria, e alcune delle più antiche specie vegetali domestiche. Lo zucchini è tra le specie vegetali di maggiore importanza economica ed è coltivato nelle regioni temperate e subtropicali (Paris, 2008). L'Italia è l'ottavo paese produttore a livello mondiale ed il primo dell'area mediterranea, con più di 566.000 tonnellate prodotte nel 2014 (FAOSTAT 2014).

L'afide *Aphis gossypii* Glover è un fitofago polifago in grado di alimentarsi su diverse specie vegetali, tra le quali cotone e diverse cucurbitacee. Questo insetto è tra i principali agenti di danno nella coltivazione dello zucchini sia in pieno campo sia in ambiente protetto, ed è in grado di provocare danni diretti alle piante, come accartocciamento fogliare, avvizzimento e riduzione della crescita. In condizioni di gravi infestazioni possono comparire aree clorotiche e necrosi su foglie e frutti che provocano una forte riduzione della resa. Inoltre, *A. gossypii* è vettore di numerosi virus, come ad esempio lo ZYMV (Zucchini Yellow Mosaic Virus) che è tra i più dannosi per lo zucchini. I frutti prodotti da piante infette da ZYMV sono generalmente malformati, producono pochi semi e sviluppano alterazioni del colore rendendo il prodotto non commercializzabile (Zellnig *et al.*, 2014).

Ad oggi non sono stati ancora pubblicati studi relativi alle modifiche trascrizionali in zucchini attivate in risposta all'attacco di afidi e ai meccanismi molecolari associati al danno. Le nuove tecnologie di sequenziamento (NGS) rappresentano un importante strumento per studiare nel dettaglio i meccanismi molecolari che regolano l'interazione pianta-afide. In modo particolare, l'uso dell'RNA-seq offre il vantaggio di ottenere una visione globale delle modifiche trascrizionali che si verificano in seguito all'infestazione afidica. La strategia RNA-seq risulta molto efficiente per lo studio di trascrittomi di specie vegetali per le quali le risorse genomiche a disposizione sono scarse in confronto alle specie vegetali modello.

In questo scenario, il lavoro qui presentato si pone come principale obiettivo quello di indagare le modifiche trascrizionali in piante di zucchini in seguito all'attacco di afidi mediante approccio RNA-seq e di identificare le vie metaboliche attivate dall'infestazione e legate alla difesa diretta e indiretta della pianta.

Il trascrittoma utilizzato come riferimento in questo studio è stato assemblato *de novo* a partire dai tessuti fogliari infetti e non, ed è stato usato per l'identificazione dei geni la cui espressione è risultata alterata dalla presenza degli afidi.

Piante di zucchini della varietà campana “San Pasquale”, suscettibile ad *A. gossypii*, sono state infestate con 10 afidi adulti dopo tre settimane dalla semina. Per monitorare cambiamenti nell’espressione genica, le foglie sono state raccolte dopo 24, 48 e 96 ore dall’infestazione e congelate immediatamente in azoto liquido, rimuovendo prima gli afidi presenti. Foglie corrispondenti sono state prelevate da piante controllo (non infestate) allevate nelle stesse condizioni ambientali di quelle attaccate. Dai tessuti fogliari raccolti è stato estratto l’RNA totale, che è stato successivamente sequenziato utilizzando la piattaforma Illumina HiSeq 2500. L’esperimento di sequenziamento ha prodotto ~34 milioni di *paired-end read* di 101 nucleotidi per ciascun campione. Le *read* ottenute sono state ripulite dagli adattatori e filtrate per eliminare le sequenze corte (< 75 nt) e quelle con un valore di qualità inferiore alla soglia stabilita (*Q score* ≤ 30). Le *read* di alta qualità ottenute sono state utilizzate per condurre l’assemblaggio *de novo* del trascrittoma di zucchini. Il programma di assemblaggio selezionato (Velvet/Oases, Schulz *et al.*, 2012) ha consentito di ricostruire 122.507 *contig* che sono poi stati “collassati” per ridurre la ridondanza e gli eventuali errori di assemblaggio utilizzando il software CAP3 (Huang and Madan, 1999). È stato quindi ottenuto un trascrittoma costituito da 71.648 sequenze, con una lunghezza media di 1.331 nucleotidi. Circa il 94% dei trascritti assemblati è stato classificato come ipotetica sequenza codificante, in quanto presentava elementi tipici di una ORF. Questo risultato è un importante indice della qualità dell’assemblaggio ottenuto. Inoltre, i trascritti sono stati annotati al fine di attribuire una funzione biologica al maggior numero di essi. Analisi BLAST sono state condotte confrontando il trascrittoma assemblato con le banche dati di sequenze proteiche di *C. sativus*, *C. melo*, *Arabidopsis* e con la banca dati UniProt. Mediamente il 70% delle sequenze assemblate ha trovato almeno una corrispondenza all’interno dei quattro database interrogati. Il programma Blast2GO (Götz *et al.*, 2008) ha poi permesso di associare almeno un termine “*gene ontology*” (GO) a 51.398 sequenze in modo da poter descrivere le funzioni di tali trascritti utilizzando un vocabolario univoco. In totale sono stati associati 276.601 termini GO, di questi il 50% è stato assegnato al dominio funzionale “processo biologico”, il 27% al dominio “funzione molecolare” ed il 21% al “compartimento cellulare”. È stato inoltre confrontato il trascrittoma assemblato con quello disponibile in rete (pubblicato da Blanca *et al.*, 2011). Il risultato dell’analisi BLASTn ha evidenziato la presenza di 1.313 nuovi trascritti le cui funzioni ricadono anche in processi metabolici noti per essere attivati in risposta al danno biotico (insetti e/o patogeni). Quest’ulteriore dato conferisce un valore aggiunto alla risorsa trascrittomica generata, rafforzando anche la scelta di costruire un proprio riferimento per consentire di descrivere al meglio la risposta molecolare della pianta all’attacco afidico.

I cambiamenti nell’espressione genica durante le prime fasi dell’interazione compatibile tra zucchini “San Pasquale” e *A. gossypii* sono stati analizzati nei tre diversi tempi di infestazione. A 24, 48 e 96 h dall’attacco di *A. gossypii*, 766 geni sono stati identificati come differenzialmente espressi (DEG) e quindi influenzati dall’afide. In particolare, dopo 24 h di infestazione sono stati identificati 158 trascritti (115 up- e 43 down-regolati). A 48 h il numero di DEG è aumentato a 565 (420 up- e 145 down-regolati), mentre a 96 h dall’infestazione è stata osservata una variazione nell’espressione di 179 sequenze (62 up- e 117 down-regolati). Quindi, il numero di geni coinvolti nella risposta della pianta all’attacco dell’afide raggiunge il suo massimo a 48 h, implicando l’attivazione di una risposta dinamica e crescente della pianta nelle prime fasi di attacco per poi assistere ad una attenuazione della risposta nell’ultimo punto temporale.

Nelle condizioni sperimentali saggiate, durante le prime fasi dell'interazione, la pianta di zuccino percepisce la presenza dell'afide e attiva geni coinvolti in vie di segnalazione mediate da incrementi di calcio citosolico, e geni che codificano per enzimi associati alla detossificazione delle specie reattive dell'ossigeno (ROS). Il metabolismo primario, in particolare il metabolismo proteico e la fotosintesi, risultano attivati a differenza di quanto riportato in diversi studi di interazione condotti su altre specie vegetali. Questa risposta potrebbe essere direttamente legata alla presenza di effettori afidici che dirigono il metabolismo della pianta ospite alle esigenze nutrizionali degli afidi.

Inoltre, la sintesi di acido salicilico (SA) è attivata grazie alla sovra-espressione del gene *ICS1* coinvolto nel pathway biosintetico di tale ormone. Per quanto riguarda, invece, un altro ormone chiave nella risposta a insetti, l'acido jasmonico (JA), non è stato identificato tra i DEG nessun trascritto associato alla sua biosintesi. Invece, geni codificanti per inibitori di proteasi, che sono generalmente attivati in risposta alla sintesi di JA, sono risultati sotto-espressi.

Anche il metabolismo secondario è influenzato già a 24 h. Un gene codificante per l'enzima 4CL, che catalizza l'ultima reazione del pathway generale dei fenilpropanoidi che regola la produzione di molti composti tra cui lignina e flavonoidi, è risultato sovra-espresso.

A 48 h è stata osservata l'induzione di geni dipendenti dall'acido salicilico e correlati alla patogenesi (PR), mentre geni dipendenti dal JA quali inibitori delle serine proteinasi sono sempre down-regolati. Inoltre, un trascritto annotato come *NIMIN1* è stato identificato tra i geni sotto-espressi. Questo gene codifica per un regolatore negativo di NPR1/NIM1 che svolge un ruolo fondamentale nel regolare la risposta sistemica acquisita (SAR) e nella via di segnalazione mediata dall'acido salicilico (Weigel *et al.*, 2001). Tali modifiche trascrizionali potrebbero essere legate al tentativo della pianta di attivare risposte di difesa mediate dall'acido salicilico, e queste, secondo un meccanismo da chiarire, potrebbero antagonizzare la via di risposta legata all'acido jasmonico. Ad avvalorare questa ipotesi è anche il risultato di biosaggi effettuati sul comportamento degli afidi quando posti ad alimentarsi su piante di zuccino già precedentemente infestate, oppure pre-trattate con methyl salicilato (MeSA). Gli afidi posti su foglie di zuccino pre-infestate hanno mostrato un'alterazione nel comportamento. Infatti, meno del 50% degli afidi saggiati è stato ritrovato sulla foglia pre-infestata sulla quale erano stati posti, tendendo a spostarsi per cercare un nuovo sito di alimentazione. Lo stesso comportamento è stato registrato per afidi posti su piante pre-trattate con MeSA. Confrontando il numero di afidi che si arrampicavano sullo stelo di piante non trattate per alimentarsi (80%) con quello ritrovato sulle piante pre-trattate (circa il 47%), si può affermare che il MeSA ha esercitato un ruolo importante nel condizionare il comportamento degli afidi. Tale risultato è comparabile con quanto osservato su piante pre-infestate per le quali vi è stata attivazione della risposta di difesa.

A 48 h dall'infestazione il metabolismo primario risulta sempre attivato, ed in particolar modo la sintesi di proteine ribosomiali (più di 80 geni), sia citosoliche sia plastidiali, risulta fortemente up-regolata. Dopo 48 h risultano anche fortemente up-regolati diversi geni coinvolti nella sintesi e modifica di componenti della parete cellulare, che rappresentano un meccanismo di difesa diretto contro i fitomizi. A 96 h la prevalenza di geni down-regolati, insieme con la riduzione, rispetto alle 48 h, del numero di DEG, può essere legata ad un meccanismo di adattamento alla presenza dell'agente di danno e al progredire dell'infestazione. Infatti, i processi di detossificazione dei ROS risultano down-regolati e anche la via di segnalazione del

calcio è influenzata negativamente. Il numero di geni legati al metabolismo proteico è fortemente ridotto in confronto alle 48 h, anche se ad essere down-regolati sono geni codificanti per proteasi (*aspartic proteinase*). Anche i geni coinvolti nella modifica della parte cellulare sono tutti down-regolati durante l'ultimo tempo di analisi, e codificano sia per classi di enzimi coinvolti nella degradazione, ad esempio poligalatturonasi, sia per enzimi coinvolti nel rafforzamento della parete cellulare (AGPs). Tra i geni fortemente down-regolati coinvolti nel metabolismo secondario è stata anche annotata una *Terpene synthase*. Enzimi appartenenti a questa classe sono attivi nel pathway dei terpenoidi, e quindi nella sintesi di una classe composti organici volatili (VOC), i quali svolgono un ruolo importante nelle risposte indirette delle piante, richiamando i nemici naturali dell'insetto fitofago. Allo scopo di associare alla descrizione dei geni influenzati da *A. gossypii* in zuccino "San Pasquale" anche quella relativa alle molecole coinvolte nelle risposte indirette, i volatili emessi in seguito all'attacco sono stati raccolti in esperimenti di "air entrainment". Le analisi GC e GC/MS hanno rivelato la produzione di un numero ristretto di molecole volatili. Quando le piante sono state infestate con 10 afidi adulti per 48 h è stata osservata una significativa riduzione dei livelli di emissione di (*E*)-caryophyllene, ma non sono state trovate differenze significative per gli altri volatili. Inoltre, anche analizzando i volatili emessi da piante infestate con 300 afidi non sono state identificate molecole emesse in maniera differenziale rispetto alle piante controllo. L'alta densità di infestazione ha però influenzato in modo significativo l'incremento dell'emissione di (*E*)-caryophyllene a partire da 96 h dopo l'inizio dell'infestazione. Il (*E*)-caryophyllene è un metabolita appartenente alla classe dei sesquiterpeni ed è coinvolto nei processi di comunicazione delle piante con l'ambiente esterno. In particolare, numerosi studi hanno dimostrato il suo coinvolgimento in pomodoro e mais, anche in associazione ad altri volatili come il 6-methyl-5-hepten-2-one, nell'attrazione di nemici naturali degli afidi (Sasso *et al.*, 2007; Köllner *et al.*, 2008). Sono stati inoltre eseguiti biosaggi per valutare il comportamento in risposta all'(*E*)-caryophyllene sintetico di *A. gossypii* e di un suo nemico naturale, il parassitoide *Ahidius colemani*, molto utilizzato in programmi di lotta integrata. Dai saggi condotti in olfattometro a 4-vie è emerso che il parassitoide trascorre la maggior parte del tempo nella sezione relativa al (*E*)-caryophyllene rispetto alle sezioni controllo, in cui è posto il solo solvente. Pertanto il parassitoide risulta attratto in maniera significativa dal composto testato. Emerge quindi un ruolo chiave per questo volatile nell'interazione pianta-afide. Risulterà importante andare a chiarire quale meccanismo mediato dagli afidi regola la soppressione della emissione del (*E*)-caryophyllene in piante infestate con un basso numero afidi, ma che non può essere mantenuta in caso di elevata densità di infestazione.

Infine, dall'analisi trascrizionale è emerso un dato decisamente interessante ed al tempo stesso atipico, in quanto, a nostra conoscenza, non è riportato in nessuno studio di interazione pianta-afide pubblicato fino ad oggi. Fin dal primo punto temporale analizzato, sono stati identificati tra i DEG degli mRNA di origine afidica, il cui livello di espressione è poi aumentato nel tempo. Riteniamo che tali trascritti siano stati iniettati nei tessuti vegetali durante il processo di alimentazione degli afidi e siano stati prodotti dalle cellule delle ghiandole salivari, come confermato dalla identità di sequenza riscontrata in seguito al confronto con il trascrittoma delle ghiandole salivari di *A. gossypii* (Pennacchio *et al.*, unpublished).

Allo stato attuale delle conoscenze non è chiaro quale sia il ruolo di questi mRNA di origine afidica all'interno dei tessuti vegetali, ma un'ipotesi potrebbe essere quella che l'afide tenta di regolare il metabolismo e la risposta di difesa della pianta ospite a

proprio vantaggio anche attivando meccanismi di soppressione di alcune vie metaboliche chiave nella risposta a stress.

In conclusione, in questo studio è stata indagata, per la prima volta, la risposta molecolare della pianta di zuccino all'attacco afidico, mostrando quanto questa sia complessa e regolata da numerosi fattori.

Le informazioni e le conoscenze prodotte nella presente tesi rappresentano un fondamentale punto di partenza per sviluppare efficaci strategie per il controllo di *Aphis gossypii*, anche attraverso l'uso di tecniche innovative di breeding, quali il "genome-editing", per ottenere nuove varietà con un'augmentata capacità di difesa.



## SUMMARY

Zucchini (*Cucurbita pepo* L.) belongs to the *Cucurbitaceae* family and ranks among the highest-valued vegetables worldwide. It is widely cultivated in temperate region where one of the main problem related to its cultivation is the damage imposed by the cotton/melon aphid *Aphis gossypii* (Homoptera: Aphididae). *Aphis gossypii* is a polyphagous aphid which can both directly and indirectly affects host plant by inducing leaf curling and necrosis and vectoring several plant viruses. Plants can defend themselves against aphids using different strategies. A direct response activates a deep transcriptional reprogramming which leads to *de novo* synthesis of proteins and molecules implicated in defence, including production of herbivore induced volatile organic compounds (HI-VOCs), which can indirectly attract herbivore natural enemies.

In the present study, the mRNA from un-infested (control) and infested leaves by *A. gossypii* of *C. pepo* cultivar “San Pasquale” was sequenced to obtain a *de novo* transcriptome assembly to be used as reference for gene expression profiling.

Leaf material was collected from control and infested (10 adult *A. gossypii*) plants at three different time points (24, 48, 96 hours post infection; hpi) and RNA was extracted. Illumina sequencing generated ~34 million of paired-end reads of 101 nucleotides in length per sample. Short reads were pre-processed and, then, *de novo* assembled using Velvet/Oases and CAP3 tools into a non-redundant set of 71,648 transcripts. Approximately 94% of the assembled transcripts contains coding sequences that could be translated into proteins, and ~70% of transcripts was successfully annotated using BLAST similarity-based searches and Blast2GO. Furthermore, BLASTn comparisons with the publically available *C. pepo* transcriptome resulted in 1,313 transcripts exclusively assembled in the aphid-challenged transcriptome.

Following transcriptome assembly, a dataset of 42,517 sequences, in which each gene *locus* was represented only once, was filtered out and used as reference for read mapping and differentially expressed gene (DEG) call. A total of 766 transcripts was differentially expressed (FDR < 0.05;  $-2 < \log_{2}FC > 2$ ). At 24 hpi, 158 transcripts were influenced by aphid infestation. The number of affected transcripts increased to 565 at 48 hpi and declined to 179 transcripts at 96 hpi. The analysis of DEGs highlighted the modulation of genes involved in hormone-related defence pathways. Among these, SA-related genes were found mainly up-regulated assuming an important role in “San Pasquale”-*A. gossypii* interaction. Furthermore, a significant negative effect on aphid fixing behaviour was observed on zucchini plants pre-treated with synthetic methyl salicylate (MeSA). Zucchini plant response was also characterised by the overexpression of genes involved in primary metabolic processes as well as cell wall modification. Interestingly, several aphid-derived transcripts were discovered among zucchini DEGs. Even if no conclusive evidence can be drawn, we hypothesized that these mRNAs might play a role in modulation of plant direct and/or indirect response.

Finally, analysis of VOCs emitted by zucchini plants infested with 10 adult aphids, for 48 h, showed a significant reduction in (*E*)-caryophyllene emission, whereas emission levels of other volatiles were not affected. Conversely, a significant increase in (*E*)-caryophyllene emission was observed when plants were infested with 300 adult aphids, for 96 h. Moreover, olfactometer bioassays revealed that synthetic (*E*)-caryophyllene was attractive to female *Aphidius colemani* parasitic wasps, widely used as biological control agents for *A. gossypii*. Taken together, these results suggest

(*E*)-caryophyllene may play an important role in zucchini plant indirect defence responses.

Our study allows to elucidate, for the first time, the molecular mechanisms activated by zucchini plants in response to *A. gossypii* infestation during a compatible interaction.





Chapter 1

# *General Introduction*



## 1.1 Plant defence mechanisms against insect pest

Plants and insects that feed on them are actively implicated in a war since hundreds of millions of years. Insect pests cost billions of dollars in terms of crop losses and insecticides and farmers face an ever-present threat of insecticide resistance due to large use of chemical control agents (Gordon and Waterhouse, 2007). Furthermore, the massive use of insecticides involves tremendous damages to ecosystems that induce deep modification in soil microbial communities and lead to gradual contamination of soil and water resources. Hence, there is an urgent need to develop alternative pest-control strategies for crops reducing the amount of pesticides. For all these reasons the enhancement of plant endogenous defence, through the identification of new genes and molecules able to contain harmful insect population, is considered a suitable tool for crop protection in integrated pest management strategies (Le Mire *et al.*, 2016).

Plants, in the course of evolution, have developed sophisticated systems to defend themselves against the attack of insect pests with different feeding strategies. Plant defences are commonly divided into constitutive and induced defences. Constitutive defences are physical and chemical defensive traits that plants have regardless of the presence of herbivores (Wu and Baldwin, 2010). By contrast, inducible defences are turned up only after plants are attacked by herbivores. Plant survival depends on its ability to quickly recognize, decipher the incoming signal, and adequately respond to it activating efficient defences. Such defences are triggered either to directly protect the plant, improving structural features and toxic compounds, or to indirectly protect it through molecular interactions that may attract natural enemies of herbivore insects such as predators or parasitoids. Direct defences are able to interfere with insect growth, development and reproduction using physical or chemical barriers. Physical barriers on plant surfaces, such as thorns, glandular trichomes and cuticles could prevent insect colonization and limit insect movement. Moreover, trichomes may also complement plant chemical defence producing substances that are olfactory or gustatory repellents. Plant direct defences include production of many secondary metabolites that act as powerful chemical weapons. These metabolites such as alkaloids, glucosinolates and cyanogenic glucosides, function as toxins, repellent or poisons. Moreover, proteinase inhibitors (PIs) are produced following insect attack and act as anti-digestive proteins, also reducing the nutritional value of crops. The generally accepted mode of action is that PI molecules inhibit protein digestive enzymes in insect guts, resulting in amino acid deficiencies and thereby developmental delay, mortality, and/or reduced fecundity (Gatehouse, 2011). Others defence-related proteins such as arginases, ascorbate oxidases, lipoxygenases, polyphenol oxidases, and peroxidases may have anti-nutritional properties (Mithöfer and Boland, 2012).

Plants indirect defences attract natural enemies of herbivores by releasing volatile organic compounds, green leaf volatiles and extra-floral nectars (Wu and Baldwin, 2010). The emission of volatile organic compounds (VOCs) by infested plants, that mainly consist of terpenoids, fatty acid derivatives, and aromatic compounds, can attract parasitoids or predators of the feeding insect (De Moraes *et al.*, 1998; Kessler and Baldwin, 2001). Usually VOCs blend composition depends on the mode of damage but also on specific plant-herbivore interaction. The insect feeding-induced emission of volatiles facilitates the identification of target plants and the control of pest population. However, some VOCs can also serve in direct defences as repellents to the attacking insect. For example, tobacco plants emit nocturnal VOCs

that repel *Heliothis virescens* female moths from oviposition on previously damaged plants (De Moraes *et al.*, 2001).

Biosynthesis of defensive compounds and activation of defence mechanisms are expensive, and it is not surprising that plants use complex regulatory systems to balance growth and development against defence. This is a problem especially when fitness-limiting resources, like nitrogen, are invested or if the compounds produced are toxic to the plant itself, and not only to the herbivores (Fürstenberg-Hägg *et al.*, 2013). Induced defences are complex also because different types of organisms could be recognised and elicit different responses. Chewing herbivores burst vacuoles and trichomes determining the release of defensive compounds. Piercing/sucking insects, such as aphids, do less structural damage but divert nutrient flow from the plant (Bruce and Pickett, 2007). Thus, defence response of plants changes according to different type of stressors.

## 1.2 Recognition of insect herbivore attack

Plants have the ability to recognise mechanical damage and to properly respond to herbivore attacks. This feature is essential to avoid wasting precious resources, since production and release of defence compounds only benefits herbivore-challenged plants (Fürstenberg-Hägg *et al.*, 2013). Plants defence mechanisms are activated considering herbivores feeding strategy and the quality and quantity of tissue damaged by insect pests. Moreover, herbivore-derived elicitors contained in oral secretions, also described as HAMPs (herbivore-associated molecular patterns), and herbivore-induced molecules originated from plants, elicit and modulate plant defence responses. HAMPs are described as all herbivore-derived signalling compounds that might come into contact with the particular host plants during any stage of their life cycle and thereby elicit defence reactions (Mithöfer and Boland, 2008). The first fully characterized herbivore-derived elicitor has been volicitin or N-(17-hydroxylinolenoyl)-L-glutamine, a hydroxy fatty acid-amino acid conjugate (FAC), which has been isolated from the beet armyworm *Spodoptera exigua* oral secretions (Alborn *et al.*, 1997). The application of volicitin greatly enhances volatile emission in *Zea mays* seedlings, which attracts parasitoids to feeding larvae (Alborn *et al.*, 1997). Since then, FACs have been found in oral secretions of several Lepidopteran species (Spiteller and Boland, 2003; Pohnert *et al.*, 1999; Halitschke *et al.*, 2001), and also in crickets (*Teleogryllus taiwanemma*) and fruit flies (*Drosophila melanogaster*) (Yoshinaga *et al.*, 2007). FACs have been shown to play a key role in insect nitrogen metabolism (Yoshinaga *et al.*, 2008), and hence it may be demanding for insects to avoid to synthesize FACs so as to feed stealthily on plants that use FACs to detect pest attack (Wu and Baldwin, 2010). In addition to FACs, several other types of elicitors in insect oral secretions have been discovered. For example, inceptins derived from proteolysis of the plant chloroplastic ATP synthase  $\gamma$ -subunit (cATPC) was identified in *Spodoptera frugiperda* midgut (Schmelz *et al.*, 2006). Small amounts of inceptin tested on mechanically damaged leaves of *Vigna unguiculata* (cowpea) are able to significant increase intracellular levels of hormones involved in stress signalling pathways (Schmelz *et al.*, 2007). Another class of elicitors, caeliferins, has been identified in American bird grasshoppers, *Schistocerca Americana*. In contrast to previous examples, a few number of elicitors characterized from oral secretions are able to suppress plant defence responses, as described for salivary glucose oxidase (GOX) secreted by *Helicoverpa zea* that can inhibit wound-inducible nicotine production in *Nicotiana tabacum* (Musser *et al.*, 2005). More recently, several studies on effectors released by piercing/sucking insects, such as

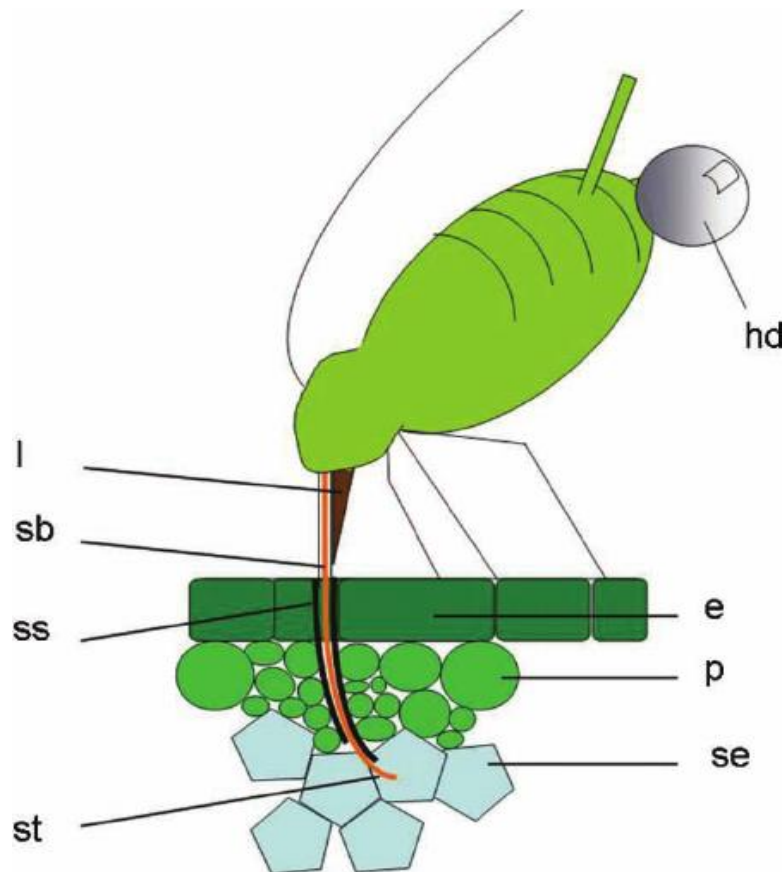
aphids, were performed. The candidate effector protein Mp10, produced by the green peach aphid *Myzus persicae*, specifically induced chlorosis and local cell death in *Nicotiana benthamiana*, indicating that this protein may trigger plant defences. Moreover, aphid fecundity assays revealed that *N. benthamiana* plants overexpressing Mp10 showed a negative effect reducing aphid fecundity (Bos *et al.*, 2010). Moreover, *N. benthamiana* plants overexpressing Mp10 activated hormone-related defence signalling and reduced susceptibility to the oomycete *P. capsici* (Rodriguez *et al.*, 2014). Aphid salivary proteins can also facilitate feeding from host plants. *Arabidopsis thaliana* plants overexpressing the *M. persicae* salivary effector Mp55 showed increased aphid reproduction in response to aphid feeding. Mp55-expressing plants also were more attractive for aphids in choice assays (Elzinga *et al.*, 2014).

Given the diversity of herbivore species and the very different fitness consequences of their attack of the plants, it is reasonable to assume that plants have developed multiple receptors and sensors that form a complex surveillance system for herbivores. Depending on the effector introduced by the insect into the plant, different signalling events may be triggered by a single or a specific combination of receptors/sensors (Wu e Baldwin, 2010).

### 1.3 Plant-aphid interaction

Aphids are major economic insect pests of plants that cause yield losses worldwide. These insects are phloem-feeders and belong to Aphididae family, which comprises more than 4300 species. Damage to plants as a consequence of aphid infestation can result in water stress, reduced plant growth and wilting. In particular, aphids can manipulate resource allocation within the plant. Aphids increase the nutritional quality of their feeding sites by enhancing the import of resources from other sites in the plant, mobilizing local resources and blocking their export to other organs (Goggin, 2007). Moreover, these insects are major vectors of economically important plant viruses. Aphids can reproduce clonally and give birth to live young, in which embryonic development begins before its mother's birth (Goggin, 2007). These traits allow for short generation times and contribute to have a tremendous negative impact on host plants. Most aphids are specialized and can only feed on one or few related plant species. However, polyphagous aphid species are considered the most dangerous because of their ability to infest many plant species, including important crops (Jaouannet *et al.*, 2014). For all these reasons it is important to identify factors that regulate plant resistance or susceptibility to these insects as background information to develop biotechnological applications for plant protection.

Prior aphids feeding activity starts, some steps are involved in initial contact with the host plant. Aphids first need to localize the host plant, usually taking advantage of volatiles emitted, and to land on it surfaces. Thin and elongated mouthpart, called stylets, enables aphids to penetrate plant tissue compartments (figure 1.1) (Powell *et al.* 2006). Moreover, a probing behaviour takes place during both aphid-host and aphid-non-host interactions, during which a molecular interaction occurred to check the compatibility of the plant species. However, during non-host interactions aphids cannot successfully feed from the phloem.



**Figure 1.1.** Schematic representation of a feeding aphid. **e:** epidermis; **hd:** honeydew droplet; **l:** labium, not participating to the piercing activity (brown); **p:** parenchyma; **sb:** stylet bundle (orange); **se:** sieve elements (blue); **ss:** stylet sheaths (black); **st:** stylet tip (from Guerrieri and Digilio, 2008).

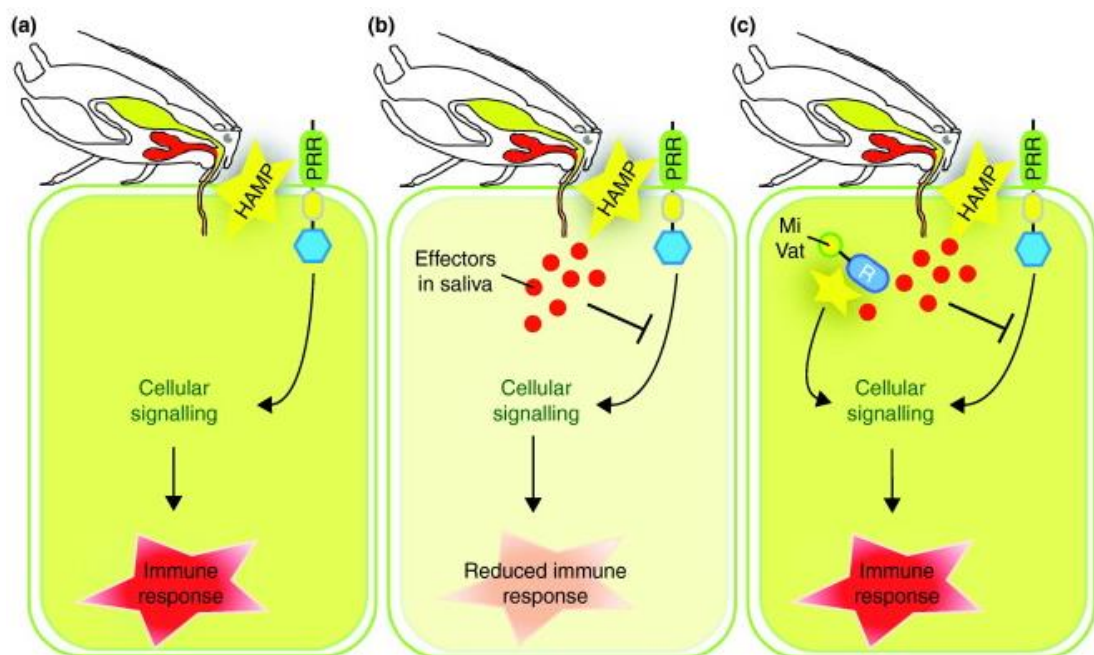
The aphid stylet must penetrate the plant epidermis and move through the cortical layer. To facilitate this process, aphids secrete proteinaceous gelling saliva which hardens to form a continuous tubular sheath encasing the full length of the stylet within the apoplast (Will and Vilcinskis, 2015). Whereas, watery digestive saliva, which is a complex mixture of enzymes (e.g. oxidases, pectinases, and cellulases) and other components (e.g. effectors) able to enhance plant defence response (Smith and Boyko, 2007), is secreted during probing and phloem sap ingestion. The stylet penetration in plant tissue is able to activate the  $\text{Ca}^{2+}$  channels in the plasma membrane of sieve elements and to promote the  $\text{Ca}^{2+}$  influx from the apoplast into the sieve element lumen. Saliva proteins represent key elements in plant defence modulation and are considered to act as herbivore-associated molecular patterns (HAMPs) (Will *et al.*, 2013).

#### 1.4 Molecular pathways activated by aphid infestation

During the constant interaction between host plant and aphids, plant defences are likely to be triggered. Progresses based on transcriptomic studies have been made in recent years identifying the molecular pathways activated during plant-aphid interaction (Thompson and Goggin, 2006; Kusnierczyk *et al.*, 2008; Coppola *et al.*, 2013). A common feature of these studies is that plants activate responses that overlaps with those related to bacterial and fungal pathogens (Zhu-Slazman *et al.*, 2005). Furthermore, the different researches, that have been carried out to study changes in gene expression induced by aphid, revealed that many of the differentially expressed genes encode proteins directly involved in defence and signalling,

oxidative burst, protein synthesis, modification and degradation, cell wall degradation and strengthening, cell maintenance, photosynthesis and secondary metabolites (Smith and Boyko, 2007; Kusnierczyk *et al.*, 2008; Delp *et al.*, 2009; Coppola *et al.*, 2013). This means that plant response to sucking insects appears to be very complex (Foyer *et al.*, 2014).

The recognition of aphid feeding by plants likely occurs through the use of transmembrane pattern recognition receptors (PRRs) or of polymorphic nucleotide-binding site-leucine-rich repeat (NBS-LRR) protein products, encoded by a majority of resistance genes that confer resistance to aphids, such as *Mi-1.2* and *Vat* (Jones and Dangl, 2006). The *Mi-1.2* resistance gene confers resistance in tomato to certain clones of *Macrosiphum euphorbiae* (potato aphid), two whitefly biotypes, a psyllid, and three nematode species (Kalosian *et al.*, 1997; Rossi *et al.*, 1998; Milligan *et al.*, 1998; Kingler *et al.*, 2005; Casteel *et al.*, 2006; Francis *et al.*, 2010), indicating that there is significant overlap in plant pathogen and aphid recognition in plants. The *Vat* gene from melon, *Cucumis melo* L., controls resistance to the cotton aphid *Aphis gossypii* Glover, and to transmission of some non-persistent viruses vectored by *A. gossypii* (Dogimont *et al.*, 2014). In addition, aphid resistance conferred by several resistance genes was shown to be race-specific (figure 1.2, (c)) (Bos *et al.*, 2010). Following the recognition of the attacker, plants activate different signal cascades that involve various signalling molecules to reprogram their phenotype (figure 1.2, (a)). Typical signal molecules include phytohormones such as Jasmonic Acid (JA), Salicylic Acid (SA), Ethylene (ET), and reactive oxygen/nitrogen species (ROS/RNS), mainly hydrogen peroxide ( $H_2O_2$ ), and nitric oxide (NO); all induce alteration in the expression of defence genes, enhancing plant defence responses (Drzewiecka *et al.*, 2014). These molecules can act separately or together, with antagonistic or synergistic interactions in the plant signalling network (Morkunas and Gabryś, 2011).



Current Opinion in Plant Biology

**Figure 1.2.** Model of the multi-layered plant defence response to aphid herbivory proposed by Hogenhout and Bos (2011). (a) Plant cells perceive aphid herbivore-associated molecular patterns (HAMPs) activating effective defence response that deters the aphid from further feeding. (b) Although plants perceive aphid HAMPs, the defence response is effectively suppressed by aphid effectors leading to aphid colonization. (c) The aphid species produces effectors that effectively suppress plant

responses, but in certain clones of this aphid species one or more effectors are being recognized by R genes leading to plant effective immune response and plant resistance to the aphid clone.

Jasmonic acid (JA) and ester-methyl jasmonate (MeJA) are linoleic acid-derived compounds and key molecules of the octadecanoid-signalling pathway (Meyer *et al.*, 1984). The JA functions in plant–aphid interactions have been described in several plants such as Arabidopsis, tobacco, tomato, wheat and sorghum (Morkunas and Gabryś, 2011). Beside JA, its precursor, 12-oxo-phytodienoic acid (OPDA), and JA methyl ester (MeJA) are also essential elements in plant defence mechanism (Korth and Thompson, 2006). Several genes encoding enzymes involved in JA synthesis and JA-mediated defence responses, such as *12-oxophytodienoate 10,11-reductase*, *Cytochrome P450* and *lipoxygenase (LOX)*, were up-regulated in aphid-resistant plants after the attack of aphids (Moran and Thompson, 2001; Voelckel *et al.*, 2004; Boyko *et al.*, 2006).

Salicylic acid (SA) promotes the development of systemic acquired resistance (SAR), a broad-range resistance against pathogens and it is involved in hypersensitive (HR) response (Smith and Boyko, 2007). The accumulation of SA and expression of SA-responsive genes following aphid feeding provided evidence of a possible involvement of this phytohormone in plant defence mechanism (Mohase and van der Westhuizen, 2002; Divol *et al.*, 2005; Zhu *et al.*, 2010). Induction of SA-pathway in aphid-resistant wheat plant challenged by *Diuraphis noxia*, and the increase in expression of SA-dependent genes in aphid-susceptible Arabidopsis, sorghum and tomato, support a predominant role of this phytohormone in resistance mechanism (Moran and Thompson, 2001; Moran *et al.*, 2002; Zhu-Salzman *et al.*, 2004; Botha *et al.*, 2010; Coppola *et al.*, 2013).

Strong production of ET was observed in susceptible cultivars of alfalfa and wheat after early infestation by *Schizaphis graminum* and the spotted alfalfa aphid, *Therioaphis maculata* (Dillwith *et al.*, 1991; Anderson and Peters, 1994). Moreover, the expression of genes encoding proteins involved in ET production or ET signalling (e.g. ethylene-responsive elements) was up-regulated in aphid-susceptible celery infested with *M. persicae* (Divol *et al.*, 2005) and in aphid-resistant wheat infested with *D. noxia* (Boyko *et al.*, 2006). ET production was enhanced in both resistant and susceptible plants in response to aphids, suggesting that ET may be involved in basal defence against phloem-feeders (Drzewiecka *et al.*, 2014).

The cross-talk of JA, SA, and ET takes place in a complex network of interconnecting signalling pathways, but it is essential to develop the best defensive strategies. SA and JA are known to antagonistically interact in plant responses to herbivore attacks, and SA is involved in suppression of endogenous production of JA when it reaches a certain level (Mur *et al.*, 2006). The synergistic interaction between SA and JA, however, has been described. Kusnierczyk and colleagues (2007; 2011) reported that both SA- and JA/ET-responsive genes were significantly induced in Arabidopsis following *Brevicotynae brassicae* and *M. persicae* attack. The signalling pathways active against aphids are driven not only by phytohormones, but also by ROS/RNS that contribute to the production of plant defence compounds. The involvement of ROS in pathogen resistance is well documented and genes involved in oxidative signal transduction through control of cellular hydrogen peroxide concentration are modulated by aphid infestation in both aphid-susceptible and aphid-resistant plants (reviewed in Smith and Boyko, 2007). Among other components that act in defence against aphids independently from phytohormones we can report *PAD4 (Phytoalexin deficient 4)* as an example. *PAD4*, which encodes a lipase-like protein, contributes to defence response to *M. persicae* in Arabidopsis plants with an effect on aphid



reproduction (Pegadaraju *et al.*, 2005, 2007). Sugars also function as messengers in plant signalling pathways after aphid infestation. Increased synthesis of sugar transporters and modification in expression of genes associated with sugar metabolism occurred during aphid feeding. These processes contribute to the creation of nutrient sinks at aphid-feeding sites due to phloem sap removal (Smith and Boyko, 2007).

While research on plant-aphid interaction has long been focused on the plant side, with resistance genes, secondary metabolites and hormones being discovered, there has been a recent shift to the aphid-side of the interaction. It is well known that aphids are able to modify host morphology (van Emden and Harrington, 2007), nutrient allocation (Girousse *et al.*, 2005) and to suppress defence responses through their feeding behaviour (Will *et al.*, 2007). Modulation of plant defence is possibly due to delivery of aphid effectors inside their hosts which act as HAMPs, enabling successful infestation of plants (figure 1.2, (b)) (Bos *et al.*, 2010). As previously reported, recognition by plants of aphid effector by specific receptors (NBS-LRR) is strongly related to the genotype plants belong to, and certain aphid clones may be able to avoid and/or suppress plant defences.

Aphids can act preventing wound-induced plugging of sieve plates so their nutrition supply is not interrupted (Will and van Bel, 2006). Two possible strategies could prevent the  $\text{Ca}^{2+}$ -dependent callose deposition: (i) reduction of calcium influx into sieve elements; (ii) sequestration of calcium ions inside sieve elements. The wound inflicted by stylet penetration is immediately sealed by sheath saliva (Miles, 1987) so that influx of cell wall  $\text{Ca}^{2+}$  is prevented. Moreover, the presence of  $\text{Ca}^{2+}$ -binding proteins in watery aphid saliva can limit or suppress the intracellular  $\text{Ca}^{2+}$ -dependent defence response (Goggin, 2007). As a weapon against callose deposition, the presence of the callose-hydrolysing enzyme 1,3- $\beta$ -glucanase in watery saliva (postulated by Dorschner, 1990) may also assist in removal of sieve-plate callose.

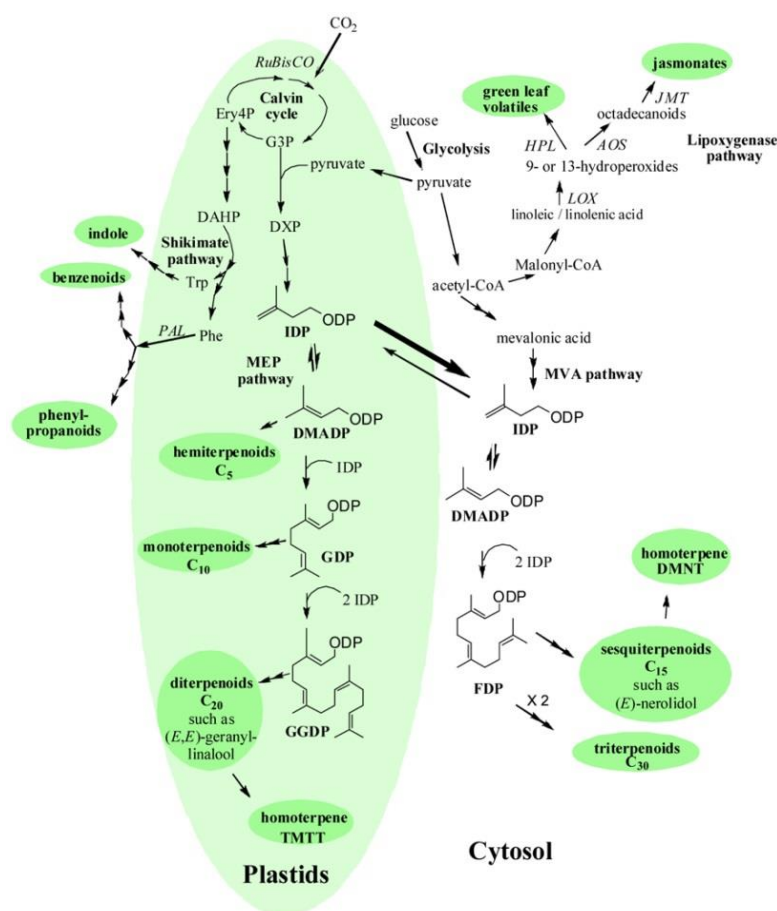
### 1.5 Plant volatile organic compounds

Plants synthesize an incredible diversity of VOCs that facilitate their interaction with the environment. Plant volatiles are typically lipophilic liquids with low molecular weight and high vapour pressure at ambient temperatures (Dudareva *et al.*, 2013). Non-conjugated volatiles can freely cross cellular membrane and be released into the atmosphere (Pichersky *et al.*, 2006). At present, more than 1700 volatile compounds have been isolated from more than 90 plant families. Moreover, recent progresses in “omics” approaches resulted in the identification of gene encoding enzymes involved in a large number of plant volatile biosynthetic pathways. Consequently, regulatory systems involved in VOC production have been elucidated. Plant volatiles strongly influence the ecological relation between plants and insects, providing important cues for insects in their search for host plant on which to feed (Bruce and Pickett, 2011). However, most volatiles are involved in species-specific interactions allowing herbivorous insects, pollinators as well as predatory insects to recognize the specific volatile blend of target host (Dike and van Loon, 2000). During evolution plants were forced to act in different directions avoiding producing recognizable molecules for herbivorous insects and emitting volatiles still attractive to beneficial pollinators and natural enemies of pests (Dike and Baldwin, 2010). Biosynthesis of VOCs depends on availability of building blocks derived from primary metabolism, demonstrating the high degree of connectivity between primary and secondary metabolism. In flowers the biosynthesis of volatiles occurs in epidermal cells, allowing an easy release in the atmosphere. In vegetative organs these molecules may be synthesized in glandular

trichomes and then secreted from the cell, or synthesized in internal structures, such as specialized cells, accumulated in storage vacuoles, and then released upon disruption (for example by herbivore) (Pichersky *et al.*, 2006). An herbivore-induced plant volatile (HIPV) blend may comprise more than 200 compounds (Dicke and van Loon, 2000), but often the same basic constituents are found as the major products. The composition of the blends also strongly depends on the type of damage such as herbivore feeding (Paré and Tumlinson, 1996) and egg deposition (Hilker and Meiners, 2002).

HIPV can directly influence insect physiology and behaviour due to their toxic, repelling, or deterring properties (Bernasconi *et al.*, 1998; De Moraes *et al.*, 2001; Kessler and Baldwin, 2001; Vancanneyt *et al.*, 2001; Aharoni *et al.*, 2003). They can also attract enemies of attacking herbivores, such as parasitic wasps, flies or predatory mites, which can protect the plant from further damage (Dicke *et al.*, 1990; Turlings *et al.*, 1990; Vet and Dicke, 1992; Paré and Tumlinson, 1997; Drukker *et al.*, 2000; Kessler and Baldwin, 2001).

VOCs constitute about 1% of plant secondary metabolites and are mainly represented by terpenoids, phenylpropanoids/benzenoids, fatty acid derivatives, and amino acid-derived products (Dudareva *et al.*, 2004). Although undamaged healthy plants constitutively emit some of these compounds, considerably higher amounts are emitted after herbivore damage. Many of them are synthesized *de novo* after stress perception (figure 1.3) (Maffei *et al.*, 2007).



**Figure 1.3.** Simplified scheme of the interactions among the biosynthetic pathways responsible for volatile and non-volatile stress metabolites in plants. Pathway names are in italics, volatile compound classes are in bold inside ellipses, and the key enzymes involved in biosynthetic pathways are next to the arrows in italics. Abbreviations: acetyl-CoA, acetyl coenzyme A; AOS, allene oxide synthase;

DAHP, 3-deoxy-D-arabino-heptulosonate 7-phosphate; DMADP, dimethylallyl diphosphate; DMNT, 4,8-dimethyl-1,3E,7-nonatriene; DXP, 1-deoxy-D-xylulose 5-phosphate; Ery4P, erythrose 4-phosphate; F6P, fructose 6-phosphate; FDP, farnesyl diphosphate; G3P, glyceraldehyde-3-phosphate; GGDP, geranylgeranyl diphosphate; GDP, geranyl diphosphate; HPL, fatty acid hydroperoxide lyases; IDP, isopentenyl diphosphate; JMT, jasmonic acid carboxyl methyl transferase; LOX, lipoxygenase; MEP-pathway, methylerythritol 4-phosphate pathway; MVA, mevalonic acid; PAL, phenylalanine ammonia lyase; PEP, phosphoenolpyruvate; Phe, phenylalanine; TMTT, 4,8,12-trimethyl-1,3(E),7(E),11-tridecatetraene (from Niinemets *et al.*, 2013).

Terpenoids compose the largest class of plant secondary metabolites with many volatile representatives (Dudareva *et al.*, 2006). All terpenoids originate from isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMPP), which are synthesized via two alternative pathways. The cytosolic mevalonate (MVA) pathway begins with the formation of IPP from three molecules of acetyl-CoA (Dewick, 1999), while the plastidial 2-C-methyl-D-erythritol 4-phosphate pathway (MEP) starts with condensation of pyruvate and glyceraldehyde-3-phosphate (Lichtenthaler *et al.*, 1997; Rohmer, 1999). The route in plastids provides precursors for the biosynthesis of isoprene, mono-, and diterpenes, while the cytosol-localized pathway for sesqui- and triterpenes. Precursors of terpenoids have been experimentally demonstrated to be transported from plastids to the cytosol (Dudareva *et al.*, 2005; Bartram *et al.*, 2006), referred to as the “cross-talk” between the MEP- and MVA-pathways.

The lipoxygenase pathway starts with the dehydrogenation of linolenic and linoleic acids at C9 or C13 position by lipoxygenases forming 9-hydroperoxy and 13-hydroperoxy derivatives of polyenic acids (Hatanaka, 1993; Howe and Schaller, 2008). These derivatives can be further metabolized by an array of enzymes, including allene oxide synthase (AOS) and hydroperoxyde lyase (HPL), which represent two branches of the lipoxygenase pathway yielding volatile compounds (Dudareva *et al.*, 2006). In the HPL branch, these compounds are further cleaved by hydroperoxide lyases into oxoacids and C6-aldehydes. These aldehydes can be converted to their isomers by spontaneous rearrangement or by alkenal isomerases, or they can be reduced into the corresponding alcohols by alcohol dehydrogenases (AOS; Akacha *et al.*, 2005). In the AOS branch of the lipoxygenase pathway, 13-hydroxyperoxy linolenic acid is converted to 12,13-epoxy octadecatrienoic acid by AOS (Feussner and Wasternack, 2002). A series of subsequent enzymatic reactions leads to the formation of jasmonic acid, which can in turn be converted into the volatile ester, methyl jasmonate, by the enzyme jasmonic acid carboxyl methyltransferase (Seo *et al.*, 2001; Song *et al.*, 2005).

Finally, phenylpropanoids and benzenoids derived from L-phenylalanine constitute a large class of structurally diverse volatile compounds involved in plant reproduction and defence (Dudareva *et al.*, 2006). Aromatic volatiles are formed via the shikimic acid pathway, starting from condensation of erythrose 4-phosphate and PEP. After numerous steps, phenylalanine (Phe) is produced and it is further converted to trans-cinnamic acid by phenylalanine ammonia lyase (PAL) (Niinemets *et al.*, 2013). Trans-cinnamic acid is a starting point for the synthesis of phenylpropanoids, (e.g., phenylethanol, phenylethylbenzoate) and benzenoids (benzaldehyde, methyl benzoate, methyl salicylate etc.; Boatright *et al.*, 2004; Dudareva *et al.*, 2006). Additionally, tryptophan (Trp), which is the precursor of volatile indole, is biosynthesized via shikimic acid pathway (Paré and Tumlinson, 1996) in chloroplast, while indole itself is synthesized in cytosol (Zhang *et al.*, 2008).

## 1.6 *Cucurbita pepo* L. characteristics and genomic resources

*Cucurbita pepo* L. ( $2n=2x=40$ ) belongs to the Cucurbitaceae family, the second-most large vegetable crop family of economic importance after Solanaceae (Esteras *et al.*, 2012). This family includes several important vegetable crops cultivated worldwide, such as watermelon (*Citrullus lanatus* L.), cucumber (*Cucumis sativus* L.), melon (*Cucumis melo* L.) and squashes (*Cucurbita* spp.). The main characteristic of this family is the rich diversity of many important traits. Precisely for this reason Cucurbitaceae family has been used as model for sex expression analyses and study of mechanisms involved in fruit development and ripening (Li *et al.*, 2009; Ezura and Owino, 2008).

Historical records report that *Cucurbita pepo* is native to North America and was dispersed to other continents during the 16<sup>th</sup> century by transoceanic travels (Paris, 2008). *C. pepo* is extremely variable in fruit-related features. The edible forms of this species can be grouped in two sub-species: ssp. *pepo* that includes Pumpkin, Vegetable Marrow, Cocozelle and Zucchini; spp. *ovifera* that includes Acorn Squash, Scallop, Crookneck and Straightneck. The great economic value of this species is based on the consumption of immature fruits as vegetables, collectively named “summer squashes”, but also Pumpkin and Acorn, known as “winter squashes”, display a large use as mature fruits (Blanca *et al.*, 2011). Zucchini types rank among the highest-valued vegetables worldwide, which are cultivated in temperate and subtropical areas and represent a rich source of nutrients, such as vitamins and minerals. More than 25 million tons of zucchini (also reported as squash), together with pumpkins and gourds, were produced in the world in 2014 on a cultivated area of about 2 million hectares. China is the largest producer followed by India and Russian Federation. Italy, with more than 566,000 tons on a cultivated area of approximately 18.000 hectares, is the eighth producer country worldwide and the first producer among Mediterranean countries (FAOSTAT 2014). The major production areas are located in Southern Italy (Sicily 3,622.55 ha; 2,941.78 Lazio ha; Puglia 1,852.62 ha; Campania 1,291.3 ha; Calabria 1,218.62 ha), but zucchini cultivation is also spread in the North (Piemonte 1,420.85 ha; Veneto 1,473.5 ha; Emilia Romagna 1,428.2 ha) (ISTAT 2011). However, high production levels can be achieved by strenuous efforts to fight against numerous diseases affecting *C. pepo*. One major issue related to zucchini cultivation, both in greenhouse and open-field, is represented by diseases caused by fungi and viruses as well as by damage caused by insects. Among fungi, powdery mildew, caused by *Podosphaera xanthii*, is a serious disease affecting leaves, stems and fruits of cucumber and zucchini squash that reduces fruit quality and yield (Cohen *et al.*, 2003). On the other hand, among insects, aphids are evaluated the most dangerous. In particular, *Aphis gossypii* Glover (Homoptera: Aphididae) is considered the major pest of cotton and cucurbit species. The cotton/melon aphid *A. gossypii* has long been regarded as a cosmopolitan, highly polyphagous species, widely distributed in warm climate regions (Singh *et al.*, 2014). Damage is direct through feeding that can induce leaves deformation and stunted growth, leading to host death and also reducing productivity long before plant death. Furthermore, indirect damage is dependent to transmission of serious viruses, especially the Zucchini Yellow Mosaic Virus (ZYMV), and to deposition on plant tissue surfaces of honeydew. Honeydew causes economic loss through physical contamination and through providing a nutrient source for fungi that contaminate products and reduce photosynthesis rates by blocking sunlight (Ebert and Cartwright, 1997).

In the past, this aphid has been controlled by a wide array of insecticides. The growing concern over the use of pesticides in agriculture is a major theme due to

environmental contamination and the economic impact of pesticide resistance. Moreover, so far, effective biopesticides, such as *Bacillus thuringiensis*, are not available for aphid control. The development of defence strategies based on the usage on natural molecules against these biotic stressors, reducing dependence on pesticides, represents one of the main and novel objectives for zucchini breeding programs. Recent advances in next-generation sequencing technologies represent an important opportunity for studying in-depth the molecular mechanisms of plant-insect interactions in non-model species and to continue controlling this pest in a sustainable manner.

Despite the agricultural and economic importance there are few genomic and genetic resources available for *C. pepo*, unlike other cucurbit plants. The whole genome of the domestic cucumber, *Cucumis sativus* var. *sativus* L., was assembled using a combination of traditional Sanger and Illumina sequencing technologies (Huang *et al.*, 2009). The complete sequence of melon (*Cucumis melo* L.) genome was published in 2012 in the frame of the Spanish project Melonomics whose aims were the sequencing and the study of the melon genome (Garcia-Mas *et al.*, 2012). Moreover, a high-quality draft genome sequence of watermelon (*Citrullus lanatus* L.) cultivar from East Asia was released (Guo *et al.*, 2013). Several others genomic resources are available for these crops and also for *Cucurbita maxima* Duchesne and *Cucurbita moschata* Duchesne. Detailed physical and genetic maps (Deleu *et al.*, 2009; Ren *et al.*, 2009; Ren *et al.*, 2012; Diaz *et al.*, 2011; Wei *et al.*, 2014; Zang *et al.*, 2015), mapping populations (Fernandez-Silva *et al.*, 2008), microarrays (Wechter *et al.*, 2008; Mascarell-Creus *et al.*, 2009), reverse genetic platforms (Dahmani-Mardas *et al.*, 2010; González-Ballester *et al.*, 2011; Frenkel *et al.*, 2012) and transcriptomes (Guo *et al.*, 2010; Guo *et al.*, 2011; Blanca *et al.*, 2011; Blanca *et al.*, 2012; Ando *et al.*, 2012; Wu *et al.*, 2014) have been developed and generated. Many of these resources are available at the Cucurbit Genomics Database (<http://cucurbitgenomics.org/>), the new web site of the International Cucurbit Genomics Initiative (ICuGI). Until 2011 only three genetic maps of *Cucurbita* have been constructed: two maps from inter-specific crosses between *C. pepo* and *C. moschata* (Lee *et al.*, 1995; Brown and Myers, 2002) and the third from an intra-specific cross of *C. pepo* (Zraidi *et al.*, 2007). These maps contain only RAPD (Random Amplified Polymorphic DNA) and AFLP (Amplified Fragment Length Polymorphism) markers. Later, to increase map density, a collection of 178 microsatellites, also referred to as simple sequence repeats (SSRs), and 105 new AFLP markers were developed (Gong *et al.*, 2008). Esteras and colleagues (2012) developed the first SNP-based (Single Nucleotide Polymorphisms) genetic map of *Cucurbita pepo* using a population derived from the cross of two varieties with contrasting phenotypes, Zucchini (subsp. *pepo*) × Scallop (subsp. *ovifera*). Moreover, this map was used to infer syntenic relationships between *C. pepo* and cucumber and to successfully map QTL that control plant flowering and fruit traits for breeding purposes (Esteras *et al.*, 2012). Production of this dense genetic map was possible thanks to a large collection of molecular markers generated by Blanca and colleagues (2011). More recently, a high-density SNP-based genetic map (more than 7,000 markers) was developed using a RIL (Recombinant Inbred Line) population derived from the cross between the two *C. pepo* subspecies Zucchini and Scallop (Montero-Pau *et al.*, 2017). Such a map improved the previously reported *C. pepo* SNP-based map released by Esteras and colleagues (2012). In the same study authors investigated the genetic control of economically important quality traits by QTL analysis, taking advantage of the new high-density map and of the first draft of

the *C. pepo* genome available at the Cucurbigene web site (<https://cucurbigene.upv.es>) (Montero-Pau *et al.*, 2017). In 2011, Blanca and colleagues generated the first transcriptome of *C. pepo* using the 454 GS FLX Titanium technology. A total of 49,610 unigenes derived from flower, leaf and root tissue of two contrasting *C. pepo* cultivars, Zucchini and Scallop, were assembled from 512,751 new EST (Expressed Sequence Tags), and used to generate the first large collection of EST-derived SSR and SNP in this species (more than 10,000 potential molecular markers) (Blanca *et al.*, 2011). Furthermore, recently other two squash *de novo* transcriptome assembly were published to study fruit quality and morphology and to identify genes related to fruit development and ripening (Wyatt *et al.*, 2015; Xanthopoulou *et al.*, 2016). All genomic resources reported above are invaluable tools useful for future mapping and diversity studies, and will be essential to accelerate the process of breeding new and better-adapted squash varieties (Esteras *et al.*, 2012).

### 1.7 Aim of the thesis

The main objective of this thesis is to investigate zucchini transcriptome reprogramming following aphid infestations by RNA-seq and to associate modified gene expressions with direct and indirect plant defence responses. Specifically, this work used a zucchini local variety, named “San Pasquale” and extensively cultivated in Campania region, which is highly susceptible to melon aphid *Aphis gossypii* Glover attacks.

The lack of a *C. pepo* reference genome until 2017 and the availability of transcriptomes assembled starting from plant tissues in physiological conditions, has led to *de novo* assembly of a custom zucchini transcriptome from aphid-challenged leaf tissues (**Chapter 2**). Furthermore, analyses to determine the quality of the assembly were performed and, when possible, a biological function was attached to assembled transcripts. This resource was subsequently used as a reference to provide insights into changes in defence-related gene expression. To this end, a time-course transcriptomic analyses based on RNA-Seq was carried out to investigate zucchini responses during a compatible interaction (**Chapter 3**). As expected the aphid feeding behaviour induced changes in expression of genes involved in both primary and secondary metabolisms. Moreover, genes involved in stress and defence response, signalling and transcriptional regulation were found influenced by *A. gossypii* infestation. In addition, volatile organic compounds emitted by un-infested and infested plants were identified to highlight the effect of infestation on zucchini indirect defence response (**Chapter 4**).

Finally, in **chapter 5** we present an overview of the results obtained underlining the importance of the new knowledge achieved.



## Chapter 2

# *De novo transcriptome assembly of the zucchini variety “San Pasquale”*

Alessia Vitiello in collaboration with  
Nunzio D’Agostino, Giandomenico Corrado and  
Rosa Rao





## Abstract

Zucchini (*Cucurbita pepo*) ranks among the highest-valued vegetables worldwide. One major issue related to zucchini cultivation, both in greenhouse and open-field, is the damage imposed by aphids such as *Aphis gossypii* (Homoptera: Aphididae). In the present study, the transcriptome of *C. pepo* cultivar “San Pasquale” was sequenced to obtain a *de novo* transcriptome assembly from leaves un-infested and infested by *A. gossypii*, using an Illumina HiSeq 2500 platform. Leaf material was collected at three different time points (24, 48, 96 h) after infestation. The sequencing generated ~34 million of paired-end reads of 101 nucleotides in length per sample. Raw reads were pre-processed to remove sequences with low quality bases and adapter contaminations. High quality reads were *de novo* assembled using Velvet/Oases and CAP3 tools into a non-redundant set 71,648 transcripts with an average length of 1,331 nucleotides. About 94% of the assembled transcripts contain coding sequences that could be translated into proteins. BLAST similarity-based searches were performed against (i) *Cucumis melo* protein complement; (ii) *Arabidopsis thaliana* proteins; (iii) *Cucumis sativus* protein complement and (iv) the UniProtKb/SwissProt database. About 70% of transcripts found at least one correspondence in the four databases queried. Furthermore, BLASTn comparisons with the publically available *C. pepo* transcriptome identified 1,313 transcripts exclusively assembled in the aphid-challenged transcriptome. Over 70% of the transcripts were functionally annotated and assigned to one or more Gene Ontology (GO) terms. The 50% of GO terms were assigned to the biological process domain, and 27% and 21% of GO terms were assigned to molecular function and cellular components domains, respectively.

The dataset of zucchini transcripts we generated provides a resource for gene discovery and for the development of novel control strategy for *A. gossypii*.

## 2.1. Introduction

### 2.1.1 RNA-Seq

The transcriptome is the set of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA, transcribed by an organism. Transcriptome analysis is essential for understanding the functional elements of the genome and highlighting the molecular processes activated during development and disease in specific group of cells or tissues.

Microarray is a technique employed until recently for its ability to measure the expression levels of thousands of genes in a single experiment, but lacks the capacity to detect novel transcripts (Unamba *et al.*, 2015).

NGS rapid development has provided a new method for transcriptome mapping and quantification. RNA sequencing (RNA-Seq) can record the repertoire of expressed sequences found in a particular tissue at a specific time point and growth stage through rapid generation of large expression datasets. In this way, it can produce a nearly complete picture of transcriptomic events in a biological sample (Strickler *et al.*, 2012). RNA-Seq analysis include the conversion of a population of RNA (total or fractionated, such as poly(A)+) in a library of cDNA fragments with adaptors ligated to one or both ends. Each molecule, with or without amplification, is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing) (Wang *et al.*, 2009). Following sequencing, raw reads are either aligned to a reference genome or transcriptome, or

are *de novo* assembled to give information on both transcriptional composition and gene expression levels of target samples.

Unlike hybridization-based approaches, RNA-Seq is not limited to detecting transcripts that correspond to existing genomic sequence. *De novo* sequencing of transcripts is a valuable method of providing genomic resources in non-model species (for which there is sufficient knowledge of evolution and ecology but little genomic resources). The RNA-Seq strategy can also reveal sequence variations, for example SNPs in the transcribed regions (Novaes *et al.*, 2008; Alagna *et al.*, 2009; Blanca *et al.*, 2011; D'Agostino *et al.*, 2013), isoform and novel splice-junctions (Xu *et al.*, 2015; Zhang *et al.*, 2017) and identification of novel transcripts (Denoeud *et al.*, 2008; Roberts *et al.*, 2011).

Moreover, RNA-Seq technologies are useful for digital gene expression profiling and detection of genes differentially expressed in specific conditions in both model and non-model species (Molina *et al.*, 2008; Vera *et al.*, 2008; Alagna *et al.*, 2009; Yang *et al.*, 2016).

### 2.1.2 *Cucurbita pepo* L. transcriptome

Transcriptome generation through RNA sequencing is a technology that can be used in the dissection of complex traits. Assembled transcriptomes also provide valuable sequence resources in species lacking a sequenced genome. However, one limitation of RNA-Seq data is that it is specific to the plant line, tissue, developmental stage and physiological condition sequenced. For this reason, it is essential to use transcriptome data relevant to the experimental question of interest (Wyatt *et al.*, 2015).

Although the economic importance and the growing attention during the last years, at present few transcriptomic studies using the RNA-Seq approach are available for *Cucurbita pepo*.

The first *C. pepo* transcriptome was sequenced using a 454 GS FLX Titanium platform and *de novo* assembled by Blanca and colleagues (2011). Two different cDNA libraries derived from leaves, flowers and roots from two *C. pepo* cultivars highly different in flower and fruit phenotypes were used (MU16 *C. pepo* subsp. *pepo* cv Zucchini, and UPV196 *C. pepo* subsp. *ovifera* cv Scallop). The assembled unigenes were functionally annotated and also screened for the identification of SSRs and SNPs. Molecular markers identified in that study constitute an important resource for mapping and marker-assisted breeding in *C. pepo* and *Cucurbita* genus (Blanca *et al.*, 2011) and were successfully used to build a SNP-based genetic map with an F2 population (Zucchini × Scallop) and to detect QTLs for the very first time (Esteras *et al.*, 2012).

More recently, other two *C. pepo* transcriptomes were *de novo* assembled using an RNA-Seq approach from Acorn squash and Pumpkin (*C. pepo* subsp. *ovifera*) respectively. The Acorn squash fruit and seed transcriptome from the cultivar “Sweet REBA” was generated to provide insight into winter squash fruit and seed development (Wyatt *et al.*, 2015).

Furthermore, Xanthopoulou and colleagues (2016) released two transcriptomes derived from sequencing of cDNA libraries obtained from leaves and female flowers of two contrasting Pumpkin cultivars, “Big Moose” and “Munchkin”, using an Illumina HiSeq™2000 device. This resource was used to perform comparative transcriptome analyses in order to identify new genes associated with fruit morphology and size, and to develop EST-SSR markers (Xanthopoulou *et al.*, 2017).

The main goal of the transcriptomic studies reported above was mainly to improve knowledge of molecular mechanisms relate to fruit quality and development, and the identification of molecular markers and QTLs useful for squash breeding programs. However, it is important to emphasize once more that RNA-sequencing is the production of information that is specific to the plant line, tissue, developmental stage and physiological condition analyzed. For this reason, tissues that will be analysed should be selected with great care and considering the experimental question of interest. Among these high valuable resources available for *Cucurbita* genus no one take into account the purpose of breeding for resistance to biotic stress, particularly to diseases and insect pests.

Breeding for disease and pest resistance is considered one of the most important goal of several crops breeding programs and also of summer squash breeding, as *C. pepo* is highly susceptible to diseases (Whitaker and Robinson, 1986). Main zucchini diseases are caused by fungi and viruses as well as damages induced by insects pest, such as aphids.

Within this scenario, our aim is to generate a *de novo* assembled zucchini transcriptome using leaves from the cultivar “San Pasquale” which was found to be highly susceptible to the cotton/melon aphid *Aphis gossypii*.

A far as we know, this is the first study in which zucchini transcriptome is *de novo* assembled using leaves form a susceptible cultivar challenged by an insect pest.

## 2.2 Results

### 2.2.1 RNA-Sequencing and raw data pre-processing

To construct *C. pepo* transcriptome suitable for this study, leaf tissue from aphid-infested and un-infested (control) plants were harvested for RNA isolation. Total RNA was isolated form 18 samples and RNA quality and quantitative analyses were performed. A good quality RNA, suitable for RNA-seq experiment, should be consistent with the following thresholds: absorbance 260nm/280nm and 260nm/230nm ratios higher than 1.8, concentration value higher than 200 ng/μl, RIN (RNA Integrity Number) value higher than 6.50. RNA concentration and quality parameters obtained for each sample are summarized in table 2.1. The lowest value of RNA concentration was 252.4 ng/μl for sample A96-2 and the highest 1950.2 ng/μl for A24-1 sample; the ratio of the absorbance at 260 nm and 280 nm was always higher than 2.1 and the absorbance ratio A260/A230 was from 1.96 to 2.32; the RNA integrity number (RIN) value ranged between 6.80 and 7.80.

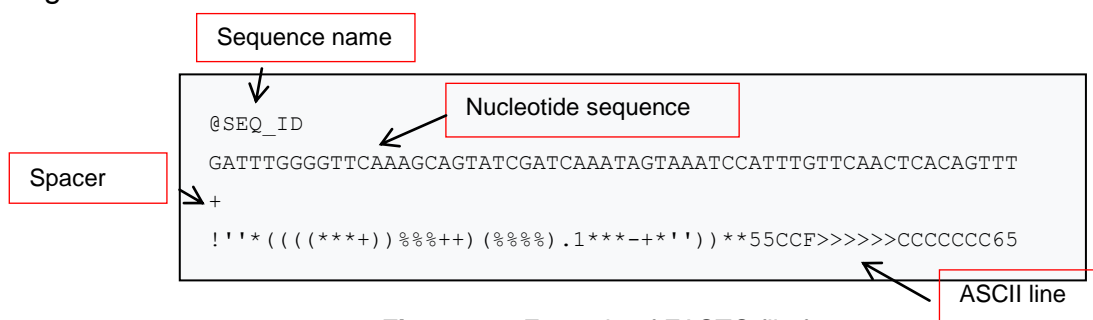
**Table 2.1.** Concentration and quality parameters of RNA extracted from infested and un-infested zucchini leaves.

Condition	Sample name	Concentration (ng/μl)	260/280	260/230	RIN
Control	C24-1	1395.8	2.15	2.15	7.50
	C24-2	1794.1	2.16	2.08	7.50
	C24-3	1376.5	2.18	2.27	7.50
	C48-1	1061.9	2.13	1.96	7.50
	C48-3	1328.9	2.18	2.27	6.80
	C48-4	1313.3	2.14	2.24	7.60
	C96-2	1532.7	2.19	2.28	7.80
	C96-3	840.1	2.12	2.15	7.80
Infested	C96-4	1438.2	2.16	2.27	7.40
	A24-1	1950.2	2.16	1.99	7.40
	A24-2	1844.5	2.15	2.18	7.20
	A24-3	1607.3	2.15	2.00	7.50
	A48-2	1924.9	2.15	2.32	7.70
	A48-3	1627.2	2.16	2.20	7.80
	A48-4	756.5	2.19	2.27	7.80
	A96-1	680.5	2.17	2.29	7.50
	A96-2	252.4	2.14	2.14	7.20
	A96-3	627.3	2.12	2.21	7.40

RNA sequencing was performed on an Illumina HiSeq 2500 device. The sequencing generated ~34 million of paired-end reads of 101 nucleotides in length per sample. Raw reads from sequencing were released in FASTQ format and two files were produced for each sample containing forward and reverse reads, respectively. FASTQ is a text-based format for storing nucleotide sequence and also information related to their quality; each sequence is normally described by four lines:

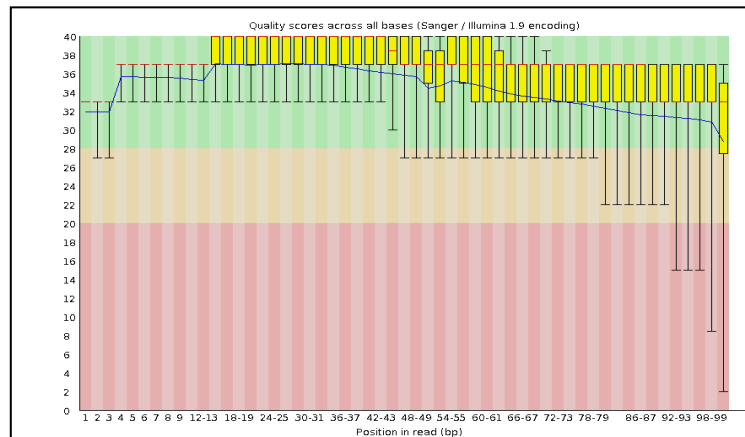
- **Line 1** begins with a '@' character and is followed by a sequence identifier and an optional description (like a FASTA title line).
- **Line 2** contains the raw sequence.
- **Line 3** begins with a '+' character and is optionally followed by the same sequence identifier (and any description) again.
- **Line 4**, representing the ASCII line, encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as nucleotides in the sequence.

A FASTQ file containing a single sequence might look like the example reported in figure 2.1.

**Figure 2.1.** Example of FASTQ file format.

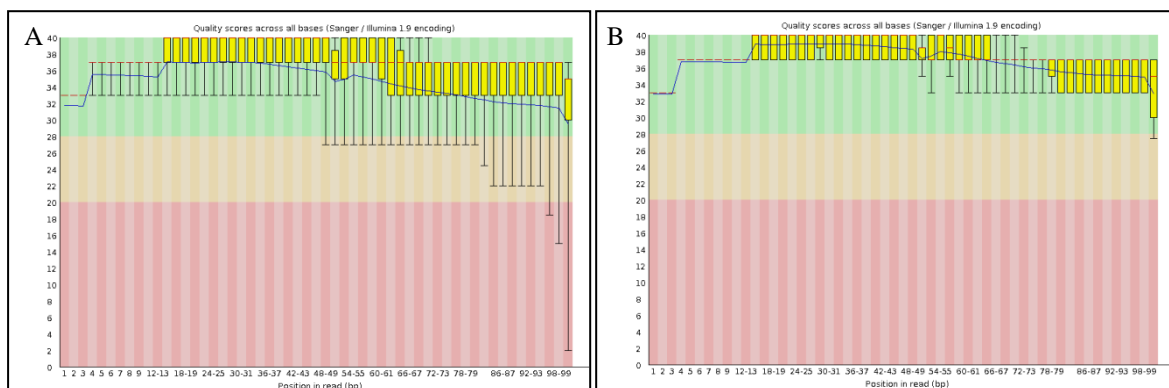
FASTQ files were analysed using the FastQC software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) for a preliminary evaluation of sequence quality. In figure 2.2 it is reported, as an example, the output

from FastQC. *Per base sequence quality* report is a graphic representation of quality values assigned to each sequenced nucleotide. On the y-axis the quality values as Phred Quality scores (or Q score) are reported; based on Q scores, the graph is divided into three coloured portions. When yellow boxes, which represent groups of nucleotides, are in the green area they are tagged as high-quality. When yellow boxes are in the orange or red areas, base calls have an intermediate and poor quality, respectively (figure 2.2).



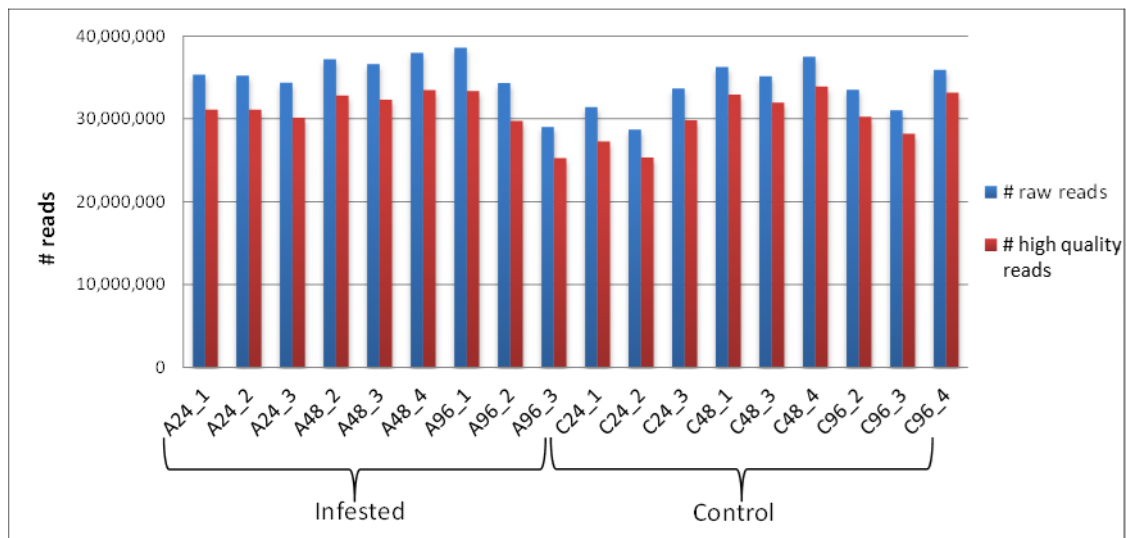
**Figure 2.2.** FastQC *Per base sequence quality* graph for the sample A48-4-R2.

Low quality reads with a Q score  $\leq 30$  in over 80% of the length of the read were removed. Illumina adapters and unassigned bases (N bases) were also removed using the Trimmomatic software (Bolger *et al.*, 2014). In figure 2.3 it is reported an example of FastQC *Per base sequence quality* graphs before and after the pre-processing step.



**Figure 2.3.** FastQC *Per base sequence quality* graphs for the sample A24-2-R2 before (A) and after (B) low quality reads removal and adapters trimming.

Raw read pre-processing resulted in the reduction of the number of reads (Appendix: table A.1). In total, about 552,4 million of high-quality reads of 75-101 nucleotides in length were obtained and about 4 million of reads were filtered out for each sample (figure 2.4). Moreover, three FASTQ files were obtained per sample: two files in which were reported paired high quality reads (R1 and R2) and one file which included all unpaired reads.



**Figure 2.4.** Pre-processing results. The number of raw reads from sequencing (blue bars) and the number of high quality reads (red bars) are reported.

### 2.2.2 *De novo* assembly and quality evaluation of zucchini transcriptome

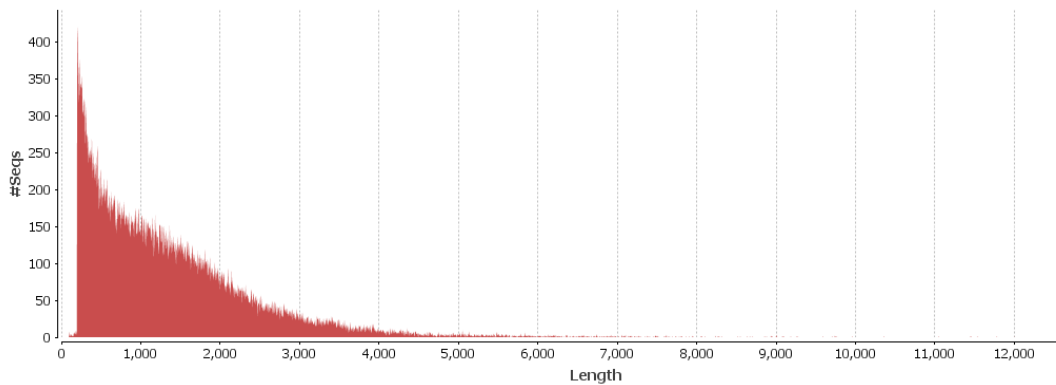
Illumina high quality reads, both paired and unpaired, from eighteen zucchini cDNA libraries were combined to build a *de novo* *C. pepo* reference transcriptome. To perform transcriptome assembly, the *insert size* of paired-end reads was calculated using the Picard software (<http://broadinstitute.github.io/picard>). This value, useful for the assembly software, was 125 nucleotides, which suggested partial overlapping of forward (R1) and reverse (R2) paired reads.

The transcriptome was built using the Velvet/Oases assembler (Schulz *et al.*, 2012) and 122,507 contigs (i.e. transcripts) were reconstructed. To reduce redundancy and potential assembly errors, the CAP3 software (Huang and Madan, 1999) was selected to collapse contigs identical for more than 70% in a single sequence. The final transcriptome resulted in 71,648 transcripts, with an average length of 1,331 nucleotides. In table 2.2 the major statistics of the *de novo* assembly are recorded. The transcripts ranged in size between the minimum set threshold of 100 bp and 12,009 bp with about 23,000 transcripts that were between 500 and 1000 bp in length (figure 2.5).

**Table 2.2.** Statistics on the *de novo* assembled *C. pepo* transcriptome.

Total # transcripts	71,648
Total # gene locus	42,517
# single sequence	22,594
# multiple variants	19,923
Total sequence length (nt)	95,354,115
average transcript length (nt)	1,331
maximum transcript length (nt)	12,009
minimum transcript length (nt)	100
median transcript length (nt)	1,084

To evaluate the quality of transcriptome assembly, ESTScan (Iseli *et al.*, 1999) was used to find coding regions within assembled transcripts. Results indicated that 67,534 sequences (about 94% of total transcripts) contain putative coding sequences that could be translated into proteins. Among these, 23,735 transcripts were categorized as complete ORF, containing defined start and stop codons. Additionally, 43,799 transcripts were classified as partial coding sequences. Specifically 25,000 sequences were classified as “5’ truncated ORF” with clear stop codon and lacking the ATG start codon; 8,220 transcripts displayed the initiating ATG codon but not termination triplet. Furthermore, 10,579 sequences encoded for truncated proteins showing neither start nor stop codons. The remaining 4,114 sequences (about 6% of all transcripts) were probably UTRs with interspersed stop codons or long non-coding RNAs.



**Figure 2.5.** Size distribution of *Cucurbita pepo* assembled transcripts.

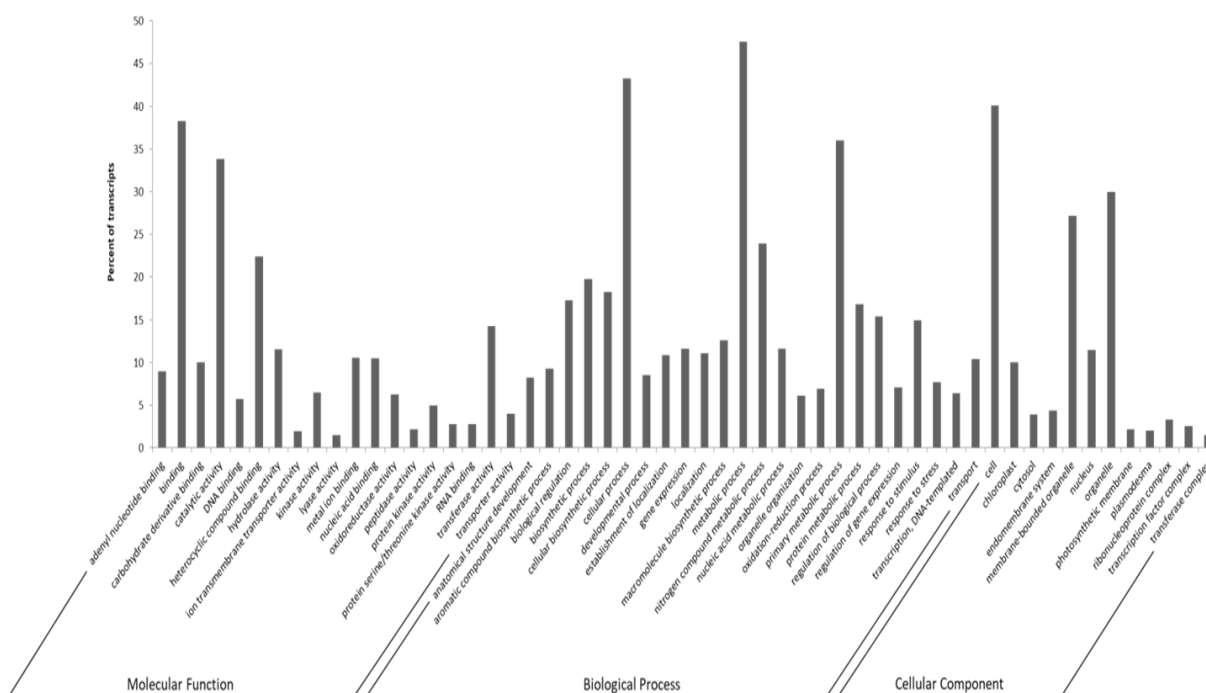
### 2.2.3 Functional annotation of *C. pepo* transcriptome

To predict and attach meaningful biological information to each transcript, similarity-based searches were performed. BLASTx analyses were carried out against melon, cucumber and Arabidopsis protein complement as well as against UniProtKB/Swiss-Prot, with a cut-off E-value of  $1e^{-5}$ . Comparing the assembled transcriptome against the TAIR10, containing complete protein sequences for the model plant *Arabidopsis thaliana*, 56,683 transcripts, corresponding to 79% of transcriptome, displayed significant BLAST hits. A total of 58,937 significant BLAST matches (72% of total transcripts) was obtained when *Cucumis melo* protein dataset (version 3.5) was used as reference database. BLAST searches against *Cucumis sativus* protein sequences (version 1.0) and the UniProt database (release 2012\_02) revealed that 48,303 (67%) and 48,597 (68%) sequences presented at least one significant match, respectively. Based on these results, approximately 71% of all transcripts presented at least one match in one of the four databases queried. Furthermore, a total of 38,087 sequences, about 53% of total, matched all four protein databases.

To functionally classify *C. pepo* transcripts in a standard and controlled vocabulary, gene ontology (GO) terms and Enzyme Commission numbers (EC numbers) were assigned to each sequence using the Blast2GO suite (Götz *et al.*, 2008). Gene ontology (GO) terms were assigned to 51,398 sequences. The number of GO terms per sequence varied between 1 and 74, with an average of seven GO terms per transcript. *C. pepo* transcripts were classified in three main GO categories: biological process, molecular function and cellular component. In total 276,601 GO terms were retrieved, with 50% assigned to biological process, 27% assigned to molecular function and 21% assigned to cellular component. The distribution of GO terms is

reported in figure 2.6. As for biological process classification highly represented categories were metabolic process (GO: 0008152) and cellular process (GO: 0009987) with respectively 47.5% and 43.2% of transcripts. Molecular functions were mainly assigned to binding (GO: 0005488; 27,544 sequences-38.2%) and catalytic activity (GO: 0003824; 24,337 sequences-33.8%), whereas many genes were assigned to cell (GO: 0005623; 40.1%) and organelle (GO: 0043226; 29.9%) in case of cellular component domain. Based on Blast2GO results, EC numbers were assigned only to 15,304 transcripts out of 51,398 GO annotated sequences. EC number is a numerical classification method used for enzymes, based on the chemical reaction they catalyse.

To survey genes involved in important pathways, annotated transcripts were also mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using Blast2GO. As a result, 10,426 sequences were mapped at least in one KEGG pathway. As expected, the pathways with the higher number of sequences were mainly related to cellular organization and primary metabolic processes. Among these, the most representative were Starch and sucrose metabolism (1087 members); Oxidative phosphorylation (326 members); Amino acid metabolism (592 members) and Fatty acid metabolism (169 members).



**Figure 2.6.** Bar chart describing the distribution of *Cucurbita pepo* transcripts into GO categories. Transcripts were annotated in three domains: cellular components, molecular functions, biological process. The y-axis indicates the percentage of sequences in a given category.

A BLASTn analysis was performed to compare the *de novo* assembled transcriptome with the *C. pepo* transcriptome available on-line (version 3.0) and identify novel transcripts in the aphid-challenged transcriptome. The comparison resulted in the identification of 1,313 transcripts with no match against the publically available transcriptome. Moreover, 548 out of 1,313 sequences presented at least one match with one of the four protein databases used for BLASTx. Through the information obtained from BLASTx and Blast2GO analyses, it was possible to describe the function of 504 novel transcripts. Among the novel transcripts in our transcriptome



several genes involved in protein metabolism and translation were identified. Specifically, 65 transcripts were annotated as “ribosomal protein”, coding for several rRNA belonging to both large (60S) and small (40S) ribosome subunits. Moreover, 9 chaperonin protein TCP-1/cpn60, 3 DNAJ protein and 5 Heat shock proteins (Hsp) were listed: 2 Hsp70 and 3 Hsp90. Genes coding for *Translation initiation factors* and *Elongation factors* were also present. Transcripts putatively involved in ubiquitination were identified such as six *Zinc finger, C3HC4 type (RING finger) family protein*, *Ubiquitin-conjugating enzyme E2*, *RING/U-box superfamily protein* and *Proteasome core complex*. Genes coding for kinases involved in DNA damage response were also identified. These sequences, coding for 5 *Ataxia telangiectasia-mutated (ATM) and RAD3-related (ATR) proteins*, are serine/threonine-protein kinases. Five sequences annotated as *LRR receptor-like serine/threonine-protein kinases (LRR-RLK)*, potentially involved in plant cell signalling, were listed. Other transcripts involved in Ca<sup>2+</sup> cell signalling processes were identified (five *Calmodulin protein*, one *Calmodulin-like protein* and one *Calcium-dependent protein kinase*). Among novel transcripts present in our transcriptome, sequences putatively involved in plant stress perception and response were described. Transcripts involved in oxidative stress were annotated such as four *Thioredoxins*, three *Glutaredoxins* and four *Glutathione S-transferases (GSTs)*. Six genes coding for *Cytochrome P450* family members were listed. Finally, a transcript annotated as *Probable WRKY transcription factor 72* was included among *C. pepo* transcripts assembled following aphid infestation.

## 2.3 Discussion

### 2.3.1 De novo transcriptome assembly

Transcriptome sequencing allows for functional genomic studies for the organism under investigation. Although several high-throughput technologies have been developed for rapid sequencing and characterization of transcriptomes, expressed sequence data are still not available for many organisms, including crop plants. Next generation sequencing technologies provide a low cost, labour saving and rapid mean for transcriptome sequencing and characterization (Morozova *et al.*, 2009).

Similar to sequencing technologies, many bioinformatic tools have also been developed for the short-read transcriptome sequence data assembly and analysis (Zerbino and Birney, 2008; Grabherr *et al.*, 2011; Robertson *et al.*, 2010).

In this study, a strategy based on the Velvet/Oases assembler was adopted for *de novo* assembly of transcriptome using short reads.

The final transcriptome assembly from aphid-challenged “San Pasquale” leaves resulted in 71,648 transcripts with a total assembly length of 95,354,115 bp and an average length of 1,331 bp. Furthermore, the assembly contained a substantial number of large transcripts with sequence length >500 bp (~76%, 55,140 transcripts; figure 2.5), which was comparable to the results obtained by previous studies with a deep sequencing method for transcriptome generation (D’Agostino *et al.*, 2013; Wu *et al.*, 2014; Wyatt *et al.*, 2015; Sudheesh *et al.*, 2016).

The *C. pepo* transcriptome generated in this study was compared with the previously published zucchini transcriptome sequenced from root, leaf and flower tissue (Blanca *et al.*, 2011). The transcriptome released by Blanca and colleagues was assembled from sequences derived from two contrasting *C. pepo* cultivars (a scallop-type and a zucchini-type squash) and consisted of 49,610 unigenes with an average length of 626 bp. The aphid-challenged leaf transcriptome had a similar number of gene *loci*

(42,517) as the previous transcriptome, though it had a longer average transcript length. The greater sequencing depth of our transcriptome may be responsible for the assembly of more full-length transcripts. These results are also highly comparable to those of other squashes, such as the transcriptome of the *C. pepo* cultivar “Big Moose”, which consisted of 84,727 total transcripts and a total sequence length of 88,473,202 bp (Xanthopoulou *et al.*, 2016). Conversely, statistics obtained for our reference transcriptome were higher than that of the transcriptome of the pumpkin cultivar “Munchkin”, which has a total of 70,574,057 bp, and higher than that of the transcriptome generated from acorn squash, which has 73,559,618 bp (Xanthopoulou *et al.*, 2016; Wyatt *et al.*, 2015). However, the average length statistic of the assembly was highly comparable to that for acorn squash, with an average length of 1,315 bp.

### 2.3.2 Transcriptome annotation and novel transcripts identification

As expected, a large number (67-72%) of assembled transcripts showed significant similarity with cucurbits (*C. sativus* and *C. melo*) at protein level. Certainly, this results are consistent with known phylogenetic relationships existing among these species. The similarity-based searches performed against the Arabidopsis protein database (TAIR10) reveals that the majority of transcripts had significant hits with the queried dataset (79%). Overall, approximately 70% of all transcripts had at least one significant hit. These results are comparable to those obtained for *C. pepo* transcriptomes previously assembled (Blanca *et al.*, 2011; Wyatt *et al.*, 2015).

GO analysis revealed that a total of 72% transcripts from “San Pasquale” were associated with at least one GO term which is higher than that assigned to other *C. pepo* transcriptomes (~60%). However, the distribution of annotated transcripts under different GO categories showed a concentration in 4-10, 3-7 and 4-8 levels respectively for biological process, molecular function and cellular component, indicating a good accuracy of the annotation.

Finally, to more closely compare the *de novo* assembled aphid-challenged leaf transcriptome with the first *C. pepo* transcriptome published a BLASTn search was performed. The 98% (70,335) of our transcripts were significantly similar to the unigenes released by Blanca and colleagues (2011), while 1,313 sequences had no match. Hence, the comparison revealed the presence of novel transcripts. Looking at the function of these novel transcripts, several genes involved in protein metabolism were annotated. They encode for ribosomal proteins, chaperonins, and proteins of ubiquitin-proteasome complex. Moreover, transcripts annotated as heat shock proteins were present. These proteins, in addition to their role in protein-folding processes, could also have a role in signal transduction and protein degradation and trafficking (Wang *et al.*, 2004). The increase of protein turnover, which includes both protein biosynthesis and ubiquitination, is considered a metabolic adaptation to environmental cues (Saibo *et al.*, 2008).

Transcripts that code for plant receptors (e.g. LRR-RLK) were identified among the novel transcripts. These are involved in plant signal perception and cell signalling. RLKs participate in a diverse range of processes, including regulation of development, disease resistance, and hormone perception (Shiu and Bleecker, 2001). Moreover, transcripts related to oxidative stress were listed as novel ones. Among these *Glutathione S-transferases* (GSTs) are best known for the detoxification of xenobiotics, but they can also act as antioxidants by tagging oxidative degradation products (especially from fatty acids and nucleic acids) for

removal or by acting as a glutathione peroxidase to directly scavenge peroxides (Dalton *et al.*, 2009).

One gene coding for a *Probable WRKY transcription factor 72* was also annotated. WRKY transcription factors play a central role in transcriptional reprogramming associated with plant immune responses. The WRKY72-type transcription factors are implicated in basal defence in tomato and Arabidopsis, a function that has been recruited to serve *Mi-1*-dependent immunity (Bhattarai *et al.*, 2010).

## 2.4 Conclusions

The sequencing and computational strategies adopted allowed to obtain a well-structured reference transcriptome for the cultivar "San Pasquale". This resource was generated and characterised with the aim of improving the knowledge on expressed genes modulated following aphid infestation. Identification of novel transcripts involved in stress response mechanisms can give an added value to this resource. The transcriptome will be useful for the development of novel tools for aphid control as well as for future marker-assisted selection strategies in *C. pepo* breeding programs.

## 2.5 Materials and Methods

### 2.5.1 Biological materials

Seeds of the aphid-susceptible *Cucurbita pepo* cultivar "San Pasquale" were obtained from the seed company "La Semiorto Sementi", in the frame of the project named "GenHORT—adding value to elite Campania horticultural crops by advanced genomic technologies (PON02\_00395\_3215002)" (OR1: Qualità e sostenibilità delle produzioni mediante strumenti di genomica strutturale e funzionale). Zucchini plants were sown in plastic pots with 10 cm diameter and were enclosed in cages equipped with an anti-insect net (50 mesh). Plants were grown in a dedicated climatic chamber under a 16 hours day cycle at a temperature of  $22 \pm 1$  °C, and with a relative humidity of  $75 \pm 5$  % as environmental settings.

Melon/cotton aphid *Aphis gossypii* Glover (Homoptera: Aphididae) was obtained from a population infesting watermelon in Terracina (Latina, Central Italy) and reared on "San Pasquale" plants in cages provided with an anti-insect net. Aphid rearing was maintained at the Department of Agricultural Sciences, University of Naples Federico II in a dedicated climatic chamber, using the same environmental conditions previously reported (temperature  $22 \pm 1$  °C; RH  $75 \pm 5$  %; photoperiod L16:D8).

### 2.5.2 Aphid infestation bioassays and plant material collection

Zucchini plants were transferred in a clean climatic chamber (Temperature:  $22 \pm 1$  °C; Relative Humidity:  $75 \pm 5$  %; Photoperiod: L16:D8), and were individually placed in insect-proof cages for the following infestation assay (figure 2.7).



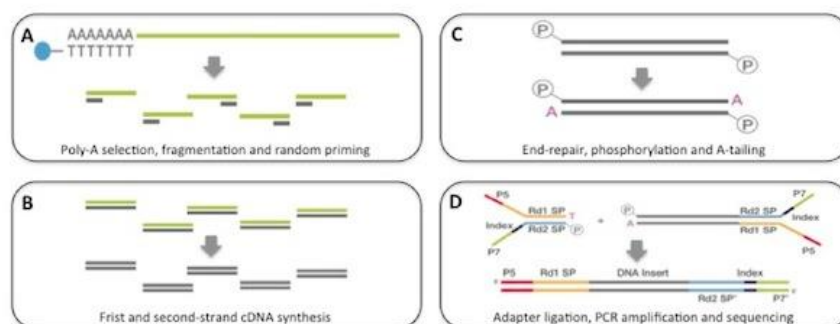
**Figure 2.7.** Zucchini plants individually arranged in insect-proof cages for aphid infestation assay.

First and second leaves were infested with ten *A. gossypii* adults. Five aphids per leaf were transferred onto adaxial surface with the help of a paintbrush, and their number was daily monitored. Control plants were grown under the same environmental conditions and moved in the same climatic chamber of the infested ones. Also control plants were individually enclosed in insect-proof cages. Aphids were left to feed for 24, 48 and 96 hours after that they were manually removed using a fine paintbrush. Leaf tissue was sampled and immediately frozen in liquid nitrogen. At the same time points, corresponding leaf tissues were sampled from aphid-free control plants. Three biological replicates were collected per time point for both infested and control plants, and leaves of a single replicate were pooled for downstream analysis.

### 2.5.3 RNA isolation and sequencing

Total RNA was isolated from 100 mg of tissue previously grinded in liquid nitrogen using the mi-RNeasy Mini kit (Qiagen), according to manufacturer's protocol. RNA samples were quantitatively and qualitatively analysed with NanoDrop ND-1000 UV-vis Spectrophotometer (NanoDrop Technologies) and Agilent 2100 Bioanalyzer (Agilent Technologies), respectively. Only samples characterized by a 260/280 nm absorbance higher than 1.8, a 260/230 nm absorbance higher than 2 and an RNA Integrity Number (R.I.N.) higher than 8 were used for RNA-sequencing. Next generation sequencing experiments were performed by Genomix4life S.R.L. (Baronissi, Salerno, Italy). Indexed libraries were prepared from 2  $\mu$ g of RNA with TruSeq Stranded mRNA Sample Prep Kit (Illumina) following the manufacturer's instructions schematically described in figure 2.8.

## Illumina Tru-Seq RNA-seq protocol



**Figure 2.8.** Schematic representation of Illumina library preparation. mRNA is poly-(A) selected using magnetic beads (A); mRNA is then fragmented and used for cDNA synthesis (B). cDNA molecules are phosphorylated at 5' end and A-tailed at 3' end (C). The library is PCR amplified using adapter oligos (D) before clustering and sequencing (<http://bitesizebio.com>).

Libraries were quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies) and pooled such that each index-tagged sample was present in equimolar amounts, with final concentration of the pooled samples of 2nM. The pooled samples were subjected to cluster generation and sequencing using an Illumina HiSeq 2500 System in a 2x101 paired-end format at a final concentration of 8pmol. The files generated contained the nucleotide sequence for each sample in FASTQ format. In paired-end experiments reads are typically split over two ordered files, one with the first-end and the other with the second.

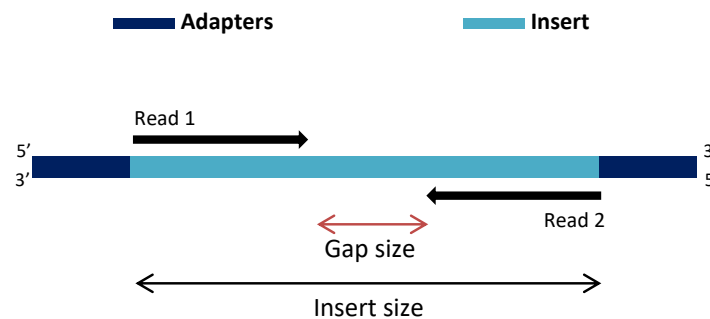
Raw sequence files (FASTQ format) were subjected to quality control analysis using the FastQC software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). It provides easy graphical outputs with information about sequence such as sequence quality score, sequence GC content and sequence length distribution.

### 2.5.4 Data analysis

#### 2.5.4.1 Pre-processing of raw reads and de novo transcriptome assembly

After Illumina sequencing, raw reads were pre-processed using the "fastq\_quality\_filter" from the FASTX Toolkit (Gordon and Hannon, 2010) and the Trimmomatic software (Bolger *et al.*, 2014) to perform a quality score based filtering and Illumina adapters trimming, respectively. Briefly, "fastq\_quality\_filter" tool was used to remove reads with Phred Quality Score per base value lower than 30 ( $\leq$  Q30) for 80% of read length. Phred quality scores (or Q score) are defined as  $Q =$

$-10\log_{10}(P)$ , where  $P$  is the probability of the base call being incorrect (Cock *et al.*, 2009). This cut-off allowed to select reads for which the chances that each base was called incorrectly was 1 in 1000. Trimmomatic, thanks to “ILLUMINACLIP” function, scans each read from the 5' end to the 3' end to determine if any of the user-provided adapters are present and to trim them. Using the “HEADCROP” function was possible to cut unassigned bases (N bases) at the 5' or 3' end, if present. Moreover, all reads shorter than 75 nucleotides at the end of quality control and trimming steps were removed. All high-quality reads were combined into a single dataset for *de novo* transcriptome assembly using Velvet/Oases (Schulz *et al.*, 2012). Before assembly, Picard package (<http://broadinstitute.github.io/picard>), and specifically Picard-Collect Insert Size Metrics tool, was used to calculate *insert size* value and distribution in our paired-end libraries. *Insert size* refers to average length of fragments selected for sequencing experiment, excluding Illumina adapters size (figure 2.9).



**Figure 2.9.** Schematic representation of *insert size*. Gap size is identified by the distance, in nucleotides, between Read 1 (forward read) and Read 2 (reverse read) generated during paired-end sequencing.

Velvet assembler was run using the following *k-mers*: 65, 67, 69, 71 and 73. The *hash tables* generated for each *k-mer* were used to build de Bruijn graphs which were then organized into a scaffold, divided into *loci* and finally analysed to extract transcript assemblies or transfrags. Once all the individual *k-mer* assemblies were finished, they were merged into a final assembly using Oases. Then, the resulting contigs were fed into CAP3 (Huang and Madan, 1999). Thanks to the CAP3 software all contigs which showed equal or more than 70% of identity, were collapsed into a single sequence, thus eliminating the redundancy present within the assembled transcriptome.

#### 2.5.4.2 ORF identification and functional annotation

*De novo* assembled transcriptome quality assessment was performed using different approaches. ESTScan (Iseli *et al.*, 1999) was used for the identification of potential protein-coding regions (Open Reading Frames, ORFs) in the *C. pepo* transcriptome using *Arabidopsis thaliana* training matrix for peptide prediction. The software can correct any sequencing errors that may lead to incorrect stop codons creation, and it is also able to identify sequence insertions and deletions that can lead to an incorrect reading frame prediction. Automated functional annotation of assembled transcripts was carried out using the BLASTALL package (release 2.2.25; Altschul *et al.*, 1990). Similarity-based search (E-value <  $1e^{-5}$ ) was performed against *Cucumis melo* (version 3.5; [https://melonomics.net/files/Genome/Melon\\_genome\\_v3.5\\_Garcia-Mas\\_et\\_al\\_2012/](https://melonomics.net/files/Genome/Melon_genome_v3.5_Garcia-Mas_et_al_2012/); Garcia-Mas *et al.*, 2012), *Cucumis sativus* (version 1.0; <http://genome.igi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=Phytozome>) and the *Arabidopsis thaliana* (version TAIR 10; <https://www.arabidopsis.org/>) protein dataset (BLASTX), UniProtKB/Swiss-Prot database (<http://www.uniprot.org/downloads>; release 2012\_02) (BLASTX) and *C. pepo* draft transcriptome (version 3.0; [https://cucurbigene.upv.es/db/transcriptome\\_v3/](https://cucurbigene.upv.es/db/transcriptome_v3/)) (BLASTN). BLAST-formatdb tool was used to format protein and nucleotide source databases before running BLASTALL.

Functional annotation was refined using the Blast2GO suite (Götz, *et al.* 2008). Briefly, for each sequence a BLASTX similarity search (E-value <  $1e^{-3}$ ; word size: 6) against the *nr* NCBI database (Non-redundant GenBank CDS translations including RefSeq, PDB, SwissProt, IR and PRF) was performed to retrieve a maximum of 20 top homologous hits per query. The GO-term mapping was obtained from gene identifiers using NCBI as well as non-redundant reference protein database (PSD, UniProt, Swiss-Prot, TrEMBL, RefSeq, GenPept, PDB Full Gene Ontology DB) using as threshold an e-value lower than  $1e^{-6}$ . The selection of GO terms from the GO pool obtained by the mapping step was performed at default parameters. Additional annotations (e.g. the recovery of implicit “Biological

Process” and “Cellular Component” GO-terms from “Molecular Function” annotations) were implemented using ANNEX 5.0. Completion of the functional annotation with protein domain information was performed with InterProScan 5.0. Sequencing with a BLAST hit that could not be annotated were then blasted (BLASTX) against the *Arabidopsis thaliana* protein sequences, and if not annotated, against the *Oryza sativa* protein sequences, the maize database and finally the SwissProt database. After removal of top-level annotations, a functionally-based sequence labelling was performed using the Blast Descriptor Annotator (DBA). EC numbers was assigned based on the Blast2GO results. To determine metabolic pathways, Kyoto Encyclopedia of Genes and Genomes (KEGG) mapping was used.

Chapter 3

*Identification of  
zucchini (Cucurbita  
pepo L.) differentially  
expressed genes in  
response to Aphis  
gossypii feeding*

Alessia Vitiello in collaboration with  
Nunzio D'Agostino, Maria Cristina Digilio,  
Mariangela Coppola and Rosa Rao





## Abstract

Aphids are among the most destructive pest for cultivated crops. The melon aphid *Aphis gossypii* (Homoptera: Aphididae) is a serious pest for different plant species, including zucchini. *Cucurbita pepo*, widely cultivated in temperate regions, was selected for this study because of its growing economic importance among *Cucurbita* species. The aim of this study is to investigate, through a time-course transcriptomic analysis based on RNA-seq, the transcriptional reprogramming of zucchini plants (cv. “San Pasquale”) during *A. gossypii* infestation. Zucchini plants were infested with ten adult aphids. Leaves were collected from infested and un-infested (control) plants after 24, 48 and 96 hours. Total RNA was extracted from each sample and sequenced using the Illumina HiSeq 2500 platform. Following quality assessment and read pre-processing, short paired-end reads were aligned versus the *de novo* assembled zucchini transcriptome (Chapter 2). To ensure that each gene was represented only once in the dataset, we filtered out the longest transcript per gene *locus*, obtaining a dataset of 42,517 sequences, we used as reference for read mapping and differentially expressed gene (DEG) identification. The filtering criteria used for DEG call were: a log Fold change in expression greater than 2 and a FDR < 0.05. Considering the three time points, a total of 766 transcripts was differentially expressed. After 24 hours, 158 transcripts (115 up and 45 down) were influenced by aphid infestation. The number of affected transcripts increased to 565 after 48 hours (420 up and 145 down) and declined to 179 transcripts (62 up and 117 down) after 96 hours from the infestation. The analysis of DEGs involved in hormone-related defence pathways, showed that SA-signalling and SA-related genes play a dominant role in “San Pasquale”-*A. gossypii* interaction. In addition, aphid dispersion behaviour was observed on zucchini leaves pre-treated with methyl salicylate (MeSA). Zucchini plant response was also characterised by the overexpression of genes involved in primary metabolic processes as well as cell wall modification. Our study allows to elucidate, for the first time, the molecular mechanisms activated by zucchini plants in response to aphid infestation during early stages of a compatible interaction.

## 3.1 Introduction

Plant-aphid interaction is a dynamic system subjected to continual variation and change (Mello and Silva-Filho, 2002). Actually, aphids are an example of “stealthy” pest because they are adapted to feed on phloem sap and produce a small mechanical damage, compared with chewing herbivores. However, short generation time can result in severe increase in aphid population and elevated depletion of phloem sap (Kuśnierczyk *et al.*, 2008). After landing on a plant, the first activity of aphids is to determine whether the plant is suitable for them or not. During probing phase aphids’ stylets transiently puncture the epidermis, mesophyll, and parenchyma cells to the phloem, and this mechanical damage, combined to aphid effectors introduced with oral secretions, may influence the activation of plant responses to infestation (Morkunas and Gabryś, 2011). Constant development of RNA-sequencing technologies enabled the identification of specific genes in a target sample and provided qualitative and quantitative description of gene expression. At present, these technologies represent an essential tool for understanding the molecular events occurring in plants following aphids infestation.

As described in Chapter 1, plant survival upon aphid attack relies on a complex protection strategy, based on recognition, cell signalling and defence response activation. Plants that carry a particular resistance gene (R gene) resulted in an

incompatible interaction (Thompson and Goggin, 2006). However, numerous studies focused on transcript profiles of compatible interactions indicated that events such as cell wall modification, protein phosphorylation, calcium flux, ROS generation and phytohormone changes take place in plants infested by aphids, leading to relevant transcriptional regulation in response to phloem-feeding insects (Liang *et al.*, 2015).

Plant transcriptional responses are partially determined by the coordinated regulation of SA and JA/ET signalling pathways, that can show both synergistic and antagonistic interactions. The SA-dependent cascades stimulate expression of genes associated with SA defence signalling pathway, including pathogenesis-related (PR) proteins, such as PR1, PR2, chitinases, and  $\beta$ -1,3-glucanases. Tomato plants susceptible to *Macrosiphum euphorbiae* showed an increase in PR1 gene expression (Coppola *et al.*, 2013), and *Rhopalosiphum padi* infestation elicited synthesis of  $\beta$ -glucanase, PR1 and thaumatin-like proteins in susceptible *Hordeum vulgare* plants (Delp *et al.*, 2009). Furthermore, Kaloshian (2004) reported that aphid-resistant tomato plants lost resistance when transformed with *NahG*, a gene encoding a SA-degrading enzyme.

One of the key regulatory elements in SA-dependent response is the activation of NPR1 (NON-EXPRESSOR OF PR GENE 1) protein. SA promote NPR1 deoligomerization into its active monomeric forms, which can interact with transcription factors (e.g. TGA, WRKY) and activate PR gene expression and subsequent defence response (Verma *et al.*, 2016). These findings suggest that SA plays a role in activation of plant defence.

Aphid-induced methyl salicylate (MeSA), a volatile compound derived from SA, has been reported as a strong aphid repellent that may deter aphids from settling on plants with already high aphid densities as in *A. fabae* on broad bean plants (Hardie *et al.*, 1994) and *R. padi* on barley (Glinwood *et al.*, 2009). Digilio and colleagues (2012) reported that at 24 h after the introduction of *M. euphorbiae* in a cage containing a tomato plant pre-treated with MeSA, an alteration of aphid acceptance behaviour occurred. A low number of aphids with their stylets inserted was present on treated tomato plants compared with controls, indicating a negative effect of synthetic methyl salicylate on aphid fixing behaviour (Digilio *et al.*, 2012).

JA-responsive gene expression is mainly mediated by transcription factors such as JASMONATE INSENSITIVE 1/MYC2 (JIN1/MYC2), and several members of APETALA2/ETHYLENE-RESPONSIVE FACTOR (AP2/ERF) family. The activation of JA-pathway can show a significant effect reducing aphid population, as reported for *Arabidopsis* mutants *cev1*, in which JA and ET signalling is constitutively activated (Ellis *et al.*, 2002). By contrast, suppression of JA signalling in tomato mutant *jai1* had no effect on *M. euphorbiae* population (Thompson and Goggin, 2006). ET role in plant defence is mainly mediated by the transcription factor EIN3 (ETHYLENE INSENSITIVE3) which was suggested to induce *ERF1* gene expression and mediate ET-signalling pathway (Solano *et al.*, 1998). However, compared with chewing pest and mechanical wounding, aphids had a weak influence on JA- and ET-related gene expression. In general, SA antagonizes JA-induced pathway during plant-aphid interaction, whereas ET can have both positive and negative effects to achieve tailored defence responses.

Few transcriptomic studies on cucurbit-phloem feeders interaction have been published so far. A research performed on melon (*Cucumis melo* L.) plants had the purpose to determine whether the ET pathway was induced by *A. gossypii* feeding and whether that induction differs in susceptible and resistant plants (Anstead *et al.*, 2010). The expression level of genes involved in ET synthesis (*ACS* and *ACO*),

perception and signal transduction (*ETR1*, *ETR2*, *ERS1*, *EIN2*, *EIN3*, *EIL1*, *ERF1*), and downstream response (*SSA-13*, *Type1-PI*, *SAG-21*), was analysed in susceptible (*Vat*) and aphid-resistant (*Vat<sup>t</sup>*) melon plants using Real-Time PCR after aphid treatments. Evidence of a stronger induction of ET pathway was highlighted in resistant melon, when compared with susceptible plant. In particular, the authors reported that ET signalling pathway and responsive genes were highly up-regulated in resistant plants, indicating ET as player in *Vat*-mediated host-plant resistance. Moreover, the strong up-regulation of *ERF1* gene, coding for a transcription factor involved in activation of JA pathway, during incompatible interaction between *A. gossypii* and *Vat<sup>t</sup>*, may indicate an ET-dependent activation of the JA pathway (Anstead *et al.*, 2010). Jasmonic acid-ethylene synergism has also been observed in induction of defence response in *Cucurbita moschata* Duchesne to feeding by the silver leaf whitefly *Bemisia argentifolii* Bellows and Perring (van de Ven *et al.*, 2000). Expression levels of two genes, *SLW1* and *SLW3*, usually expressed in silver leaf whitefly-infested leaves, were monitored. Exogenous treatments with wound and defence signals, such as MeJA, and ethylene indicate a strong activation of *SLW1* gene expression and a possible role of these products in defence against the leaf-silvering disorder in squash (van de Ven *et al.*, 2000). Liang and colleagues (2015) described transcriptional reprogramming of cucumber plants (*Cucumis sativus* L.) infested by *A. gossypii* to identify genes associated with resistance to aphid-induced damage. The aphid-resistant cucumber cultivar “EP6392” was used to monitor, via Illumina-based RNA-Seq, responses to aphid infestation. A total of 49 differentially expressed genes belonging to processes such as signal transduction, plant-pathogen interaction and sugar metabolism, were putatively found involved in cucumber aphid resistance (Liang *et al.*, 2015).

Here, we present the first zucchini gene expression profiling analysis of compatible *C. pepo*-*A. gossypii* interaction using an RNA-Seq approach. Description of transcriptional reprogramming occurred in *C. pepo* plants following *A. gossypii* infestation represents an important step to define the defensive capacities of plants against phloem feeding insects. Moreover, understanding of the molecular mechanism activated following aphid infestation is the most effective way to decrease aphid damage, reduce pesticide use, and produce higher-quality fruits. (Liang *et al.*, 2015).

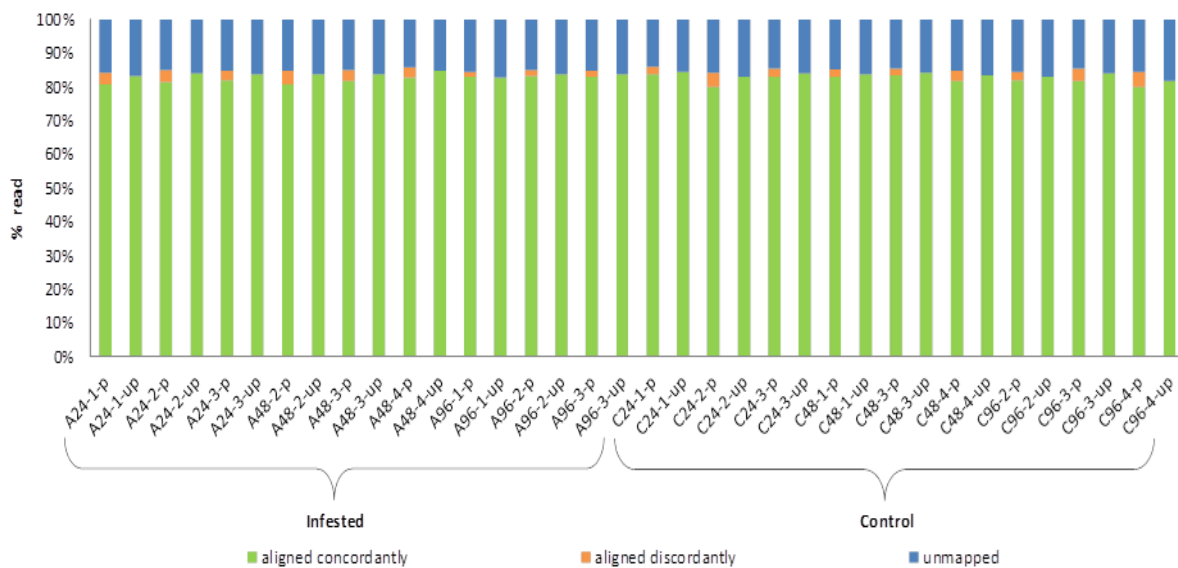
## 3.2 Results

The zucchini transcriptome assembled *de novo* in this study (Chapter 2) was used as high quality reference for read alignment and for the identification of differentially expressed genes influenced by aphid feeding.

### 3.2.1 High quality reads alignment to the reference transcriptome

A subset of the *C. pepo* transcriptome was used as reference for read mapping. This subset includes 42,517 transcripts, namely all the transcripts expressed from a single gene *locus* (for those *loci* that express multiple transcript isoforms the longest transcript was considered). This strategy was adopted to reduce the number of reads with multiple matches on the reference. Read alignment was performed using Bowtie2 (Langmead and Salzberg, 2012). As reported in figure 3.1, paired and unpaired reads, for each sample, were independently aligned to the reference. About 84% of high quality reads globally aligned on reference. As for paired-end reads, ~2.9% of reads aligned discordantly, while ~82.7% aligned concordantly. A pair of reads that aligns with the proper relative mate orientation and with the correct

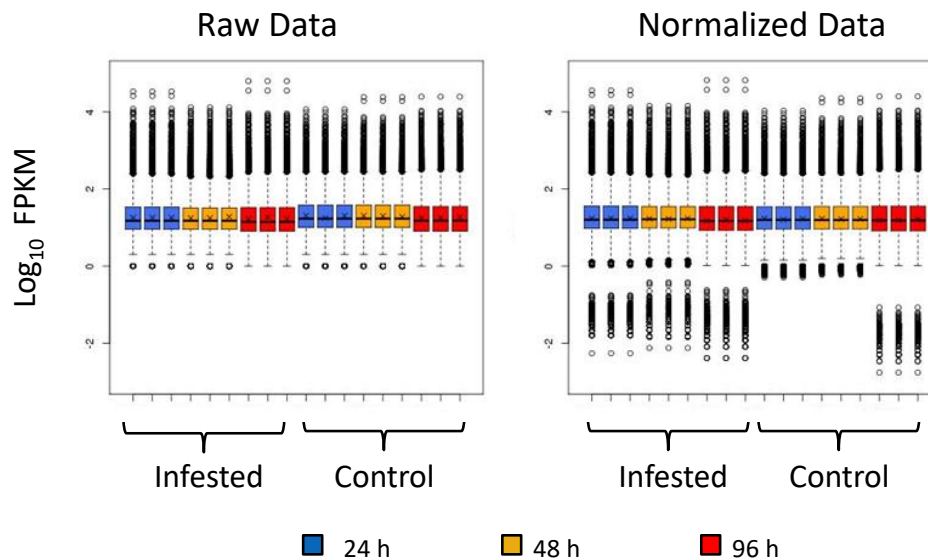
distance between mates is said to align concordantly. In case paired-end reads have unique alignments, but the alignments do not match paired-end expectations (i.e. unexpected relative orientation, unexpected distance range, or both), mate pairs are said to align discordantly. Finally, ~16% of reads did not find any match.



**Figure 3.1.** Summary of read-to-reference alignment results. *De novo* *C. pepo* transcriptome was used as reference. “\*-p” indicates files containing paired reads; “\*-up” indicates files containing unpaired reads.

Read mapping represents the first step before gene expression profiling analysis. Read summarization was performed using the eXpress software that allowed us to count how many reads map per transcript (raw read count). eXpress was selected thanks to its ability to work with more than one alignment file at the same time and to manage multiple-matches. An additional filtering step was performed on eXpress output before the downstream normalization phase. Transcripts represented by low read count were removed from the list. Such filter reduced the number of expressed transcripts to 13,956.

Inter-sample normalization was carried out adopting the *R* package edgeR. The Trimmed Mean of M values (TMM) method considers RNA library size to computing a scale factor which minimizes gene expression differences between samples. Box-plots displaying read count distribution before and after TMM normalization are reported in figure 3.2. The graphs allowed to identify biological replicates for each sample with dispersion values higher than the estimates, which could lead to identification of false-positives.  $\text{Log}_{10}$ FPKM values showed a highly conserved distribution both before and after normalization.



**Figure 3.2.** Box plot of read count distribution before (Raw data) and after read normalization (Normalized data). Biological replicates are shown as boxes with the same colour. The black line reported inside the boxes represents the median and the X represents the average read count value.

Moreover, pairwise comparison of gene expression values between replicates at each experimental point was performed by formulating *scatter plots*. For each pairwise comparison Person coefficient ( $r^2$ ) was calculated using the *cor()* function in *R* package (figure 3.3 and Appendix figure A.1). Correlation values were between 0.80 and 0.99 (table 3.1).

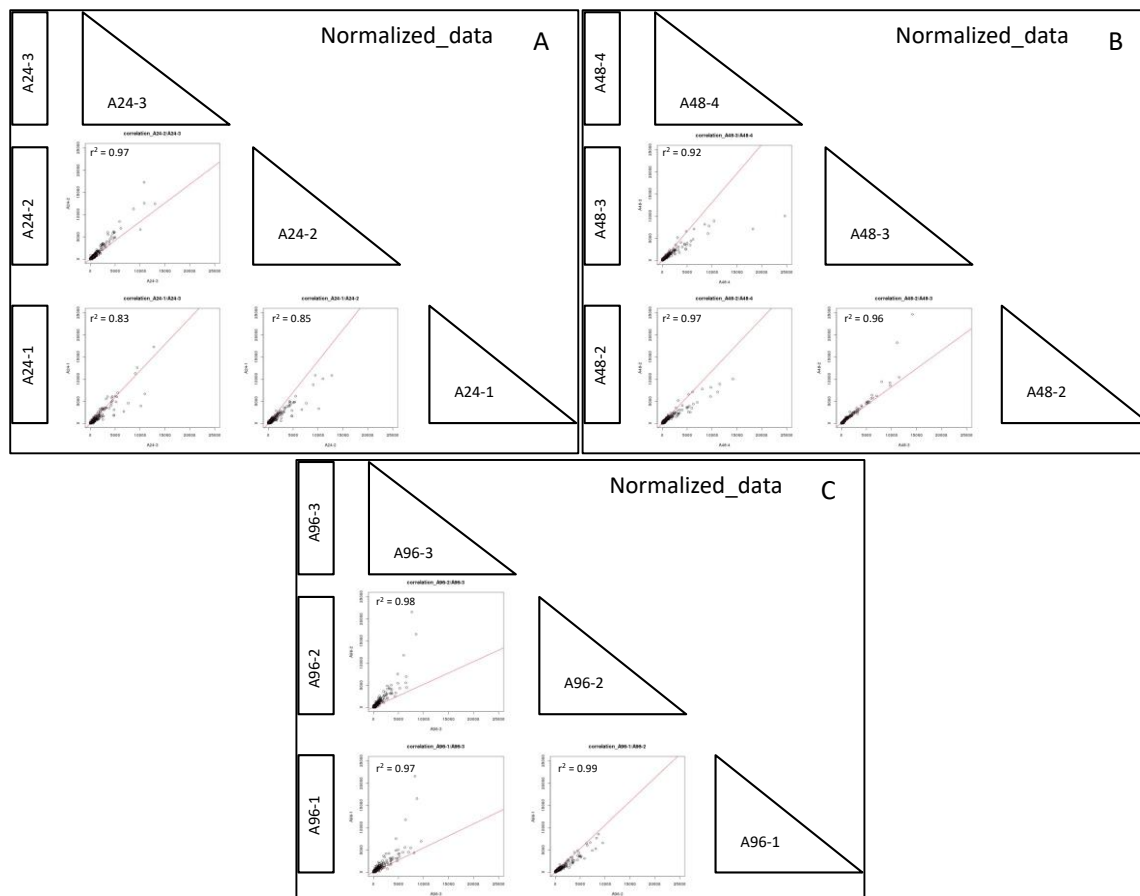
After summarization, filtering and normalization processes, the list of 13,956 transcripts was analysed for the identification of genes affected by *A. gossypii*.

**Table 3.1.** Comparisons of genes expression values of the three biological replicates obtained at each sampling time.

Condition	Time (hpi)	Comparison	$r^2$ **
Infested	24	A24-1 vs A24-3	0.83
		A24-1 vs A24-2	0.85
		A24-2 vs A24-3	0.97
	48	A48 -2 vs A48-3	0.96
		A48 -2 vs A48-4	0.97
		A48-3 vs A48-4	0.92
	96	A96-1 vs A96-2	0.99
		A96-1 vs A96-3	0.97
		A96-2 vs A96-3	0.98
Control	24	C24-1 vs C24-2	0.92
		C24-1 vs C24-3	0.91
		C24-2 vs C24-3	0.96
	48	C48-1 vs C48-3	0.97
		C48-1 vs C48-4	0.99
		C48-3 vs C48-4	0.97
	96	C96-2 vs C96-3	0.80
		C96-2 vs C96-4	0.95
		C96-3 vs C96-4	0.89

\*hours post infestation

\*\*Pearson coefficient values



**Figure 3.3.** Scatter plots and correlation coefficients ( $r^2$ ) of normalized read count values of biological replicates of infested samples collected at 24 hpi (A), 48 hpi (B) and 96 hpi (C).

### 3.2.2 Identification of differentially expressed genes during aphid infestation

To highlight the variation in gene expression in zucchini plants after the compatible interaction with *A. gossypii*, the *R* package edgeR (Robinson et al., 2010) was used to compare samples collected at 24, 48 and 96 hours post infestation (hpi) with uninfested plants collected at the same time points. Considering whole time course, 902 transcripts were identified as differentially expressed using as threshold a two-fold change in transcript level and an FDR value lower than 0.05. DEG lists were compared to identify common transcripts among the three tested conditions. A total of 766 transcripts were listed after duplicate removal, and more than 60% of these genes were up-regulated. In table 3.2 it is reported the number of genes affected by infestation in our experimental conditions. Specifically, the data indicated that at 24 hpi a total of 158 transcripts were influenced by aphid feeding. The number of affected transcripts increased to a total of 565 after 48 hours and declined to 179 transcripts at 96 hpi.

**Table 3.2.** *C. pepo* differentially expressed genes in response to aphid at each experimental point.

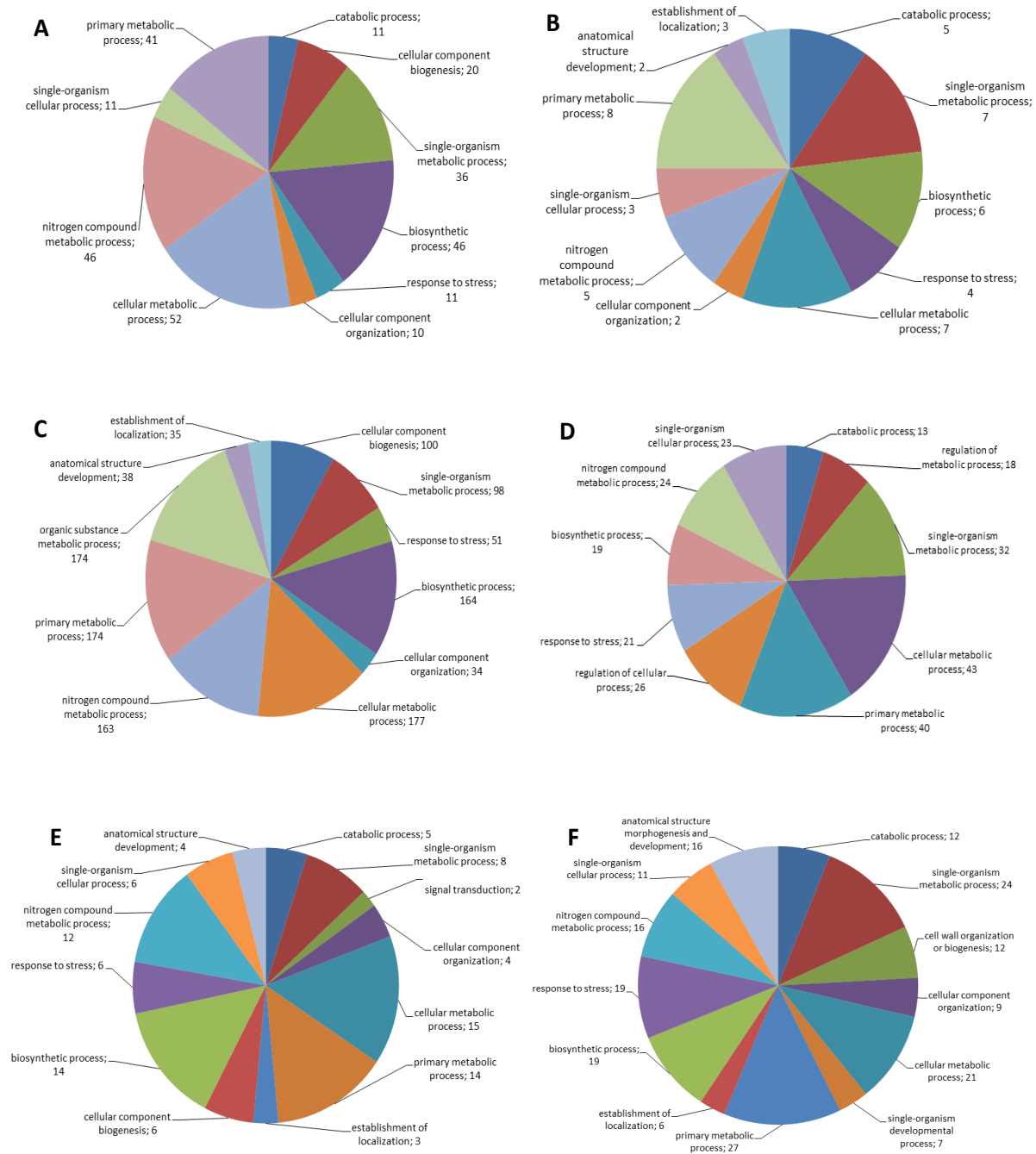
Time (hpi)*	# transcripts		
	Up-regulated	Down-regulated	Total
24	115	43	158
48	420	145	565
96	62	117	179

\*hours post infestation

Up and down-regulated genes, identified at each experimental point, were listed separately in order to associate them with the available annotation for easy analysis. Through this approach it was possible to assign a GO term to 109 (86 up and 23 down), 442 (355 up and 87 down) and 118 (38 up and 80 down) transcripts affected by aphid feeding, respectively at 24, 48 and 96 hpi. The lists of differentially expressed genes identified at each time, including their expression levels, FDR,  $p$ -values and the description of gene function are reported in Appendix table A.2\_A, A.2\_B and A.2\_C.

Analysis of GO terms successfully associated with DEGs indicated the transcriptional reconfiguration of several biological processes. Using Blast2GO it was possible to build combined graphs highlighting the major processes influenced by *Aphis gossypii* feeding activity after the GO-slim analysis. GO-slim allows to obtain cut-down versions of the GO ontologies that contain a subset of terms and offer an overview of the ontology content. Level 3 pie charts of GO Biological Process domain were built for both up-regulated and down-regulated sequences at each time point and are shown in figure 3.4.

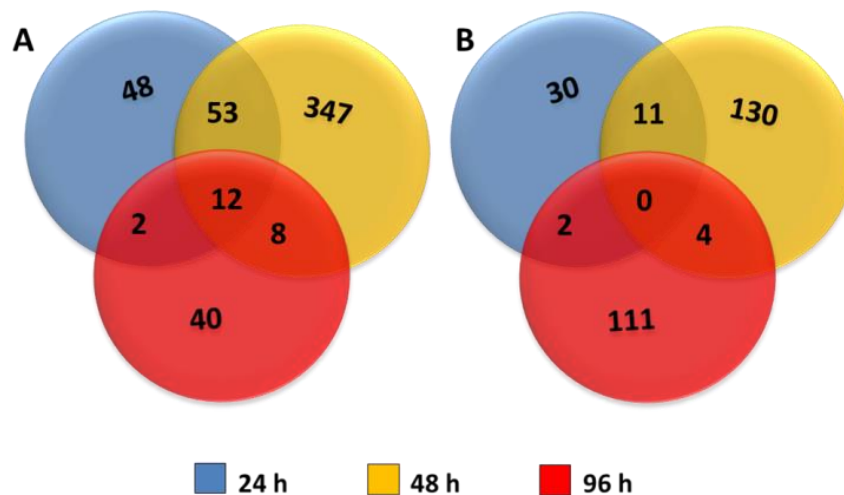
The biological processes considerably affected by aphids, for both up and down-regulated sequences, were related to chemical reactions and pathways involving those compounds which are formed as a part of the normal anabolic and catabolic processes. Specifically, “cellular metabolic process” and “primary metabolic process” were the most present GO terms showing the highest percentage of sequences at each experimental point. Also “response to stress” category was largely influenced during the time course. For this last process the percentage of sequences increase from ~9% at 24 hpi to ~13% at 48 hpi for over-expressed and under-expressed genes. “Response to stress” category peaked at 96 hpi for down-regulated genes with the highest percentage of sequences (16.2%). In this category genes involved in response to external stimulus as well as response to biotic and abiotic stresses categories are included. Genes affected by infestation were also categorized as “nitrogen compound metabolic process”. At 24 hpi and 48 hpi ~40% of up-regulated genes were present in this category and declined to 20% at 96 hpi. “Signal transduction” category for up-regulated genes and “cell wall organization or biosynthesis” category for down-regulated genes appeared at 96 hpi, and were not highlighted for the remaining time points (figure 3.4, E and F).



**Figure 3.4.** Multilevel distribution of the differentially expressed sequences annotated by GO category. For each category, it is indicated the number of annotated sequences. Up-regulated genes after 24 hpi (A); down-regulated genes after 24 hpi (B); up-regulated genes at 48 hpi (C); down-regulated genes at 48 hpi (D); up-regulated genes at 96 hpi (E); down-regulated genes at 96 hpi (F).

Venn diagrams (figure 3.5) showed the intersection among the differentially expressed genes at the three time points.





**Figure 3.5.** Venn diagrams illustrating the number of zucchini up-regulated (A) and down-regulated (B) differentially expressed genes considering the three different time points after *A. gossypii* infestation.

A total of 14 genes affected by aphids were identified as common at the three sampling times, of which 12 genes were overexpressed during the whole time span (figure 3.5, A). Only 7 of them were associated with at least one GO category. Among common genes, listed in table 3.3, it was possible to highlight one *EF hand calcium-binding protein*, which encodes a calmodulin-like protein. Also a *Polyketide cyclase/dehydrase and lipid transport superfamily protein*, up-regulated in our experimental condition, was listed. This protein is involved in lipid transport across the membrane due to the presence of a StAR-related lipid transfer (START) domain. An additional differentially expressed gene is a member of the Actin-depolymerizing factor (ADF/cofilin) family protein that alter actin dynamics and mediate actin depolymerisation. Two genes encoding for *Ribosomal proteins* and one *Enolase* were recorded among transcripts influenced during the entire time span. Finally, one *Cysteine proteinases protein* was also present. The expression value of this transcript increased progressively during the time course (table 3.3). Surprisingly, this sequence showed a high sequence identity (95%) with a *cathepsin B* sequence (cds) belonging to the brown citrus aphid, *Toxoptera citricida* (BK006332.1).

**Table 3.3.** Differentially up-regulated genes shared among three sampling times in response to *A. gossypii* feeding.

Description	log Fold Change			GO annotation
	24h	48h	96h	
EF hand calcium-binding protein	2.5611	2.9110	2.5449	BP: oxidation-reduction process; MF: 2-alkenal reductase [NAD(P)] activity; MF: calcium ion binding
Actin-depolymerizing factor	4.1208	3.4960	3.0474	BP: actin filament depolymerization; MF: actin binding; CC: actin cytoskeleton
Cysteine proteinases superfamily protein	5.1299	6.1715	7.0696	BP: proteolysis; MF: cysteine-type endopeptidase activity; BP: regulation of catalytic activity; CC: integral component of membrane
Enolase	6.6036	7.2536	5.6975	BP: L-phenylalanine biosynthetic process; MF: DNA binding; BP: tyrosine biosynthetic process; MF: phosphopyruvate hydratase activity; BP: gluconeogenesis; BP: glycolytic process; BP: tryptophan biosynthetic process; BP: phosphopyruvate hydratase complex
Polyketide cyclase/dehydrase and lipid transport superfamily protein	3.7278	4.0399	2.7219	MF: lipid binding; CC: vacuole
Zinc-binding ribosomal protein family protein	5.5378	6.9271	5.7928	CC: ribosome; BP: ribosome biogenesis; BP: translation; MF: structural constituent of ribosome
Ribosomal L29e protein family	5.3487	7.1162	6.3195	CC: ribosome; BP: ribosome biogenesis; BP: translation; MF: structural constituent of ribosome

BP: Biological Process; MF: Molecular Function; CC: Cellular Component

### 3.2.3 Evaluation of aphid-derived sequences among *C. pepo* differentially expressed genes

To verify the possible presence to other aphid-related sequence in our DEGs dataset, a BLASTx analysis was performed comparing the assembled transcripts with the *Acirrhosyphon pisum* protein database (version 2.0). The resulting BLAST output file was parsed to filtered out matches with percentage identity lower than 70. Thirty-five out of 766 DEGs displayed significant BLAST hits. By taking advantage of the *A. pisum* gene annotation (version 2.1 available on-line) we attached functional annotations to these sequences. More than 50% of sequences was annotated as ribosomal proteins. Then, the 35 genes identified were manually searched against the nr NCBI database (BLASTn) to confirm their aphid origin. This step reduced the initial dataset to 10 genes as truly aphid-derived sequences. As reported in table 3.4, it is possible to highlight two aphid-derived sequences annotated as Cathepsin B.

Additionally, to confirm the aphid nature of amplified transcripts, these sequences were compared with the *A. gossypii* transcriptome *de novo* assembled from salivary gland tissue (Pennacchio *et al.*, unpublished data).

All the sequences showed a similarity percentage score between 98.8% and 100%, as reported in table 3.5. An example of alignment is reported in figure 3.6. The transcript CUCPM\_L16538\_T\_1, annotated as Ribosomal Protein L29, and derived from *C. pepo* DEGs, showed 100% identity with an *A. gossypii* transcript with the same annotation. The remaining alignments are reported in Appendix figure A.2 to A.5.

**Table 3.4.** List of aphid-derived transcripts identified among *C. pepo* differentially expressed genes after *A. gossypii* infestation, and their GO annotation.

Query ID	ACYPI Description	logFC			GO annotation
		24 h	48 h	96 h	
CUCPM_L9364_T_1	elongation factor 1 alpha	/	6.7441	/	CC: ribosome; BP: regulation of translational elongation; MF: translation elongation factor activity; MF: GTP binding
CUCPM_L16538_T_1	60s ribosomal protein l29	5.3486	7.1162	6.3194	CC: ribosome; BP: ribosome biogenesis; BP: translation; MF: structural constituent of ribosome
CUCPM_L15699_T_1	enolase	6.6035	7.2535	5.6974	MF: magnesium ion binding; BP: L-phenylalanine biosynthetic process; MF: DNA binding; BP: tyrosine biosynthetic process; BP: gluconeogenesis; BP: glycolytic process; CC: phosphopyruvate hydratase complex;
CUCPM_L15356_T_1	isoform b	5.4465	6.8937	/	BP: ATP synthesis coupled proton transport; CC: mitochondrial proton-transporting ATP synthase complex, catalytic core F(1); MF: proton-transporting ATPase activity, rotational mechanism
CUCPM_L16719_T_1	nadh dehydrogenase	/	6.1259	/	BP: metabolic process; MF: catalytic activity
CUCPM_L16501_T_1	ribosomal protein l37a	/	6.7824	/	BP: ribosome biogenesis; BP: translation; CC: cytosolic large ribosomal subunit; MF: structural constituent of ribosome
CUCPM_L16529_T_1	ribosomal protein s18	/	6.9599	/	CC: ribosome; BP: ribosome biogenesis; BP: translation; MF: structural constituent of ribosome; MF: rRNA binding
CUCPM_L16634_T_1	tpa_inf: cathepsin b	5.1298	6.1715	7.0695	BP: sensory perception of taste; BP: proteolysis; MF: cysteine-type endopeptidase activity; CC: integral component of membrane; BP: regulation of catalytic activity
CUCPM_TC17109	tpa_inf: cathepsin b	/	4.5270	6.4456	BP: sensory perception of taste; BP: proteolysis; MF: cysteine-type endopeptidase activity; CC: integral component of membrane; BP: regulation of catalytic activity;
CUCPM_L16896_T_1	V-type proton ATPase subunit B	/	6.7847	/	BP: cellular iron ion homeostasis; MF: magnesium ion binding; MF: ATP binding; BP: intracellular signal transduction; BP: oxidative phosphorylation; CC: endomembrane system; MF: protein binding; BP: interaction with host; CC: integral component of membrane; BP: proton-transporting V-type ATPase, V1 domain

**Table 3.5.** Similarity-based search of aphid-infested *C. pepo* DEGs against *A. gossypii* transcriptome.

ID <i>C. pepo</i> transcripts <sup>1</sup>	ID <i>A. gossypii</i> transcripts <sup>2</sup>	e-value	% identity
CUCPM_TC17109	DN11910_c0_g3_i1	0.0	98.84
CUCPM_L16634_T_1	DN11910_c0_g1_i1	0.0	100
CUCPM_L16529_T_1	DN2214_c0_g1_i1	0.0	99.82
CUCPM_L16501_T_1	DN18663_c0_g1_i1	0.0	100
CUCPM_L16538_T_1	DN4152_c0_g1_i1	0.0	100
CUCPM_L9364_T_1	DN9661_c0_g1_i1	0.0	100
CUCPM_L16719_T_1	DN7973_c0_g1_i1	2.00E-107	100
CUCPM_L15356_T_1	DN8538_c0_g1_i2	5.00E-113	100
CUCPM_L15699_T_1	DN11656_c0_g1_i2	4.00E-109	100
CUCPM_L16896_T_1	DN12363_c0_g1_i1	6.00E-123	100

1: ID assigned to *de novo* assembled *C. pepo* transcripts

2: ID assigned to *de novo* assembled *A. gossypii* transcripts (Pennacchio et al., unpublished)

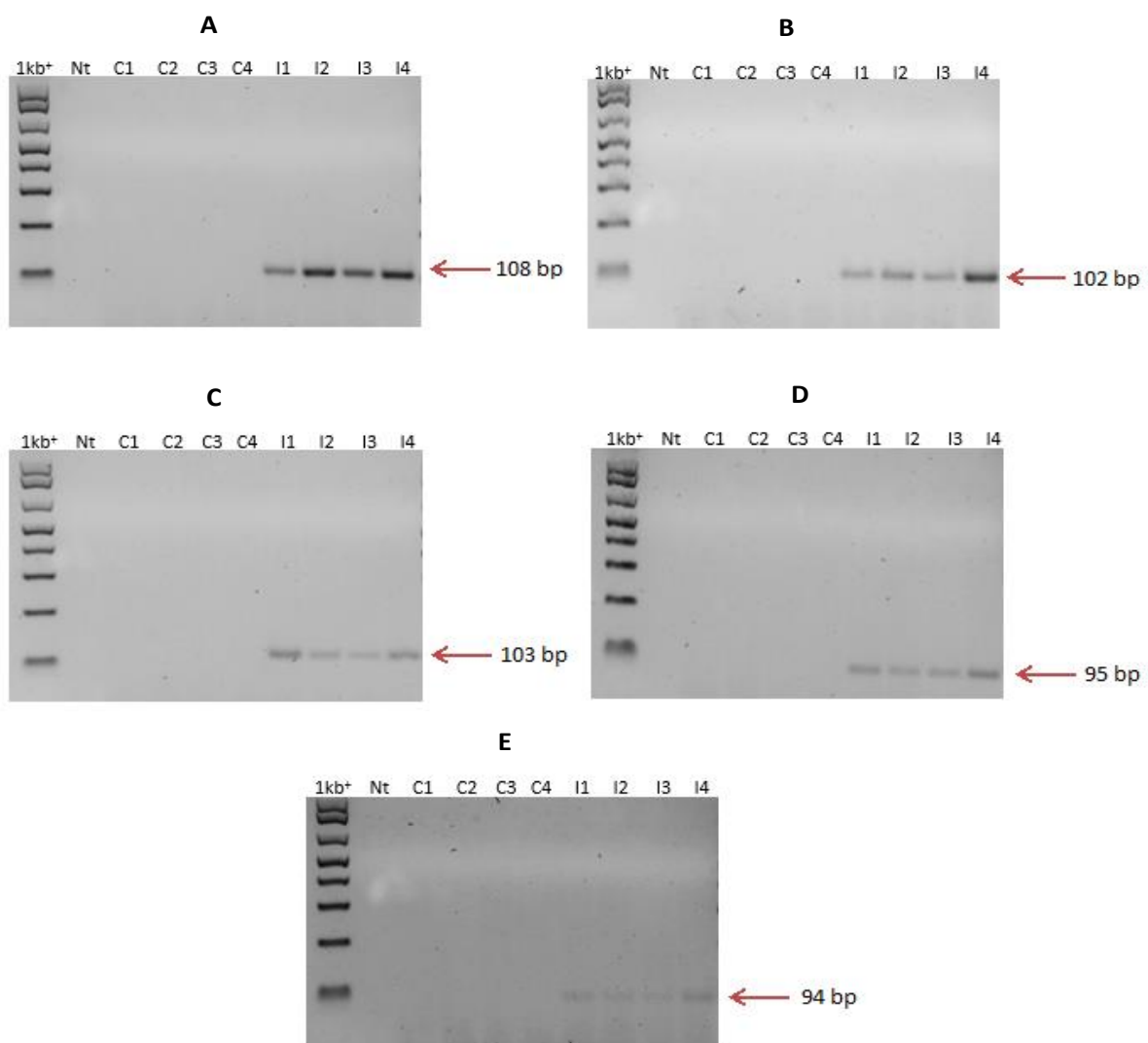
Sequence name ( <i>C. pepo</i> )	Sequence name ( <i>A. gossypii</i> )	Identities	Gaps	e-value	Score
CUCPM_L16538_T_1 (531 nt)	DN4152_c0_g1_i1 (500 nt)	351/351 (100%)	0/351 (0%)	0.0	649 bits (351)
DN4152_c0_g1_i1- consensus	ATTCTTGTTTTAAGTAGTTTTGTTTATTAATTGCTTAAAAACAGGAAGAATTCATAAAAA ATTCTTGTTTTAAGTAGTTTTGTTTATTAATTGCTTAAAAACAGGAAGAATTCATAAAAA				
DN4152_c0_g1_i1- CUCPM_L16538_T_1+	AAATAAAGTTTAAATGTATTAAATTCGTTTAAAAATAAACGAAAAGGAAATTTATTGATGT TGT				
consensus	AAATAAAGTTTAAATGTATTAAATTCGTTTAAAAATAAACGAAAAGGAAATTTATTGATGT				
DN4152_c0_g1_i1- CUCPM_L16538_T_1+	CAAACAATTTTTTAAAGTCAGAATTATGGTTTTTGGTTGCGGCTTCTCTCTCTGAAT CAAACAATTTTTTAAAGTCAGAATTATGGTTTTTGGTTGCGGCTTCTCTCTCTGAAT				
consensus	CAAACAATTTTTTAAAGTCAGAATTATGGTTTTTGGTTGCGGCTTCTCTCTCTGAAT				
DN4152_c0_g1_i1- CUCPM_L16538_T_1+	TCAGCAAACTTGGTCCTTAAGGCTAACCTCTTTTCACTTCTTTTATTAGCCCTGGCCAAC TCAGCAAACTTGGTCCTTAAGGCTAACCTCTTTTCACTTCTTTTATTAGCCCTGGCCAAC				
consensus	TCAGCAAACTTGGTCCTTAAGGCTAACCTCTTTTCACTTCTTTTATTAGCCCTGGCCAAC				
DN4152_c0_g1_i1- CUCPM_L16538_T_1+	TGATCTGCGGTGGACAGATTTCCTTTTAAAGCATGTTTCTGGTTTCTCAAAAAATTTTGA TGATCTGCGGTGGACAGATTTCCTTTTAAAGCATGTTTCTGGTTTCTCAAAAAATTTTGA				
consensus	TGATCTGCGGTGGACAGATTTCCTTTTAAAGCATGTTTCTGGTTTCTCAAAAAATTTTGA				
DN4152_c0_g1_i1- CUCPM_L16538_T_1+	CAAACACCACGACGTGATTCATGCTGTATTCTTTGGCCATAAATACCATTACGATGG CAAACACCACGACGTGATTCATGCTGTATTCTTTGGCCATAAATACCATTACGATGG				
consensus	CAAACACCACGACGTGATTCATGCTGTATTCTTTGGCCATAAATACCATTACGATGG				
DN4152_c0_g1_i1- CUCPM_L16538_T_1+	TCCTTACGATTTTGATTGTGATTGGTATGATTCTTTGACTTGCCATTTTGAACGGTTAT TCCTTACGATTTTGATTGTGATTGGTATGATTCTTTGACTTGCCATTTTGAACGGTTAT				
consensus	TCCTTACGATTTTGATTGTGATTGGTATGATTCTTTGACTTGCCATTTTGAACGGTTAT				
DN4152_c0_g1_i1- CUCPM_L16538_T_1+	ACGAATATAATTTCTTAAACACGAAAATGGACAATACTGCCGTTTACACGACGACGAAAA ACGAATATAATTTCTTAAACACGAAAATGGACAATACTGCCGTTTACACGACGACGAAAA				
consensus	ACGAATATAATTTCTTAAACACGAAAATGGACAATACTGCCGTTTACACGACGACGAAAA				
DN4152_c0_g1_i1- consensus	CTGGAGAATGTCTATACTAT CTGGAGAATGTCTATACTAT				

**Figure 3.6.** Sequence alignment of CUCPM\_L16538\_T\_1 and DN4152\_c0\_g1\_i1, from *C. pepo* and *A. gossypii* respectively, annotated as Ribosomal Protein L29. The portion highlight in red represents the amplicon sequence.

All together these observations strongly suggest that the above mentioned transcripts belong to aphid. However, to exclude possible contamination during sample preparation, a new infestation experiment was performed. The experimental conditions of this assay were the same as reported for RNA-Seq experiment, but zucchini leaves were collected only after 48 hpi considering that all aphid transcripts were expressed at this time point (see table 3.4). Before leaf sampling, all aphids and new-born neanids were carefully removed using a paintbrush.

RT-PCR analyses, on the newly prepared cDNA, performed using specific primers designed on aphid-derived sequences, resulted in amplification of five out of 10 tested genes. Specifically, the two *Cathepsin B* (figure 3.7, A-B) and the three *Ribosomal protein* sequences (figure 3.7, C-E) were successfully amplified in infested samples. The remaining five genes did not show any amplification.

These findings strongly support the hypothesis that at least these five transcripts belong to aphids.



**Figure 3.7.** Result of electrophoretic analyses after RT-PCR on aphid-derived sequences. A: Cathepsin B (ID: CUCPM\_TC17109); B: Cathepsin B (ID: CUCPM\_L16634\_T\_1); C: Ribosomal protein L29 (ID: CUCPM\_L16538\_T\_1); D: Ribosomal protein S18 (ID: CUCPM\_L16529\_T\_1); E: Ribosomal protein L37 (ID: CUCPM\_L16501\_T\_1). 1kb+: molecular marker; Nt: no template; C1-C4: cDNA of un-infested replicates; I1-I4: cDNA of infested replicates.

### 3.2.4 Zucchini transcriptional response to aphids

“San Pasquale” genes influenced by *A. gossypii* were classified in ten categories: stress and defence responses, signal transduction, phytohormone-related response, transcription, cell wall modification, photosynthesis, primary metabolism, secondary metabolism, cellular organization and transport. For the purpose of clarity, genes that participate in more than one process are presented only once considering their prevalent role in plant metabolism. Moreover, aphid-derived transcripts previously identified were removed from DEGs list.

#### 3.2.4.1 Signalling related, stress and defence response genes

Following aphid attack several genes involved in oxidative stress responses, defence molecule biosynthesis, signalling pathways and biotic and abiotic defence responses were regulated at transcriptional level.

A total of 30 genes associated with oxidative stress were differentially expressed, suggesting that the cell redox state was altered. Among the overexpressed genes, ROS-detoxifying enzymes, which are important in the redox balance maintaining during oxidative stress following insects and/or pathogens infection, were influenced. One sequence encoding for *Catalase 2* was overexpressed after 24 hpi, and a total of seven *Peroxiredoxin* were up-regulated at 24 hpi and 48 hpi. Always at 48 hpi, one *Cu/Zn Superoxide dismutase*, one *Peroxidase* and two *Glutaredoxins* were up-regulated. Conversely, two *Glutathione S-transferases* and five *Peroxidase* genes were down-regulated at the same time point. Furthermore, ROS cellular damage could result in lipid membrane oxidation and accumulation of toxic compounds (e.g. reactive aldehydes). Two *Aldehyde dehydrogenases* and one *Aldo/keto reductase*, active in detoxification mechanism of toxic aldehydes, were up-regulated at 48 hpi and 96 hpi, respectively. Mechanical damage of cell membranes that occurred after aphid puncturing leads not only to ROS formation, but also to an increase of cytosolic  $Ca^{2+}$  concentration. A total of five gene encoding *Calcium-binding* and *Calmodulin-like* proteins were up-regulated at different time points and one gene coding for an *EF-hand calcium-binding protein* was overexpressed during the whole time course. Three calcium-binding proteins down-regulated at 48 hpi resulted overexpressed during the last experimental point. Moreover, aphid feeding produced deregulation of genes coding for kinase receptor proteins. When plants are attacked by insects, FACs in insect oral secretions and other compounds can be recognized by multiple receptors and act as elicitors of plant response. In this study six genes coding *Leucine-rich repeat (LRR) family proteins* and *Probable LRR receptor-like serine/threonine protein kinases* were differentially regulated at 96 hpi, three up-regulated and three down-regulated, and only one *LRR-RLK* gene was down-regulated at 48 hpi. A total of 14 genes linked to abiotic stress response was differentially regulated during the three time points, and included ten genes encoding for *Heat shock proteins* and *DNA-J chaperone proteins*, involved in cellular and protein stability during stress. Moreover, four *Late embryogenesis abundant (LEA) hydroxyprolin-rich family proteins* were up-regulated at 24 and 48 hpi, which are involved in protecting higher plants from damage caused by environmental stresses.

#### 3.2.4.2 Transcription-related genes

A total of 28 differentially expressed genes putatively involved in regulation of transcription was identified. Genes annotated as *Pentatricopeptide repeat-containing proteins (PPR)* were overexpressed (5 genes) during the time span, and other two members of the same family were down-regulated at 24 and 48 hpi, respectively.

Four transcripts annotated as *BTB/POZ* and *TAZ domain-containing proteins*, which mediate transcriptional regulation in response to signal molecules such as  $\text{Ca}^{2+}$ ,  $\text{H}_2\text{O}_2$ , and SA, were down-regulated at 24 (1 gene) and 48 hpi (3 genes). An additional transcription factor down-regulated at 24 and 48 hpi was a RING finger and Zinc finger domain-containing protein. Finally, classes of genes involved in transcriptional regulation associated with abiotic stress response were mainly identified among down-regulated by *A. gossypii* infestation. Specifically, at 48 hpi three *NAC domain-containing proteins*, two *NAC domain transcriptional regulator superfamily proteins* and three *MYB-like transcription factors* were identified. Two *GATA transcription factors*, which are mostly implicated in light-dependent gene regulation, were strongly down-regulated at 96 hpi.

### 3.2.4.3 Primary metabolism and photosynthesis related genes

*A. gossypii* infestation determined alteration in gene expression related to photosynthesis and primary metabolic processes. Genes putatively involved in photosynthesis (25 genes), and associated with photosystem I and II complexes, were affected especially at 48 hpi. After 24 and 48 hpi from infestation genes coding for four *Protochlorophyllide reductases*, two *Phototropin-1* proteins, a *Photosystem I reaction center subunit V, chloroplastic* and two *ATP synthase proteins* were up-regulated. At 48 hpi were also over-expressed genes encoding for *Photosystem II reaction center PsbP*, *Photosystem I reaction center subunit K*, *Chlorophyll A/B binding protein*, seven *ATP synthase proteins* and one *ATP binding protein*, and two *RuBisCO large subunit-binding protein subunit alpha, chloroplastic*. Moreover, two genes associated with electron transport were identified: one gene coding for a *Plantacyanin* (Blue protein) was up-regulated at 48 hpi; a *Blue copper protein* was found down-regulated at 96 hpi. Zucchini plants response to *A. gossypii* infestation included regulation of primary metabolism. Indeed genes putatively related to carbohydrates, amino acid (e.g. methionine, cysteine, etc.) and lipids metabolism were influenced. Genes related to sugar metabolism were mainly up-regulated at 24 and 48 hpi such as *Probable sucrose-phosphate synthase*, *Glycosyl transferase*, *UDP-glucuronate decarboxylase* and *Sucrose synthase*. Also genes involved in glycolysis such as *Fructose-1,6-bisphosphatase*, *Glyceraldehyde 3-phosphate dehydrogenase*, and in TCA cycle such as *Succinyl CoA-ligase* were up-regulated. Moreover, two genes encoding for *beta-galactosidase* and *beta-xylosidase*, and involved in polysaccharides hydrolysis, were up-regulated at 48 hpi and then down-regulated at 96 h post aphid infestation. Fifteen genes involved in lipid metabolism were regulated. Among them, five *Acyl carrier proteins*, a *3-ketoacyl-CoA synthase 19 protein* and a *Fatty acid desaturase* were overexpressed at 24 and 48 hpi. At the second time point two genes coding for *Phosphomethylethanolamine N-methyltransferase* enzymes (PEAMT; EC 2.1.1.103) which catalyses the key step in choline (Cho) biosynthesis, were up-regulated. Two *GDSL esterase/lipase 1* enzymes were also deregulated at 48 hpi. Differences of expression for genes related to protein metabolism were observed at each time point. Notably, plant response was limited at 24 hpi, but increased after 48 h from the beginning of infestation. A total of 91 genes coding for *ribosomal proteins (RP)*, both cytosolic and plastidic, was identified as overexpressed at 24 and 48 hpi. All ribosomal proteins identified at 96 hpi (5 genes) were removed from the list of DEGs because classified as aphid-derived sequences. A total of five genes related to different classes of peptidases were down-regulated at 96 hpi: two *Cysteine peptidases*, two *Aspartic proteases* and a *Serine carboxypeptidase-like protein*. Three *RING/U-box E3-*

*ubiquitin ligase proteins* were also down-regulated at 24 and 48 hpi, as well as a *Proteasome alpha subunit type-1* protein and an *Ubiquitin-conjugating enzyme E2*. Five genes encoding for *Cyclophylin peptidil/prolyl cis-trans isomerases* were up-regulated at 48hpi. These genes seem to participate in protein folding process not only as prolyl isomerases but also as chaperones.

#### 3.2.4.4 Cell wall modification

Biotic stress agents such as pathogens and herbivorous insects are able to influence the expression of genes related to cell wall metabolism. A total of 23 genes active in cell wall metabolism and remodelling was differentially expressed after *A. gossypii* infestation. At the first and second time point majority of genes were up-regulated, whereas all genes associated with this category were down-regulated at 96 hpi. Over-expressed genes included three *Cellulose Synthase* enzymes, which are directly involved in cellulose microfibrils synthesis, and two *Extensin* proteins. Six *Arabinogalactan (AGP)* proteins were up-regulated at 48 hpi, specifically three AGP proteins and three *Fascilin-like Arabinogalactan* genes were annotated, which are usually secreted at the wound/infection sites. At the same time point one gene involved in callose deposition, coding for a *Sucrose synthase 6*, was also up-regulated. Conversely, genes coding for enzymes involved in cell wall degrading such as a *Polygalacturonase 2* and two *Glycosyl hydrolases (Endoglucanase 6 and Beta-glucosidase)* were down-regulated at 96 hpi. Two *Fascilin-like Arabinogalactan* genes were also down-regulated. Notably, two *Xylem cysteine peptidases (XCP1 and XCP2)* were down-regulated at the last time point. These genes play a crucial role in regulating the events of xylogenesis and secondary wall thickening (Avci *et al.*, 2008).

#### 3.2.4.5 Secondary metabolism related genes

Genes associated with secondary metabolism showed significant differential expression. After 24 hpi genes involved in vitamin metabolism such as *Myo-inositol oxygenase 4*, *Riboflavin synthase* and *Thiamin C phosphomethylpyrimidine synthase* were up-regulated. The first two genes were also found overexpressed at 48 hpi. After 96 h from infestation another *Thiamine thiazole synthase*, involved in biosynthesis of the thiamine (vitamin B1) precursor thiazole, was up-regulated. A gene coding for *4-coumarate:CoA ligase (4CL: EC 6.2.1.12)* was over-expressed at 24 and 48 hpi. This enzyme is the last of the general phenylpropanoid pathway that catalyses the formation of a number of natural products, such as flavonoids, stilbenes, and lignin, which serve diverse functions such as phytoalexins that protect against fungal infections, ultraviolet (UV) protectants, flower and fruit pigments, and structural components of cell walls (Wang *et al.*, 2016). Two genes involved in Non-Mevalonate Pathway (MEP) were influenced by aphid feeding. Specifically, one gene coding for *4-hydroxy-3-methylbut-2-enyl diphosphate reductase* was down-regulated at 24 hpi. The second gene encodes for a *2-C-methyl-D-erythritol 4-phosphate cytidyltransferase, chloroplastic* which was found up-regulated at 48 hpi. Among genes negatively affected by *A. gossypii* infestation at 96 hpi a gene encoding for a *Terpene synthase* was strongly down-regulated. Five genes coding for *Polyketide cyclases/dehydrases*, which are involved in polyketide synthesis, were overexpressed during the whole time course. Furthermore, genes related to xanthophyll biosynthesis were up-regulated at 48 hpi such as *Zeaxanthin epoxidase* and *Beta-carotene hydrolase 2*. Finally, five genes encoding for *Cytochrome p450* were up- and down-regulated at 48 and 96 hpi. These enzymes have a role in



defence response and are involved in detoxification of toxic molecules as well as in the biosynthesis of a wide range of molecules associated with defence and signal transduction (Li et al., 2002).

#### 3.2.4.6 Transport and cell maintenance related genes

Genes putatively involved in transport processes were strongly affected by aphid infestation after 48 and 96 h. A total of 26 genes involved in water, ion, heavy metal and metabolite transport was annotated. At 24 hpi two *Aquaporin*, one *Probable anion transporter 2* and a *Mitochondrial carnitine/acylcarnitine carrier-like protein* were overexpressed. The latter gene is connected to fatty acid metabolic pathway in plant tissues. Five genes coding for *Aquaporin plasma membrane intrinsic proteins (PIPs)*, which show an important role in controlling membrane water permeability, were up-regulated at 48 hpi. Two genes encoding for *Mitochondrial import receptor subunit TOM5 homolog*, component of the TOM (translocase of outer membrane) receptor complex, were overexpressed at 48 hpi. The same expression profile was observed for a *Copper transport protein, Metal ion binding, ABC-2 type transporter family protein* and a *Magnesium transporter*. After 96 hpi a member of the HMA family, *Putative cadmium/zinc-transporting ATPase HMA4*, was up-regulated. At the same time point, one *Heavy metal-associated isoprenylated plant protein 26 (HIPP26)* involved in heavy metal homeostasis and detoxification mechanisms was down-regulated. Twenty-one genes involved in cell maintenance showed differential expression following aphid infestation. This category grouped genes coding for proteins implicated in cell cycle, cellular component organization and cell differentiation. The majority of genes of this category was up-regulated at 48 hpi. One gene coding for an *Actin depolymerizing factor 2* was overexpressed during the whole time course. Genes involved in cytoskeleton organization and intracellular movement were also recorded as differentially expressed. One *Kinesin-like protein*, important in intracellular transport, mitosis and meiosis, was overexpressed at 24 hpi, and a *Formin-like protein*, involved in the organization and polarity of the actin cytoskeleton, was down-regulated at 48 hpi. Moreover, a gene coding for *Microtubule-associated protein TORTIFOLIA1* (Plant-specific microtubule-associated protein-MAP), that regulates the orientation of cortical microtubules by modulating microtubule severing and the direction of organ growth (Wightman et al., 2013), was down-regulated at 96 hpi. Finally, genes involved in cell cycle and nucleotide and nucleic acid metabolism were identified. Two *Cyclin-dependent kinases* were up-regulated at 48 hpi, and two *Tubulin protein* were down-regulated at 96 hpi. A total of six genes involved in chromatin structure was over-expressed at 48 hpi: three *Histone H2A* proteins and a single *Histone H2B, H3 and H4*.

#### 3.2.4.7 Phytohormonal-related genes

Genes related to SA- and JA-signalling pathways were modulated by *A. gossypii* feeding.

Isochorismate synthase 1 (ICS1) is a key enzyme of the SA biosynthetic pathway in response to pathogen infection (Macaulay et al., 2017). Overexpression of an *ICS1* gene was observed at 24 and 48 hpi. However, genes associated with SA were mainly influenced at 48 hpi. Nine genes encoding for Pathogenesis-Related (PR) proteins have been identified in response to insect feeding. Two *Pathogenesis-related Thaumatin-like proteins (PR5)* and three *Glucan endo-1,3-beta glucosidases (PR2)* were up-regulated at 48 hpi, as well as a *Subtilisin-like protease* which encode for a PR protein. Two genes encoding for *Pathogenesis-related proteins Bet v 1*

family (*PR10*) were identified at 96 hpi (one up-regulated and one down-regulated). One transcript annotated as *NIM1-interacting 1* (*NIMIN1*) was down-regulated at 48 hpi. In *Arabidopsis* *NIMIN1* is able to strongly bind *NPR1/NIM1*, a key regulator of systemic acquired resistance (SAR) and of SA signal transduction pathway (Weigel *et al.*, 2001). Among SA-related genes, one *Ankyrin repeat protein* gene was up-regulated at 24 and 48 hpi, as well as a gene encoding for a *Nudix hydrolase 8* (*NUDX8*) was up-regulated at 24 hpi. However, other three *NUDX hydrolase* genes were down-regulated at 48 hpi.

JA/Et-related and wounding-related genes were mainly down-regulated during the whole time course. Among these genes, a total of four *Serine proteinase inhibitor Type-I* and *Wound-induced proteinase inhibitor 1* was strongly down-regulated at 24 and 48 hpi, and a *Serine carboxypeptidase-like* was down-regulated at 96 hpi. One gene encoding for a *Plant defensin 1.2* (*PDF 1.2*) was also down-regulated at 48 hpi, as well as three genes annotated as *Senescence-associated gene* (*SAG*).

One up-regulated transcription factor involved in plant response to stress and related to ethylene metabolism, was the *Ethylene-responsive transcription factor related to AP2-7*. Additionally, at 24 hpi a *bHLH transcription factor* gene was overexpressed compared with control plants, and two members of the same gene family were down-regulated at 48 hpi. Some evidence shows that these transcription factors participate in the activation of JA-induced defence genes in tomato and *Arabidopsis* (Boter *et al.*, 2004).

### 3.2.5 Aphid behaviour response to synthetic MeSA

“San Pasquale” plants treated with the synthetic MeSA were used to investigate the effect of exogenous application on aphid fixing behaviour. Moreover, these results were compared with observation on aphid behaviour performed on pre-infested zucchini plants. Observation performed on zucchini plants feed on by *A. gossypii* for 48 h and after aphid removing, showed that previous infestation elicited dispersal behaviour in newly-laid nymphs. Indeed, less than 50% of newborn nymphs stay on the same leaf where their mother (a pre-reproductive adult) was placed, which correspond to the leaf that had received the treatment (aphid feeding), and were found distributed on the whole plant. Conversely, on control plants, for which no pre-infestation was performed, a significant higher number of newborn nymphs remained near their mothers (97.27%;  $P = 2.17E^{-81}$ ).

To test the hypothesis that SA is associated with the observed plant-mediated alteration of aphid dispersal behaviour, we treated the plant with the synthetic MeSA. A significant difference ( $P = 4.60E^{07}$ ) was evident starting from 24 h, when less number of aphids (35%) climb the MeSA-treated plants to start their feeding activity compared with control plants (53.6%). This difference was even higher at 48 h when 47% of pre-reproductive *A. gossypii* tested start to feed on MeSA-treated zucchini plants compared with controls (80%;  $P = 1.70E^{-19}$ ). It seems important to underline that the aphids were never in contact with exogenous MeSA, only the plant was.

### 3.3 Discussion

In this study gene expression profiling was performed on zucchini leaves after aphid infestation using the RNA-Seq technique. To use RNA-Seq data to compare expression between samples, it is necessary to turn millions of short reads into a quantification of expression (Oshlack *et al.*, 2010). Taking advantage of the *de novo* transcriptome assembly from *A. gossypii*-infested leaves of “San Pasquale” cultivar, ~84% of high quality reads successfully aligned on reference transcriptome and the

remaining 16% did not find match. The unaligned reads probably derived from alternative splicing isoforms, which were not represented in the subset used as reference where the longest transcript for each gene *locus* was present. Summarization and normalization of mapped reads represented essential steps in the analysis of RNA-Seq data, and allowed to identify 776 DEGs. Moreover, the biological information attached to previously assembled transcripts during annotation phase was essential for categorization of DEGs (Chapter 2).

Our study provides the first insight into *Cucurbita pepo* response to the melon aphid *A. gossypii*. Zucchini plants were analysed during the early stages of infestation, before symptoms development, providing information on reprogramming of several biological processes. Interestingly, number of genes affected by feeding considerably increased from 24 to 48 hpi, and declined at 96 h post infestation. This trend is consistent with observation performed on tomato plants following *M. euphorbiae* infestation (Coppola *et al.*, 2013), and could be explained accounting for an “adaptation” response that takes place during the last stage of the compatible interaction analysed. After 96 h plants could reduce magnitude of response due to their ability to handle progression of infestation in zucchini leaves. Moreover, during defence response, plants always try to balance an effective defence activation and the reduction of negative impact on plant fitness (Walling, 2008; Coppola *et al.*, 2013).

Generation of ROS, such as hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>), superoxide anion (O<sub>2</sub><sup>-</sup>), hydroxyl radicals (HO·), is a common phenomenon following recognition of insect/pathogen derived elicitors, and plays an essential role in plant signalling. Major ROS-scavenging enzymes of plants include catalases (CAT), superoxide dismutases (SOD), ascorbate peroxidases (APX), glutathione peroxidases (GPX) and peroxiredoxins (PrxR). Sequences coding proteins involved in protection against and detoxification of ROS were up-regulated in zucchini leaves since early stage of infestation. This result is consistent with those obtained for celery and wheat plants (Divol *et al.*, 2005; Boyko *et al.*, 2006). In general, plants seems to balance ROS generation as defensive mechanism and synthesis of ROS-detoxifying enzymes to cope with their own oxidative damage (Zhu-Salzman *et al.*, 2004). However, oxidative stress-related genes are not uniformly regulated in response to aphids. At 48 hpi aphid feeding induced the expression of certain antioxidant enzymes and suppressed others. Specifically, different peroxidase genes were both up- and down-regulated. A similar response, for example, has been observed in Arabidopsis following green peach aphid infestation. Aphids induced expression of one gene encoding for superoxide dismutase, a *Cu/ZnSOD*, but down-regulated a *FeSOD* (Moran *et al.*, 2002).

Genes associated with calcium signalling showed altered expression pattern in *A.gossypii*-infested plants. Ca<sup>2+</sup> ions serve as secondary intracellular messengers mediating cell homeostasis and initiating oxidative and signalling stress cascade (Maffei *et al.*, 2007). After sensing aphid feeding, Ca<sup>2+</sup> sensors activate downstream defence by increasing expression of calmodulin, clamodulin-binding and calcium-dependent proteins (Lecourieux *et al.*, 2006). Zucchini response includes calcium-binding proteins, calmodulin-like and calmodulin-binding proteins mainly up-regulated during time course. In addition, parallel up- and down-regulation of genes encoding for calcium-binding proteins at 48 hpi could be explained in relation to an aphid method to counteract to sieve-plate occlusion probably through active removal of Ca<sup>2+</sup> ions (Will *et al.*, 2009).

It has been proposed that phloem-feeders are perceived by plants as pathogens due to similarities in mode of penetration and in hydrolytic enzymes released in plant tissues (Walling, 2000; Zhu-Salzman *et al.*, 2004). In accordance with this hypothesis we observed induction of SA-related genes by aphid feeding, but a weak effect on JA-related genes. A transcript encoding for *ICS1* gene was induced at 24 and 48 hpi. This enzyme is involved in reversible conversion of chorismate in isochorismate in pathogen-derived SA synthesis. In plants, chorismate is synthesized in the plastid, and the isochorismate (IC) pathway is considered the primary route for SA production in *Arabidopsis* (Dempsey *et al.*, 2011). Unlike wt plants, the *ics1* mutant exhibits little to no increase in SA levels following exposure to UV light, treatment with ozone or PAMPs, or pathogen infection (Dewdney *et al.*, 2000; Wildermuth *et al.*, 2001; Ogawa *et al.*, 2005; Garcion *et al.*, 2008; Tsuda *et al.*, 2008). Furthermore, aphid feeding induced expression of PR (PR2, PR5 and PR10) proteins in zucchini plants as also reported for sorghum, tomato, *Arabidopsis*, tobacco and other plant species (Moran *et al.*, 2002; De Vos *et al.*, 2005; Park *et al.*, 2006; Kuśnierczyk *et al.*, 2008; Coppola *et al.*, 2013). These proteins are known to be induced also by pathogenic fungi, and were significantly induced at 48 h from infestation, suggesting that this time period is required to develop a SA-related response. Overexpression of a member of nudix hydrolase family (*NUDX8*) at 24 h suggested the activation of SA-pathway, due to evidence of positive regulation on defence genes expression such as *ICS1* and *NPR1* in *Arabidopsis* (Fonseca and Dong, 2014). Among SA-dependent genes, an *Ankyrin binding protein* is induced at 24 and 48 hpi. This protein is putatively associated with plant response to pathogens and regulation of antioxidant metabolism (Lu *et al.*, 2003). On the other hand, few SA-related genes, were negatively regulated by aphid feeding. Among these genes a transcript coding for a NIMIN1 protein, was down-regulated at 48 hpi. NIMIN1 is one of structurally related NPR1-interacting proteins, associated with NPR1 in the nucleus, characterised in *Arabidopsis* (Weigel *et al.*, 2001). Analysis of *Arabidopsis* plants overexpressing NIMIN1 revealed that SA-mediated PR gene induction was repressed and that these plants showed abrogation of SAR. In a complementary approach, a NIMIN1 knockout mutant showed enhanced PR-1 gene induction after SA treatment, even if this hyperactivation of gene expression did not coincide with enhanced resistance. Collectively, these data suggest that NIMIN1 acts as a repressor of NPR1 (Weigel *et al.*, 2005).

Unlike SA-related genes, expression of genes associated with ET-signalling were not changed except for the *AP2-7/ERF transcription factors*, whereas a low number of JA-related genes was repressed. Serine proteinase inhibitors were strongly repressed at 24 and 48 hpi. As already demonstrated by Green and Ryan (1972) the role of these proteins in plant defence response to wounding and insect feeding is well-known. The mechanism of action of these proteins is the competitive inhibition of serine proteinase produced in insect guts by causing a reduction in availability of essential amino acids decisive for insect growth and development (Jamal *et al.*, 2013). Down-regulation was observed also for a plant defensin *PDF1.2* gene and for *bHLH transcription factors*. This result is inconsistent with that obtained for *Arabidopsis* plants fed on by *M. persicae*. Accumulation of PDF1.2, a peptide involved in the JA-/ET-dependent response pathway, was induced by aphid feeding, due to stimulation of response pathways associated with both pathogen infection and wounding (Moran and Thompson, 2001).

In our work the prevailing activation of SA-related genes during the establishment of a zucchini-aphid compatible interaction, could be associate with a prevalent role of

this hormone in zucchini defence response, along with a possible antagonistic crosstalk between SA- and JA-signalling pathways, as reported for tomato plants fed on by *M. euphorbiae* (Coppola *et al.*, 2013). Furthermore, to evaluate a possible role of SA on aphid behaviour, foliar applications of synthetic MeSA were performed. In our study, significant less aphid infestation was observed on pre-infested zucchini plants as well as on MeSA-pretreated leaves. This suggests an effect of plant defence elicited by MeSA against aphids. Aphid altered behaviour is mediated by the plant, since the aphids do not come in contact with exogenous MeSA, but only with compounds (e.g. proteins, metabolites, volatiles) produced by the plant itself under the induction of exogenous MeSA.

Although the number of differentially expressed genes and proteins that can directly affect aphids was low in percentage, the zucchini response includes modification of cell wall structure and plant metabolism that can have an effect on aphid infestation. As reported also in other studies on plant-aphid interactions, several genes involved in cell wall modification were up-regulated in infested plants. Genes encoding cellulose synthase and extensin proteins were induced since early stage of infestation. Following wounding, increased extensin deposition and cross-linking with cell wall components should lead to a more impenetrable cell wall barrier (Showalter, 1993; Rashid, 2016). Up-regulation of these genes was also reported for celery plants in response to green peach aphid feeding on leaves (Thompson and Goggin, 2006). Arabinogalactan protein (AGP) gene expression was induced at 48 hpi. AGPs were found to be secreted at wound/infection sites, and it has been suggested that they produce physical barrier against invading organisms (i) by creating a gel plugs, and/or (ii) by producing cross-link with cell wall structure in association with extensins (Rashid, 2016). Furthermore, the overexpression of a gene associated with callose deposition represents an important result and may be indicative of the plant effort to induce sieve tube occlusion and avoid phloem sap loss (Furch *et al.*, 2007). On the other hand, down-regulation at 96 hpi of genes encoding polygalacturonase and endoglucanase proteins, associated with catabolism of cell wall components, was also observed in tomato plants in response to aphid infestation (Coppola *et al.*, 2013).

The phloem is the principal route used for translocation of photoassimilates, under control of various carriers and transporters regulated by a number of stimuli, including stress (Divol *et al.*, 2005). To avoid severe damage after aphid invasion, plants can respond by increasing production of housekeeping sequences involved in photosynthesis, protein synthesis, antioxidant production and maintenance of cell homeostasis (Smith and Boyko, 2007). Genes involved in photosynthesis (encoding photosystem I and II component, chlorophyll a/b binding proteins, RuBisCo) were induced by *A. gossypii*. Overexpression of photosynthesis genes was observed for celery infested by *M. persicae*, as well as for *N. attenuata* infested by *M. nicotianae* (Divol *et al.*, 2005; Voelckel *et al.*, 2004). Such response has been reported as result of reallocation of plant metabolites from normal growth processes to defensive function following interaction with phloem-feeders (Smith and Boyko, 2007). Aphid feeding on zucchini leaves also leads to up-regulation of a consistent number of genes involved in protein synthesis during the whole time course. Specifically, several genes related to the synthesis of ribosomal components were overexpressed. Actually, ribosomal proteins (RP), in addition to their universal role of stabilizing the ribosomal complex and mediating polypeptide synthesis, show extra-ribosomal functions such as involvement in environmental stress response (Moin *et al.*, 2016). As examples, transcripts level of Arabidopsis *RPS15a*, *RPS14*, *RPL13*

and *RPL30* increased in response to phytohormone and heat treatments (Cherepneva *et al.*, 2003; Hulm *et al.*, 2005), and *RPL13* gene was up-regulated in transgenic potato plants, leading to up-regulation of genes coding for antioxidant and defence enzymes and to tolerance against pathogens (Yang *et al.*, 2013).

On the other hand, transcripts related to protein catabolism (encoding for ubiquitin and proteasome related proteins) were both up- and down-regulated in infested leaves. It seems aphid feeding stimulated changes in phloem composition that were nutritionally advantageous to aphid themselves (Zhu-Salzman *et al.*, 2004).

Aphid feeding affected the expression of enzymes required for secondary metabolite synthesis, including phenylpropanoids and terpenoids. A gene encoding for 4CL was induced by *A. gossypii*. This enzyme catalyses the last step of general phenylpropanoid pathway for the formation of a number of natural products, such as flavonoids, stilbens and lignin, which serve diverse functions among which defence against pathogens and UV light and structural components of cell wall (Wang *et al.*, 2016). In zucchini, genes involved in MEP pathway were affected during the time course and a *Terpene synthase* gene was strongly down-regulated at 96 hpi. Terpenes constitute the largest class of plant specialised secondary metabolites, and low-molecular terpenes easily volatilise at room temperature (Yahyaa *et al.*, 2015). Terpenes synthesis is related to phenylpropanoid pathway, due to condensation of C5 precursors IPP and DMAPP into mono-, di- and sesquiterpenes by the activity of terpene synthase enzymes (Tholl, 2006). However, our results are inconsistent with those obtained, for example, in rice, for which brown planthopper feeding down-regulated several genes involved in phenylpropanoid biosynthesis and up-regulated a gene required for sesquiterpene synthesis (Zhang *et al.*, 2004; Cho *et al.*, 2005).

An additional high valuable result presented in this study is the identification of aphid-derived transcripts in plant tissues. These mRNAs are presumably actively transferred in association with aphid saliva and salivary proteins during the feeding process. The hypothesis of contamination during the sampling phase was rejected comparing the effective amplification of aphid sequences in infested leaf samples with control leaves derived from a new infestation bioassay. Moreover, salivary origin was attested comparing identified transcripts with *A. gossypii* salivary glands transcriptome (Pennacchio *et al.*, unpublished). To our knowledge this is the first study that reports mRNA transfer into plant cells through aphid saliva.

Several proteomic studies have appeared in recent years documenting the putative role of salivary proteins in the interaction between plants and phloem-feeding insects, as well as transcriptomic studies on salivary glands (Mutti *et al.*, 2006; Carolan *et al.*, 2009; Carolan *et al.*, 2011; Su *et al.*, 2012; Nicholson *et al.*, 2012; Ji *et al.*, 2013; Rao *et al.*, 2013; Rodriguez and Bos, 2013). Van Kleeff and colleagues (2016) reported for the first time the presence of sRNA from the whitefly *Bemisia tabaci* in tomato phloem of leaflets where nymphs were feeding. The transfer of putative salivary non-codings RNAs from whitefly postulate that they might target tomato host proteins. Small RNAs could facilitate the interaction between organisms by improving the attackers chance of survival (van Kleeff *et al.*, 2016).

Recently, researches on analysis of mRNA in human saliva have been published for non-invasive diagnostic applications. The existence of saliva RNA is a remarkable finding because RNA is more labile than DNA and particularly because ribonucleases are known to be present in saliva (Park *et al.*, 2006) to degrade RNA molecules. For this reason, salivary RNA may be protected against degradation when it is complexed to specific macromolecules such as lipids, proteins and lipoproteins (Park

*et al.*, 2006; Whitelegge *et al.*, 2007). Therefore, RNA in the saliva may not be as fragile as it was previously assumed to be (Palanisamy and Wong, 2010).

Proteomic studies on aphid saliva revealed the presence of a range of enzymes (i.e. oxidoreductases, glucose dehydrogenases, and proteases) that potentially reflects the need for aphids to detoxify plant defence components in order to successfully feed on their hosts (Harmel *et al.*, 2008, Nicholson *et al.*, 2012). Interestingly, for long time aphids were thought to lack proteolytic activity in their digestive tract because of their feeding on phloem sap, which contains large amounts of sucrose plus some amino acids and minerals (Tagu *et al.*, 2008). However, some plants, such as cucurbits, have a protein-rich phloem sap providing an adequate source of amino acids and nitrogen for phloem-feeding insects (Deraison *et al.*, 2004). Additionally, aminopeptidase, cathepsin L-like cysteine protease, and other proteases have been identified in the midgut of several aphids (Rahbe' *et al.*, 1995; Cristofolletti *et al.*, 2003; Kutsukake *et al.*, 2004). In social aphids of the genus *Tuberaphis*, it was shown that a cathepsin B protease was specifically produced in the gut of soldier individuals as venomous protease for attacking enemies (Kutsukake *et al.*, 2004).

Two aphid-derived transcripts encoding for cathepsin B proteins were found among zucchini overexpressed DEGs. Cathepsin B constitutes a family of cysteine proteases mainly located in the gut of insects. These enzymes have been shown to be involved in several biological processes: digestion of food proteins in midgut (Houseman *et al.*, 1984; Houseman *et al.*, 1985; Terra *et al.*, 1988), degradation and mobilization of yolk proteins during embryogenesis (Yamamoto *et al.*, 1994; Liu *et al.*, 1996), and self-destruction in programmed cell death during metamorphosis (e.g. Shiba *et al.*, 2001). However, the cysteine proteases family still remain less characterised than other classes of enzymes, such as cathepsin L (Deraison *et al.*, 2004).

To date, we have no definitive explanation for how and why these mRNAs are transmitted from aphids to plant. These transcripts might play a role in modulation of plant direct and/or indirect responses. Understanding the role of these transcripts in plant tissues, their putative translation and interaction with zucchini proteins or plant cellular components will be elucidated in future *ad hoc* experiments.

### 3.4 Conclusions

Globally, the results presented here show that *Aphis gossypii* feeding alters “San Pasquale” transcriptome extensively. SA-signalling pathway appears to play a more important role than JA pathway, which is probably antagonised during zucchini response.

Finally, overexpression of genes involved in primary metabolic processes, such as photosynthesis and protein metabolism, as well as the regulation of expression of genes related to cell wall modification, could be associated with plant counteract to aphids manipulation of plant response to ensure food supply. Moreover, the presence of aphid-derived transcripts among DEGs could be relate to a specific aphid strategy to negatively influence complete activation of zucchini plant response. The identification of genes related to aphid defence response represents an important resource for *Cucurbita pepo* and for *A. gossypii* control strategies.

### 3.5 Materials and Methods

#### 3.5.1 Plant material, aphid culture and infestation

The aphid susceptible *Cucurbita pepo* cultivar “San Pasquale” were sown and grown in dedicated climatic chamber, and the melon/cotton aphid *Aphis gossypii* Glover was reared on “San Pasquale”

plants in insect-proof cages at the Department of Agricultural Sciences, University of Naples Federico II, as previously reported (Chapter 2). To monitor changes in zucchini gene expression, three-weeks old “San Pasquale” plants were infested with 10 adult *Aphis gossypii*, following the same experimental design illustrated in the previous chapter. Leaf tissue collected from aphid-infested and un-infested (control) zucchini plants were immediately frozen in liquid nitrogen for downstream analysis. Three biological replicates were analysed for each condition at each time point. List of samples collected is reported in table 3.6 (For detailed description see Chapter 2, Materials and Methods).

**Table 3.6.** List of leaf samples collected from un-infested and *A. gossypii*-infested plants.

Condition	Time (hpi)	Sample Name
Infested	24	A24-1
		A24-2
		A24-3
	48	A48-2
		A48-3
		A48-4
	96	A96-1
		A96-2
		A96-3
Control	24	C24-1
		C24-2
		C24-3
	48	C48-1
		C48-3
		C48-3
	96	C96-2
		C96-3
		C96-4

\*hours post infestation

### 3.5.2 RNA-sequencing and data processing

Leaf samples obtained from infested and control plants were used as starting materials for total RNA extraction using the mi-RNeasy Mini kit (Qiagen), according to manufacturer’s protocol. RNA samples were quantitatively and qualitatively analysed and used for the RNA-sequencing experiment that was performed by Genomix4life S.R.L. (Baronissi, Salerno, Italy). Raw sequences were transformed into high quality reads after the following data-processing steps: i) removal of reads with Phred Quality Score per base value lower than 30 ( $\leq$  Q30) for 20% of read length; ii) 3’ and 5’ adapter sequence removal; iii) unassigned bases (N bases) at the 5’ or 3’ end removal; iv) removal of reads shorter than 75 nt. High quality reads obtained were used for digital gene expression analysis.

### 3.5.3 Differentially expressed genes identification

Zucchini transcriptome previously assembled (see Chapter 2) was used as high quality reference for read mapping and digital gene expression profiling. Bowtie2 (Langmead and Salzberg, 2012) was selected for aligning sequencing reads to reference sequences. High quality read alignment was performed on a reference dataset of 42,517 *C. pepo* transcripts, which includes the longest transcript for each gene *locus*. Reference transcriptome was first indexed using “bowtie-build” tool from Bowtie2 (Langmead and Salzberg, 2012). Bowtie2 was run in “end-to-end” mode using the “very-sensitive” option. These options provided a slow, sensitive and accurate alignment, searching for alignments which involve all read characters. Bowtie2 is also able to analyse reads generated by a paired-end sequencing. For this, files containing reads still paired at the end of cleaning process were analysed together. Conversely, files containing unpaired clean reads were individually aligned to the reference. Read summarization was carried out using eXpress (Roberts and Pachter, 2013). This software was used to estimate transcript abundances and to resolve multi-mappings of reads. eXpress was run with “fr-stranded” option, allowing to accept alignments (paired or single-end) where the first (or only) read is aligned to the forward target sequence and the second read is aligned to the reverse-complemented target sequence. Using eXpress was also possible to analyse alignment files resulting from paired and unpaired read mapping. Inter-sample normalization (Trimmed mean of M-value, TMM) and correlation analysis were performed using edgeR (Robinson *et al.*, 2010). TMM normalization method was applied on eXpress output files containing FPKM values (Fragments Per Kilobase of exon model per



Million mapped fragments) after being rounded and filtered. Specifically, transcripts characterized by FPKM values lower than 10 in more than three biological replicates were discarded. The edgeR package was also selected for the call of differentially expressed genes (DEGs). The analysis was performed comparing control and infested samples for each time point. The result of each comparison represented a list of differentially expressed genes, where, for each gene, positive and negative expression values in Log Fold Change (logFC), P-value (<0.05) and False Discovery Rate (FDR) values were recorded. Final gene lists were filtered fixing to two-fold change value in transcript levels between un-infested and infested plants and FDR values lower than 0.05.

### 3.5.4 Identification of aphid-derived sequences among *C. pepo* DEGs

The presence of aphid sequences was evaluated performing a similarity-based search (BLASTx; E-value < 1e-5) against the *Acirthosyphon pisum* protein database (<https://www.aphidbase.com/Downloads>; ACYPI proteins v2.0). Automated analysis was carried out using the BLASTALL package (release 2.2.25; Altschul *et al.*, 1990), and BLAST-formatdb tool was used to format protein and nucleotide source databases before running BLASTALL. *A. pisum* proteins annotation file (<https://www.aphidbase.com/Downloads>; Blast2GO predictions (v2.1)) was used to attach functional information to BLAST results, which were filtered fixing identity percentage value equal or higher than 70%. Moreover, the sequences obtained from BLASTALL analysis were manually compared (BLASTn) against the nr NCBI database to confirm their aphid origin.

Finally, a similarity-based search (BLASTn) was performed against the *Aphis gossypii* salivary glands transcriptome (Pennacchio *et al.*, unpublished).

To monitor presence of aphid mRNA in zucchini leaves, three-weeks old “San Pasquale” plants were infested with the melon aphid *Aphis gossypii*. Zucchini plants were placed individually in insect-proof cages in a clean climatic chamber (Temperature: 22 ± 1 °C; Relative Humidity: 75 ± 5 %; Photoperiod: L16:D8). A total of ten adult aphids was transferred onto adaxial surface using a paintbrush, and their number was daily monitored. Control plants were enclosed in insect-proof cages and were grown under the same environmental conditions. Aphids were left to feed for 48 hours after that they were counted and manually removed using a fine paintbrush. Leaf tissue was sampled from aphid-free and control plants and immediately frozen in liquid nitrogen. Four biological replicates were collected for both infested and control plants, and leaves of a single replicate were pooled for downstream analysis. Total RNA was isolated from 100 mg of tissue previously grinded in liquid nitrogen using the mi-RNeasy Mini kit (Qiagen), according to manufacturer’s protocol. RNA samples were quantitatively and qualitatively analysed with NanoDrop ND-1000 UV-vis Spectrophotometer (NanoDrop Technologies) and Agilent 2100 Bioanalyzer (Agilent Technologies), respectively. The first strand cDNA fragments were synthesized from 2 µg of total RNA, previously treated with 2 units of DNase I Amplification Grade (Invitrogen, Carlsbad, California, USA), using 200 units of SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, California, USA) according to manufacturer’s protocol. All cDNA samples were used as template for a PCR reaction performed with GeneAmp®PCR System 2700 thermal cyclers (Applied Biosystem, Foster City, California, USA) using Ef\_Cpepo Fw and Ef\_Cpepo Rv (table 3.7) as primers for amplification of *Elongation Factor 1-α* (*EF1-α*). The reaction mixture consisted of 1 µl cDNA, 4 µl GoTaq Buffer 5X (Promega, Madison, Wisconsin, USA), 0.4 µl dNTP mix (25mM), 1 µl of each primer (10 mM) and 0.3 µl GoTaq DNA Polymerase (5 U/µl) (Promega, Madison, Wisconsin, USA) in a total volume of 20 µl. The thermal cycling conditions were as follows: cDNA was denatured by a pre-incubation for 5 min at 95 °C; the template was amplified for 30 cycles of denaturation for 30 sec at 95 °C, annealing of primers at 60 °C programmed for 45 sec and extension at 72 °C for 45 sec, followed by a final extension at 72 °C for 7 min. Using the same reaction mixture composition and the same amplification conditions, the cDNA from control and infested samples was used as template for PCR reactions performed using the primers reported in table 3. Gene-specific primers were designed using the Prime3 software (<http://sourceforge.net/projects/primer3/>).

**Table 3.7.** List of primers used for aphid-derived cDNA amplification.

Primer Name	Primer Sequence	T° m	Product size
ATP synthase Fw	5'-TTCTGGCTTCAACACCCTCC-3'	59.89	89
ATP synthase Rv	3'-CCATGAGTTACTGGAGGGCC-5'	59.82	
RP_S18 Fw	5'-TGCTGACCACGTACTCTCAA-3'	58.68	85
RP_S18 Rv	3'-TGC GTGAAGATTTGGAAAGGT-5'	58.42	
RP_L29 Fw	5'-CTGTCCACCGCAGATCAGTT-3'	60.04	103
RP_L29 Rv	3'-TGGTTGCGGCTTTCTTCTTTC-5'	59.66	
RP_L37 Fw	5'-CATCTTACGCAGCGAGGCA-3'	60.52	94
RP_L37 Rv	3'-GGATTTACCGCCGTCAAAA-5'	59.41	
NADH dehydrogenase Fw	5'-GCTGGACTTACTGGAGGTTCT-3'	59.1	80
NADH dehydrogenase Rv	3'-AATCAAGAGCCGTGGGAAGA-5'	59.02	
ATPase Fw	5'-TGCTTCTTCTCCAACAACAGC-3'	59.05	112
ATPase Rv	3'-TGGTGAAGGAATGACTCGCA-5'	59.32	
Enolase Fw	5'-TCCTTTGTATCCAGCCTTTTCA-3'	57.62	82
Enolase Rv	3'-TGGTTTTGCTCCCAACATTCT-5'	58.32	
EF 1a Fw	5'-TGTCTCCAGCAACGAAACCA-3'	59.82	85
EF 1a Rv	3'-GAAGCTGTTCCCGGAGACAA-5'	59.97	
_TC17109 catepsina B Fw	5'-CATTCGTTTGTGCCTCGTCG-3'	60.18	108
_TC17109 catepsina B Rv	3'-CGGTTGGGGTGAACAATACG-5'	59.2	
_L16634 catepsina B Fw	5'-ACATTCATTTGTGCCTCGTCG-3'	59.54	102
_L16634 catepsina B Rv	3'-GGTGAAGAATACGGAACCCCA-5'	59.72	
Ef_Cpepo Fw	5'-ATTCGAGAAGGAAGCTGCTG-3'	60.2	129
Ef_Cpepo Rv	3'-TTGGTGGTCTCAAACCTCCAC-5'	59.8	

### 3.5.5 Aphid dispersal behaviour bioassays

The bioassays were designed in order to evaluate the biological performance/behaviour of *A. gossypii* after host plant conditioning by a previous aphid infestation or by application of methyl salicylate. Three week old zucchini plants were used for the bioassay, and each plant was individually confined in an insect-proof cage (30 x 30 x 30 cm) (Vermandel, The Netherlands).

The first basal leaf was infested by 50 3<sup>rd</sup> instar *A. gossypii* that were allowed to feed for 48 h. After aphids removal, a single pre-reproductive adult was placed on the same leaf, and daily checked for a total of 10 days. The bioassay was replicated on 10 pre-infested plants, and 10 controls (plants without previous infestation) were set up. After 10 days dispersal behaviour was assessed by counting the number of nymphs that remained fixed on the leaf where they were born or moved towards other feeding sites.

Methyl Salicylate (> 99% purity, Sigma-Aldrich) was applied as a pure compound in a Perspex cage perfectly sealed containing a single *C. pepo* plant, spotted on filter paper (10 µl), as described by Digilio and colleagues (2012). After 24 h, the filter paper was removed and the MeSA was allowed to diffuse out of the cage for 20 min. Then, a vial containing 20 pre-reproductive adult aphids was placed at the base of the plant, so that the aphids were able to climb the plant looking for a feeding site. Aphid dispersal behaviour was checked at 3, 24 and 48 h as “remained in the tube”, “wandering in the cage” and “feeding on the plant”. Overall 18 replicates were performed for each of MeSA and control treatment and for each time point. The distribution of aphids on different part of the plant were compared by G–test of independence (McDonald, 2014).

Chapter 4

*Identification of  
semiochemicals  
released by “San  
Pasquale” leaves  
upon infestation by  
Aphis gossypii*

Alessia Vitiello in collaboration with  
Toby J. A. Bruce and Rosa Rao



## Abstract

Plants are able to defend themselves against insect pests using different strategies, including production of herbivore induced volatile organic compounds (HI-VOCs). These molecules attract parasitic and predatory insects that are natural enemies of the herbivores. *Aphis gossypii* is a polyphagous aphid which feeds on several host plants, and it is considered the principal insect pest of zucchini (*Cucurbita pepo* L.). Volatiles were collected by air entrainment from zucchini cultivar “San Pasquale” foliage with and without *A. gossypii* infestation. When plants were infested with 10 adult aphids, for 48 h, there was a significant reduction in (*E*)-caryophyllene emission but emission levels of other volatiles were not significantly different. Conversely, a significant increase in (*E*)-caryophyllene emission was observed when plants were infested with 300 adult aphids, for 96 h. Olfactometer bioassays revealed that synthetic (*E*)-caryophyllene was attractive to female *Aphidius colemani* parasitic wasps. Taken together, these results suggest (*E*)-caryophyllene may play an important role in zucchini plant indirect defence responses because its increased emission from plants with substantial aphid infestation could attract parasitoid wasps. It is unclear why its emission was reduced with 10 aphids per plant. One possible explanation for reduced (*E*)-caryophyllene emission with low numbers of aphids was that the aphids suppressed plant defence but that defence suppression could not be maintained with higher numbers of aphids.

## 4.1 Introduction

Plants are exposed to a multitude of attackers and have evolved direct and indirect defence strategies in response to these organisms. Herbivore induced volatile organic compounds (HI-VOCs) released in response to insect feeding can directly repel phytophagous insects, and can indirectly act as attractive chemical signals for herbivore natural enemies (Heil, 2008; Tamiru *et al.*, 2011). This tritrophic interaction with beneficial insects, such as predators and parasitoids that feed on the insect pest and are attracted to HI-VOCs released from damaged plants, is a well-studied interaction (Heil 2008 and references therein). HI-VOCs released from the attacked plant parts can also act as phytohormones inducing defence responses in the non-attacked tissues of the same plant (e.g. Heil and Silva Bueno 2007) or of neighbouring plants (e.g. Baldwin *et al.*, 2006). Knowledge of volatiles involved in plant responses to insect feeding may allow improved biological control of many economically important crop pests.

Each plant species is able to synthesize a unique blend of volatiles, and variation in volatile profile emitted was even observed in individual plants of the same genotype grown under the same conditions (Webster *et al.*, 2010). Vegetative tissue usually releases a low level of volatiles, compared with the volatile organic compound (VOC) emission rates from flowers in the majority of flowering plants (Knudsen *et al.*, 2006), and with the production of leaf volatiles from herbs plant such as peppermint or basil (McConkey *et al.*, 2000; Vassao *et al.*, 2006). However, VOC emission can be highly induced by mechanical damage. VOC synthesis can also be influenced by environmental conditions and the circadian clock (Gershenzon *et al.*, 2000; Dudareva *et al.*, 2004). VOC biosynthetic rate is correlated with expression levels of genes coding for enzymes involved in their synthesis, and with the quantity of substrate available for these enzymes (Pichersky *et al.*, 2006).

In the past decade, several studies elucidated the genes, enzymes, and pathways activated for VOCs production. Numerous VOCs have been described, which belong to a few broad compound classes, including volatile terpenoids, volatile products of

shikimic acid pathway (phenylpropanoids, benzenoids, indole), amino acid-derived and fatty acid cleavage products (Niinemets *et al.*, 2013).

Although numerous studies have focused on the induction of volatiles in response to feeding by chewing insects, such as caterpillars (Alborn *et al.*, 1997; Hilker and Meiners, 2002; Schmelz *et al.*, 2006; Alborn *et al.*, 2007), recently, growing attention has been focused on plant volatiles released following infestation by phloem-feeding insects such as aphids. Aphids have been extensively studied with respect to predator and parasitoid attraction to olfactory cues emanating from their aphid prey, the host plant on which they feed, or the combination of the two. Most of these studies have measured the attraction of parasitoids (Guerrieri *et al.*, 1999; Lo Pinto *et al.*, 2004; Sasso *et al.*, 2007) or predators (Han and Chen, 2002; Francis *et al.*, 2004) to aphid-infested plants using wind tunnel and olfactometer bioassays. A growing number of studies have detected differences in the volatile compounds emitted from aphid-infested plants compared with un-infested control plants (Han and Chen 2002; Zhu *et al.*, 2005; Zhu and Park, 2005; Pareja *et al.*, 2009; Schwartzberg *et al.*, 2011) although it is recognized that HI-VOC production in response to aphid feeding is less than that in response to damage caused by chewing herbivores (Turlings *et al.* 1998).

Although considerable attention has been directed towards plant-insect interactions, almost all of these studies were performed using model plant species (e.g. *Arabidopsis*, tomato, broad bean) and analysing the effect of phloem-feeding insects (e.g. *M. persicae*, *A. pisum*, *M. euphorbiae*) on plant volatile emission.

To date, little has been done on the chemical ecology of *Cucurbita pepo* and associated sucking insect pests such as *Aphis gossypii*, and few studies on the identification of volatiles emitted from *Cucurbita pepo*, or from species belonging to *Cucurbita* genus, have been published.

Andrews and colleagues (2007) focused their work on floral volatiles emitted by *Cucurbita moschata* Duchesne (butternut squash) in the context of attracting pollinators and herbivores. The striped cucumber beetles, *Acalymma vittatum*, which feed on several cucurbits, is attracted to different volatiles emitted by *Cucurbita* blossoms. Because *Cucurbita* relies on pollinators for reproduction (McGregor 1976), the attraction of pollinators has fitness consequences. Each volatile tested showed different effects attracting just pollinators, just beetles, or both pollinators and beetles. The work just mentioned reported a new tool for plant breeders: selection of plants that produce less indole could reduce beetle population without the use of pesticides and limit damage to local bee populations (Andrews *et al.*, 2007). Further research on floral volatiles emitted by the wild cucurbit *C. pepo* subsp. *texana* was performed by Theis and colleagues (2009). Following leaf damage performed in a manner that mimicked beetle damage, increased volatiles production was detected only in male flowers. Female flowers which were bigger and produced more fragrance than males, were unaffected by leaf damage (Theis *et al.*, 2009). This study was the first to demonstrate a quantitative effect of mechanical damage on *C. pepo* floral scent.

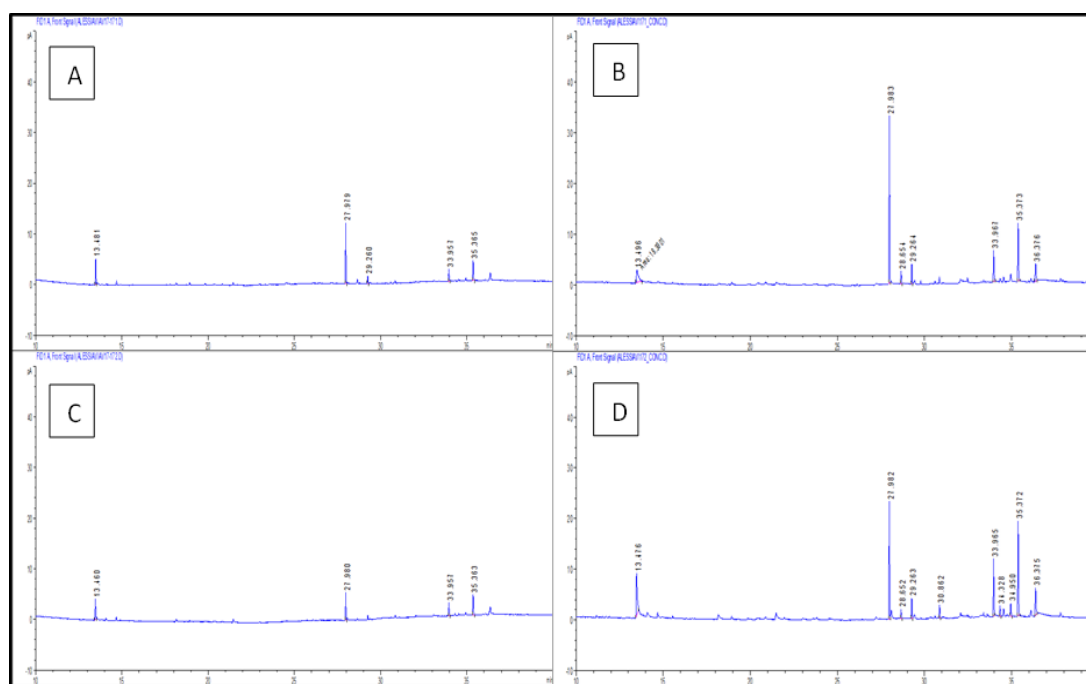
Finally, pathogens may influence volatile emission. For example, the effect of *Cucumber mosaic virus* (CMV) on the quality and attractiveness of *C. pepo* cultivar "Dixie" for two aphid vectors, *M. persicae* and *A. gossypii*, was documented. CMV is able to attract vectors deceptively to infected plants from which they then disperse rapidly, a pattern highly advantageous to the non-persistent transmission mechanism employed by CMV (Mauck *et al.*, 2014). *C. pepo* subsp. *texana* was also used to investigate the effect of the beetle-transmitted bacterial pathogen *Erwinia tracheiphila* and the *Zucchini yellow mosaic virus* (ZYMV) on foliar and floral volatile emission

(Shapiro *et al.*, 2012). Bacterial infection alters the foliage volatile emission in ways that attract vectors to infested plants to allow subsequent dispersal on healthy plants. In this study we have explored, for the first time, how aphid feeding affects volatile emission in *C. pepo* foliage. Volatiles were collected from zucchini plants after infestation with a low number (10 adult *A. gossypii*) and a high number of aphids (300 *A. gossypii*), and compared with emission levels from un-infested control plants. Qualitative and quantitative changes in VOCs emitted by zucchini cultivar “San Pasquale” plants after *Aphis gossypii* infestation were compared in order to improve our understanding of zucchini plant indirect defence responses.

## 4.2 Results

### 4.2.1 Volatile Organic Compounds produced by infested plants

Gas chromatography (GC) analysis of volatiles collected from un-infested plants and plants infested with aphids, revealed that there were only subtle changes in the production and release of volatile organic compounds (VOCs). Furthermore, leaves of control and infested plants only emitted low levels of VOCs (figure 4.1).



**Figure 4.1.** Gas chromatography analysis of VOC samples collected from (A) un-infested and (C) *A. gossypii* (10) infested zucchini plants. The same samples were analysed after concentration under a stream of nitrogen (B and D, respectively) performed for downstream analyses.

Analyses of samples collected on Tenax filters allowed identification of more peaks than samples collected on Porapak filters (Appendix figure A.6) although, even with Tenax collections, the volatile profile of zucchini plants infested with 10 and 300 aphids had only a few peaks.

GC/MS spectra acquired from VOCs collected using both Porapak and Tenax polymers were compared with select peaks shared between the two condition of infestation in order to obtain a list of compounds produced. In total, eleven main molecules emitted by zucchini plants were identified (table 4.1). Moreover, results of GC/MS analysis on VOCs collected with Tenax from control plant and plant infested

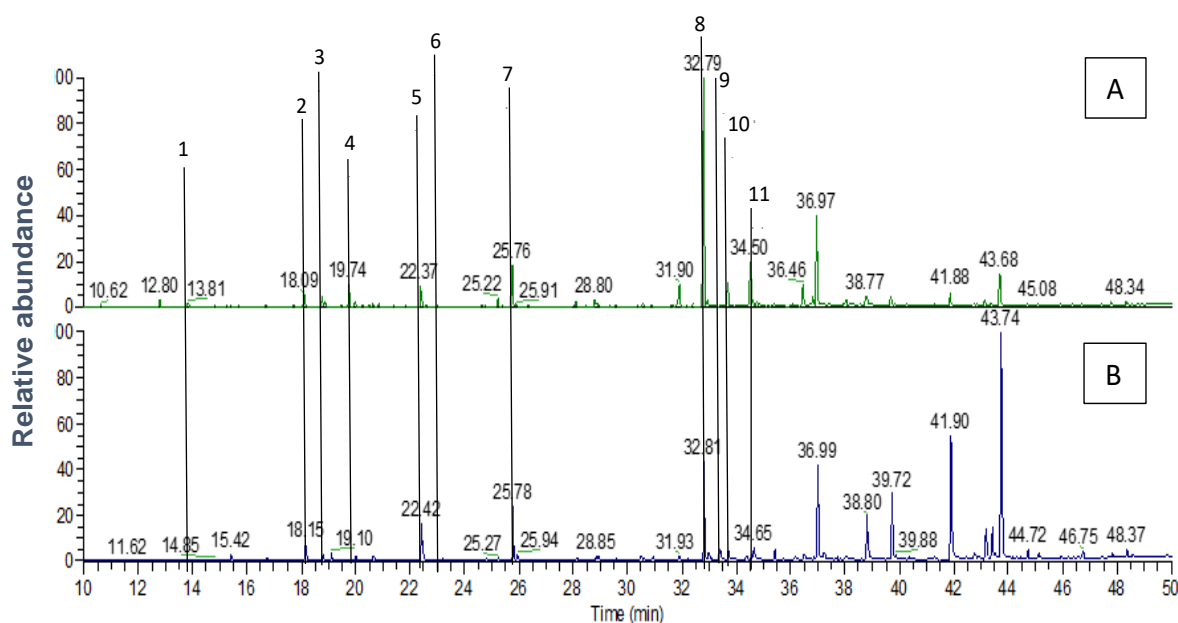
with 300 aphids were matched to highlight the identified peaks, as reported in figure 4.2 as example.

**Table 4.1.** Names and Kovats Index (KI) of volatile organic compounds collected from *A. gossypii*-infested and un-infested zucchini plants.

# Peak	Compound ID	ExpKI*	KI**
1	heptanal	883	883
2	6-methyl-5-hepten-2-one	967	966
3	3-octanone	971	968
4	acetophenone	1035	1040
5	nonanal	1084	1084
6	undecene	1090	1100
7	decanal	1187	1186
8	(E)-caryophyllene	1425	1432
9	E-beta-farnesene	1448	1450
10	humulene	1458	1465
11	beta-selinene	1488	1489

\*experimental KI

\*\*Kovartz Index



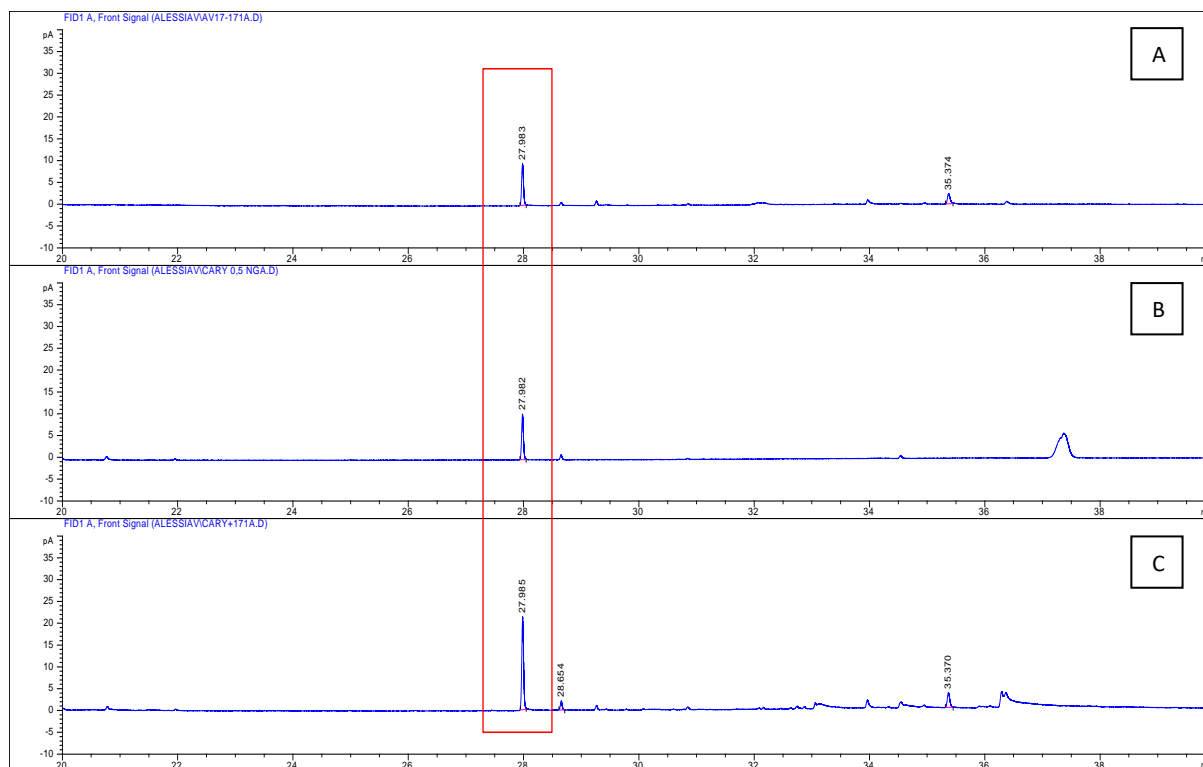
**Figure 4.2.** Chromatograms of volatile emissions (24 h collection by air entrainment of headspace) from (A) un-infested zucchini plants and (B) plants infested with *A. gossypii* for 96 h. Numbers indicate the identified peaks (see table 4.1).

The volatiles identified were emitted from both control and infested plants in all experimental treatments analysed. The only exception was (E)- $\beta$ -farnesene, which was only detected in samples collected from leaves infested with 300 aphids after 96 h and 7 days from the onset (leaves were entrained for 24 h and 3 days, respectively). Several differential peaks recorded between samples collected from control and infested plants with high retention times (figure 4.2, right part) were identified as contaminants (e.g. siloxane, benzene, etc.).



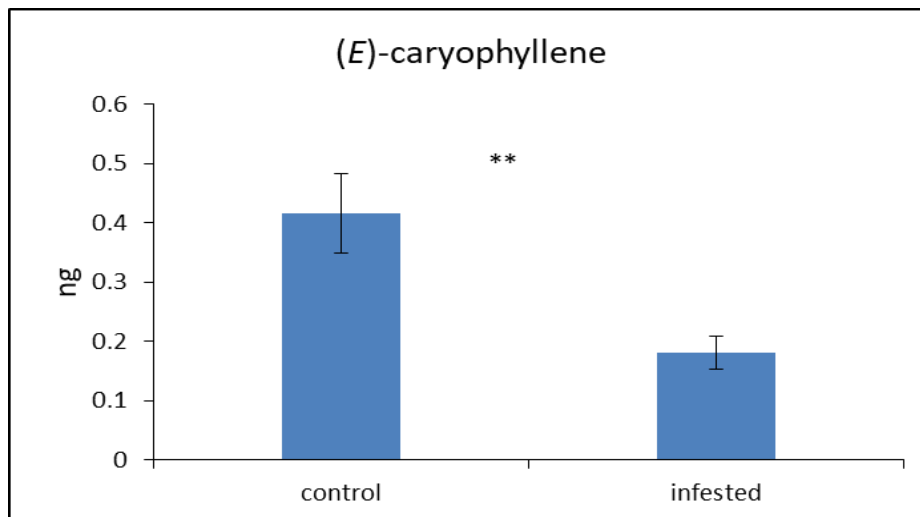
To confirm the identity of volatiles emitted by zucchini leaves in our experimental conditions, co-injection experiments were performed. Identifications of (*E*)-caryophyllene (figure 4.3), humulene, nonanal, decanal, 6-methyl-5-hepten-2-one and 3-octanone were all successfully confirmed by peak enhancement with authentic chemical standards.

Comparative analysis of VOCs obtained from undamaged samples and volatiles samples collected from plants infested with *A. gossypii* showed that monitoring quantitative change in total volatile emission was impossible. However, we observed significant qualitative differences relating to the amount of (*E*)-caryophyllene produced in both experimental conditions.



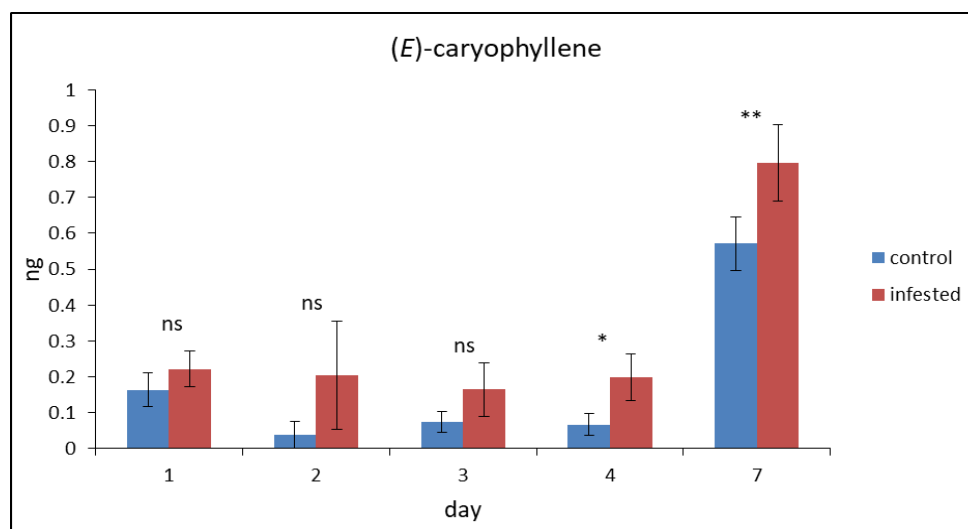
**Figure 4.3.** Chromatograms obtained after injection on HP-1 column of (A) air entrainment sample, (B) (*E*)-caryophyllene authentic standard and (C) co-injection of air entrainment sample and chemical standard. The red box highlights the (*E*)-caryophyllene peak.

(*E*)-caryophyllene emission was significantly ( $P = 0.003$ ) reduced in “San Pasquale” plants infested by 10 adult *A. gossypii* per 48 h ( $0.18 \pm 0.07$  ng) when compared with un-infested plants ( $0.41 \pm 0.17$  ng), (figure 4.4).



**Figure 4.4.** Quantitative differences in (*E*)-caryophyllene emission by zucchini control plants and infested with 10 *Aphis gossypii*. Each bar shows the mean quantity ( $\pm$  SE) collected from 7 replicate plants, expressed in ng. Data were analysed using a paired *t*-test with statistical software (Genstat Version 17, VSN International Ltd.). ns: no significant difference; \*  $P < 0.05$ ; \*\* $P < 0.01$ .

No significant increase in emission of (*E*)-caryophyllene was observed comparing control plants and zucchini plants infested with 300 aphids at 24, 48 and 72 h following infestation, although there was a trend for increased emission by aphid infested plants. However, zucchini plants infested with 300 aphids showed a significant increase in (*E*)-caryophyllene production when compared with control plants starting from the fourth day of infestation (figure 4.5). Specifically, the quantity of (*E*)-caryophyllene emitted by damaged plants ( $0.19 \pm 0.06$  ng) after 96 h from infestation was three times higher than the quantity emitted by control plants ( $0.06 \pm 0.03$  ng). The (*E*)-caryophyllene amount collected between 4<sup>th</sup> and 7<sup>th</sup> day from infested plants ( $0.57 \pm 0.07$  ng) was significantly ( $P=0.013$ ) higher than that from undamaged plants ( $0.79 \pm 0.1$  ng).

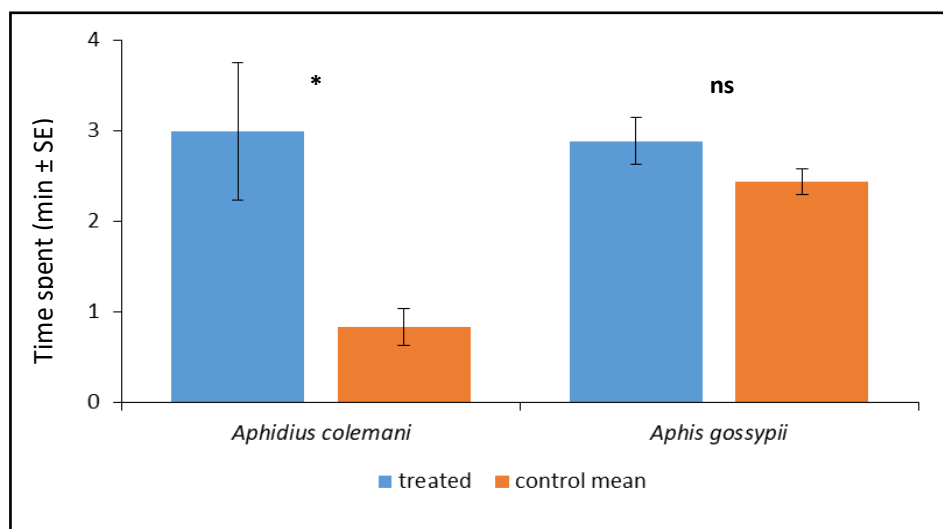


**Figure 4.5.** Quantitative differences in (*E*)-caryophyllene emission by zucchini control plants and infested with 300 *Aphis gossypii*. Each bar shows the mean quantity ( $\pm$  SE) collected from 7 replicate plants, expressed in ng. Data were analysed using a paired *t*-test with statistical software (Genstat Version 17, VSN International Ltd.). ns: no significant difference; \*  $P < 0.05$ ; \*\* $P < 0.01$ .

### 4.2.2 Behavioural response of aphids and parasitoids to synthetic (*E*)-caryophyllene

In order to evaluate the behavioural response of the cotton aphid *Aphis gossypii* and of the parasitic wasp *Aphidius colemani* to (*E*)-caryophyllene, four-arm olfactometer bioassays were performed. Insect behaviour was observed in response to synthetic (*E*)-caryophyllene, and the time spent within the arm containing the (*E*)-caryophyllene was compared with the average time spent in control arms (containing hexane solvent). The aphids did not show any significant preferences for treatment tested.

*A. gossypii* spent more time in the treated region of the olfactometer ( $2.88 \pm 0.26$  min) than in the control region ( $2.44 \pm 0.14$  min), but this difference was not significant ( $P = 0.119$ ). Conversely, *A. colemani* parasitoid females were significantly attracted by (*E*)-caryophyllene ( $P = 0.019$ ), spending more time in the treated ( $2.99 \pm 0.86$  min) than in the control sections ( $0.83 \pm 0.20$  min), (figure 4.6).



**Figure 4.6.** Behavioural responses of winged *Aphis gossypii* and parasitoid females *Aphidius colemani* in four-arm olfactometer ( $n = 12$ ) to synthetic (*E*)-caryophyllene. Data are expressed as the mean ( $\pm$  SE) time (min) spent in treatment and control arms and were analysed using a paired *t*-test with statistical software (Genstat Version 17, VSN International Ltd.). ns: no significant difference; \*  $P < 0.05$ ; \*\* $P < 0.01$ .

### 4.3 Discussion

In this study we demonstrated that aphid infestation induced an unexpected change in volatile emission of zucchini plants: at low infestation levels there was suppression of (*E*)-caryophyllene emission, whereas at high infestation levels emission of this compound was significantly induced. Volatiles were collected from control plants and plants infested by *A. gossypii* with different infestation rates. Specifically, two different infestation treatments, with 10 and 300 aphids, were used to identify differences in volatile emission rates.

The first observation was the low number and quantity of volatiles produced by uninfested plants, common to both experimental conditions. In healthy plants, vegetative tissues generally only release a small quantity of VOCs when compared with flowers (Pichersky *et al.*, 2006). However, in all plant organs VOC production and emission are developmentally regulated and show an increase during early stage of development (e.g. young and not fully expanded leaves) (Dudareva *et al.*, 2013). Volatile collection from “San Pasquale” plants was performed 3 weeks after

sowing. The volatile profile of zucchini plants infested with 10 adult aphids remained quite similar to, or even less than that of control plants.

With the exception of (*E*)-caryophyllene, our results showed that there was no difference in emission of volatile compounds, suggesting that zucchini plants do not produce HI-VOCs in response to melon aphids, in contrast to other plants fed upon by aphids (Han and Chen, 2002; James, 2003; Zhu and Park, 2005; Harmel *et al.*, 2007). Even if no studies on volatile emitted by aphid-infested zucchini plants have been published, our results were inconsistent with results obtained in available researches on pathogen-infested *C. pepo* plants (Shapiro *et al.*, 2012; Mauck *et al.*, 2014). Pathogen- and virus-infected plants showed an increase in volatile emission of infected leaves compared to healthy plants, which elicited positive behavioural responses from insect vectors. Moreover, previous studies on cucumber plants have shown that spider mite (*Tetranychus urticae*) infestation induced the emission of volatiles that attract the predatory mite *Phytoseiulus persimilis*, a major natural enemy of the spider mites (Kappers *et al.*, 2011). These differences in zucchini plant volatile emission could be the result of differential responses related to plant variety, research conditions, or aphid strain. The influence of plant cultivar and herbivore species on plant VOC response was effectively demonstrated (De Moraes *et al.*, 1998; Gouinguene *et al.*, 2001; Degen *et al.*, 2004). Moreover, it could be hypothesized that aphids cause relatively little damage to plant tissue compared with chewing herbivores, causing less induction of damage related volatiles.

The low level of HI-VOC production obtained in our experimental conditions could be related to aphid density (10 aphids). It has been hypothesised that aphids are somehow able to counteract plant defences as they feed. A recent study published by Schwartzberg and colleagues (2011) reported a negative effect on volatile emitted by broad bean plants infested by *Acyrtosiphon pisum*. They compared the VOCs emission rate of *Vicia faba* when infested with beet armyworm caterpillars with aphids, and the effect of VOC emission caused by both insects. They reported the ability of aphids to actively suppress VOC emission (Schwartzberg *et al.*, 2011). Moreover, they found that aphid feeding had a negative effect on emission of terpenoids compounds (*E*)-beta-ocimene and (*E*)-caryophyllene. The latter was also found greatly affected in our study. Among VOCs identified, (*E*)-caryophyllene was the only compound for which it was possible to find significant changes in the amount emitted in both control and infested samples. Surprisingly, a significant reduction in emission was observed in samples collected from foliage of aphid-infested plants. To verify the hypothesis of VOC suppression, and the possible relation with aphid density, a new infestation experiment was performed. Three-weeks old "San Paquale" plants were infested with 300 *A. gossypii* for a total of 7 days, and volatile samples were collected at five different time points: 24 h, 48 h, 72 h, 96 h, 7 days. Interestingly, even in this set of experiments, few changes in volatile profiles from infested plants were observed. However, there were significant changes in emission of (*E*)-caryophyllene which showed an increase in emission in plants infested with high concentration of aphids. One possible explanation for this is that defence suppression could not be maintained with higher numbers of aphids.

(*E*)-caryophyllene is a bicyclic sesquiterpene derived from isoprenoid pathways. In the biosynthesis of terpenes, the large class of terpene synthase (TPSs) enzymes converts linear prenyl diphosphate precursors into the large diversity of terpene skeletons encountered in plants. Many different TPSs have been characterized from various plant species. For example, the Arabidopsis genome contains a total of 32 intact sequences coding for TPS, which are expressed in specific plant organs and

developmental stages (Chen *et al.*, 2003). The maize Terpene Synthase 23 (TPS23) was identified as directly involved in cyclization of farnesyl diphosphate to (*E*)- $\beta$ -caryophyllene and it is regulated in leaves and roots in response to damage by different herbivores (Köllner *et al.*, 2008). Moreover, (*E*)-caryophyllene can function as chemical signal involved in attraction of natural enemies in broad bean, maize and tomato (Du *et al.*, 1998; Köllner *et al.*, 2008; Sasso *et al.*, 2007), associated with other compounds such as 6-methyl-5-hepten-2-one, contributing to the plant defence against herbivores with completely different sites and modes of attack.

Moreover, (*E*)- $\beta$ -farnesene was identified only in infested samples collected from plants infested with aphids for a longer time. Specifically, this compound was found in headspace samples collected from “San Pasquale” leaves after 96 h and 7 days from the beginning of infestation. Unlike the previous experiments (with 10 aphids), aphids were left to feed on plants for the entire time span. For this reason, (*E*)- $\beta$ -farnesene could be emitted by both aphids and plants. This compound is the key or only component of the aphid alarm pheromone which play an important role in mediating interaction between aphids and environment (plant, aphids, natural enemies, etc.) (Vandermoten *et al.*, 2012). Furthermore, (*E*)- $\beta$ -farnesene is also a widely occurring plant volatile (Webster *et al.*, 2008). The wild potato *Solanum berthaultii*, which produces (*E*)- $\beta$ -farnesene in glandular trichomes, is more repellent to the green peach aphid than are commercial varieties of *S. tuberosum* (Gibson and Pickett, 1983). No other differences were observed between 10 and 300 aphid-infested zucchini plants regarding quantity and quality of volatiles emitted.

Taken together, the observed effects of *A. gossypii* infestation on zucchini foliar volatile emissions suggest a relevant role of (*E*)-caryophyllene mediating plant-aphid interaction. As reported above (*E*)-caryophyllene is a well-studied sesquiterpene, involved in attraction of natural enemies of herbivore pests. To test the putative role of (*E*)-caryophyllene in both aphid host location and attraction of aphid natural enemies, four-arm olfactometer bioassays tested the attraction of *A. gossypii* winged forms and female parasitic wasp *Aphidius colemani* to synthetic (*E*)-caryophyllene. Olfactometer data confirmed that *A. colemani* was attracted by this volatile. Conversely, no effect was observed for this sesquiterpene on *A. gossypii*. Thus, these results are consistent with the hypothesis of parasitoid attraction.

#### 4.4 Conclusions

In summary, our results show that small amounts of volatiles are emitted by zucchini cv “San Pasquale”, and that aphid feeding does not substantially affect zucchini plant volatile emission as reported for several other plant species. However significant differences in emission of (*E*)-caryophyllene were observed, which was influenced by *A. gossypii* density on plants. (*E*)-caryophyllene could be considered a key volatile in our experimental system, as suggested by olfactometer results. However, further experiments will be necessary to assess the attractiveness of plant volatile blends towards parasitoids. In addition, coupled GC-EAG (Gas Chromatography-Electroantennography) analysis using the antennae of *A. gossypii* and of the females *A. colemani* will be important to reveal the possible perception of a number of compounds in the *A. gossypii* infested zucchini VOC samples and to identify the biologically active ones.

## 4.5 Materials and Methods

### 4.5.1 Biological materials

Seeds of the aphid-susceptible *Cucurbita pepo* cultivar “San Pasquale” were grown following the same procedure as reported in Chapter 2. The melon/cotton aphid *Aphis gossypii* Glover (Homoptera: Aphididae), derived from the aphid rearing maintained at the Department of Agricultural Sciences, University of Naples Federico II, was reared on *C. pepo* cv “San Pasquale” plants in a Perspex cage in an insectary (22 °C; 16:8 h D:L regime) at Rothamsted Research, Harpenden, Hertfordshire, UK.

*Aphidius colemani* parasitic wasps were purchased from Agralan (UK). After emergence, parasitoids were retained for 24 h and supplied with water and sugar. Female parasitic wasps were then selected and were used for behavioural assays.

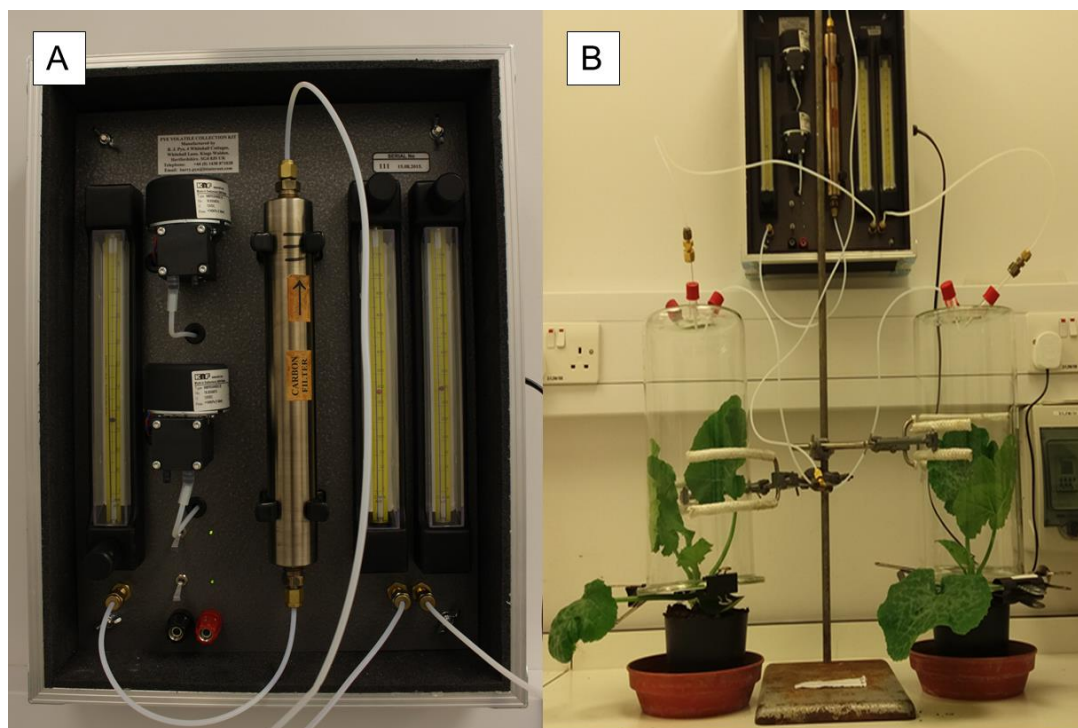
### 4.5.2 Air Entrainment of zucchini plants

Two adsorbent polymers were used for VOC collection: Porapak Q and Tenax TA. Porapak Q has a high capacity for small molecules which are collected by solvent elution. If samples are required for repeat analyses, such as GC, GC/MS and co-injection, or bioassay, such as EAG and olfactometer, Porapak entrainments should be used. The main disadvantage related to Porapak polymer is the low sensitivity to compounds produced in small quantities and with short retention times. If small amounts of volatiles are emitted or short entrainment times are required, then Tenax filters should be used. Thermal desorption allows higher sensitivity, even if stability of some analytes may be affected by temperature. Samples collected using Tenax polymer are immediately and completely analysed.

All equipment was washed with Teepol detergent (Herts Country Supplies, Herts, UK), acetone and distilled water, and baked at 160 °C for a minimum of 2 hr. Porapak Q adsorbent polymer (50 mg, 50-80 mesh, Supelco Inc., Bellefonte, P.A., USA) contained inside a 5 mm diameter glass tube was held with two plugs of silanised glass wool to form “Porapak tubes”. Similarly, Tenax TA polymer (50 mg, 60/80 mesh, Supelco Inc., Bellefonte, P.A., USA) contained inside a 5 mm diameter glass tube (GC liner) was held with two plugs of silanised glass wool to form “Tenax tubes”. These tubes were conditioned before use by washing with redistilled diethyl ether (1 ml) and heating at 132 °C and 220 °C, respectively for Porapak and Tenax, under a stream of nitrogen for 2 hours to remove all contaminants.

Volatile organic compounds produced by 3-week-old intact and *A. gossypii*-infested “San Pasquale” plants were collected enclosing zucchini leaves in glass vessels (figure 4.7, B). The first set of collections was from plants that had been infested with 10 adult *A. gossypii* for 48 h. Aphids were confined to the first real leaf, using a clip cage, and after the 48 h period were manually removed using a paintbrush. This time point was selected because after 48 h the largest number of DEGs were identified (Chapter 3), which, possibly, could be related to differences in emission of defence volatiles. Nevertheless, few number of DEGs associated with VOCs emission was identified. Infested and un-infested plants were then individually enclosed in glass vessels, excluding the infested leaf. VOCs were collected for a period of 24 h following removal of aphids. In the second set of collections, whole plants were infested with 300 aphids (mixed ages) and were entrained for a total of 7 days.

Adsorbent tubes were changed at 24 h intervals for the first 4 days. Then, a three-day collection was performed between the 4<sup>th</sup> and the 7<sup>th</sup> day, and a total of five samples were collected for each replicate. Foliage of both infested and un-infested plants was placed in glass vessels (30 cm high x 15.5 cm internal diameter) open at the bottom, and closed with three collection ports at the top (one for inlet of the air and two for outlet). The bottom was closed using two semi-circular aluminium plates that fitted around the plant stem. Air, purified by passage through an activated charcoal filter, was pumped in at 600 ml/min and drawn out at 400 ml/min through Poparak Q and Tenax TA tubes (figure 4.7, A).



**Figure 4.7.** Setup of air entrainment kit (A) and representation of the dynamic headspace collection of zucchini plants volatiles (B).

The difference in flow rates created a positive pressure preventing contaminated air from entering the collection vessel. The volatiles collected on Porapak Q were eluted with 500  $\mu\text{l}$  redistilled diethyl ether and stored at  $-20\text{ }^{\circ}\text{C}$  until required for following analysis. Tenax TA tubes, after headspace collection, were stored in sealed ampoules containing nitrogen and stored at  $-20\text{ }^{\circ}\text{C}$ . A total of seven biological replicates for each condition described above were used for downstream analyses.

#### 4.5.3 Gas Chromatography (GC) analysis

Volatile samples collected from Porapak Q filter tubes (4  $\mu\text{l}$  portions) were analysed on an Agilent 7890A GC equipped with a cool on-column (COC) injector, flame ionization detector (FID), a non-polar HP-1 bonded phase fused silica capillary column (50 m length x 0.32 mm inner diameter x 0.52  $\mu\text{m}$  film thickness, J & W Scientific) and a polar DB-WAX column (50 m length x 0.32 mm inner diameter x 0.52  $\mu\text{m}$  film thickness, J & W Scientific). The initial oven temperature was  $30\text{ }^{\circ}\text{C}$  for 0.5 min, programmed at  $5\text{ }^{\circ}\text{C}/\text{min}$  to  $150\text{ }^{\circ}\text{C}$  and held for 0.1 min, then programmed at  $10\text{ }^{\circ}\text{C}/\text{min}$  to  $230\text{ }^{\circ}\text{C}$  for 27 min for a 60-minute run time. The carrier gas was hydrogen. Volatiles collected using Tenax tubes were separated using an Agilent 6890N GC equipped with an integrated system including an Optic 2 programmable injector (ATAS GL International, Eindhoven, The Netherlands) for Thermal Desorption, a split/splitless injector, a non-polar HP-1 capillary column (50 m length x 0.32 mm inner diameter x 0.52  $\mu\text{m}$  film thickness, J & W Scientific) and a flame ionization detector (FID). The GC initial oven temperature was maintained at  $30\text{ }^{\circ}\text{C}$  for 1 min, programmed at  $5\text{ }^{\circ}\text{C}/\text{min}$  to  $150\text{ }^{\circ}\text{C}$  and held for 0.1 min and then programmed at  $10\text{ }^{\circ}\text{C}/\text{min}$  to  $250\text{ }^{\circ}\text{C}$  for 20 min. Total run time was 55 minutes, and the carrier gas was hydrogen. Samples were analysed by thermal desorption using an Optic unit programmed to start at  $35\text{ }^{\circ}\text{C}$ , then rise to  $250\text{ }^{\circ}\text{C}$  at  $16\text{ }^{\circ}\text{C}/\text{sec}$ .

Quantification of (*E*)-caryophyllene emitted was carried out on a non-polar HP-1 GC column, using a multiple point external standard method. A calibration curve (peak area vs concentration) was made using 0.1, 0.5, 1, 2 and 10  $\text{ng}/\mu\text{L}$  concentration ( $r^2=1$ ). Statistical analysis of quantity data from uninfested and infested plants was performed using a paired *t*-test function available in the GenStat software (GenStat® 2014, Seventeenth Edition, VSN International Ltd., Hemel Hempstead, UK).

#### 4.5.4 Coupled gas chromatography-mass spectrometry (GC/MS)

Tentative identification of compounds emitted by “San Pasquale” plants was achieved by coupled gas chromatography-mass spectrometry (GC/MS). Air entrainment samples collected by elution from Porapak Q filter tubes were concentrated to about 50  $\mu\text{l}$ , and 4  $\mu\text{l}$  aliquots of samples were analysed using a Micromass Autospec Ultima magnetic sector mass spectrometer, attached to an Agilent

6890N GC, equipped with a cool-on-column injector and a non-polar HP-1 capillary column (50 m length x 0.32 mm inner diameter x 0.52  $\mu\text{m}$  film thickness, J & W Scientific). Ionization was performed by electron impact (70 eV, 220  $^{\circ}\text{C}$ ). The oven temperature was maintained at 30  $^{\circ}\text{C}$  for 5 min and then programmed at 5  $^{\circ}\text{C}/\text{min}$  to 250  $^{\circ}\text{C}$ , for a 70-minute run time. Coupled gas chromatography-mass spectrometry analyses of Tenax filter tubes were performed on a Thermo Finnigan Mat95 xp magnetic mass spectrometer, equipped with a PTV unit (ATAS GL International, Eindhoven, The Netherlands) and a Thermo Finnigan Trace GC, fitted with a non-polar HP-1 column (50 m length x 0.32 mm inner diameter x 0.52  $\mu\text{m}$  film thickness, J & W Scientific). Ionization was by electronic impact (70 eV, 220  $^{\circ}\text{C}$ ). The GC oven temperature was programmed at 30  $^{\circ}\text{C}$  for 5 min, then programmed at 5  $^{\circ}\text{C}/\text{min}$  to 250  $^{\circ}\text{C}$ . Samples were analysed by thermal desorption (PTV unit programmed to start at 30  $^{\circ}\text{C}$  and then rise to 250  $^{\circ}\text{C}$  at 16  $^{\circ}\text{C}/\text{sec}$ ).

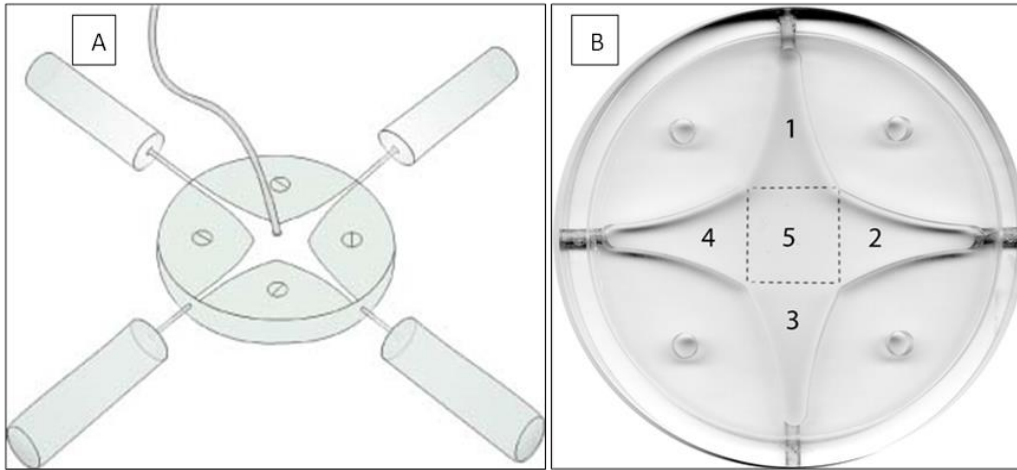
The peak fragmentation profile generated is highly characteristic and indicative of the original parent molecule, by examining the isotopic ratios, distribution and composition of these ion fragments, detailed chemical information regarding the structure and functional group present can be obtained. Tentative identifications of obtained peaks were made by comparison of peak fragmentation profile with those of authentic samples in the mass spectral database NIST (National Institute of Standards and Technology, 2011) or by spectra interpretation. Tentative identifications were confirmed by co-injection of the air entrainment sample with authentic standards on both non-polar HP-1 and polar DB-WAX GC columns, with peak enhancement indicating co-elution. Co-injection was performed for (*E*)-caryophyllene, humulene, nonanal, decanal, 6-methyl-5-hepten-2-one and 3-octanone.

#### 4.5.5 Four-arm olfactometer bioassays

A Perspex four-arm olfactometer (Pettersson, 1970) was used to determine behavioural responses of winged *A. gossypii* and parasitoid female *A. colemani* to (*E*)-caryophyllene. The olfactometer was a star-shaped arena consisting of four regions among which the insect was able to move freely. Each arm had an area of 6.2  $\text{cm}^2$  and the distance from one end of the olfactometer to the opposite end was 12 cm. Cylindrical plastic vials were connected to each of the four outer regions and used to contain the compound tested (figure 4.8, A). Perspex components were washed with Teepol solution, rinsed with 80% ethanol solution and distilled water and left to air dry overnight. The olfactometer was fitted with a filter-paper base to provide purchase for the walking insect and was illuminated from above by diffuse uniform lighting from a 18 W/35 white fluorescent light screened with greaseproof paper. It was also surrounded by black paper to remove any external visual stimuli. One-hundred ng of (*E*)-caryophyllene was tested; 10  $\mu\text{l}$  of a 10 ng/ $\mu\text{l}$  (*E*)-caryophyllene solution was pipetted onto a piece of filter paper, allowed 30 s for the solvent to evaporate, and then placed in the treated arm. Three control arms were tested, each contained a piece of filter paper with 10  $\mu\text{l}$  of redistilled hexane. A single insect was introduced through a hole in the top of the olfactometer using a fine paintbrush. Air was then drawn through the central hole at a rate of 400 ml/min and subsequently exhausted from the room. Each aphid was observed for a total of 12 minutes and the olfactometer was rotated of 90 degrees every 3 minutes to reduce any directional bias. Insects were allowed to move freely within the olfactometer for the duration of the experiment. The olfactometer was divided into five areas (figure 4.8, B) and the amount of time spent in each of the regions was recorded using the Olfa Software (F. Nazzi, Udine, Italy). A positive response to the target compound occurred when an insect spent significantly more time in the treated than in the control areas. Conversely, a negative response occurred when an insect spent significantly less time in the treated than in the control areas.

Statistical analyses were performed using a paired *t*-test function available in the GenStat software (GenStat $^{\circ}$  2014, Seventeenth Edition, VSN International Ltd., Hemel Hempstead, UK).





**Figure 4.8.** Schematic representation of four-arm olfactometer system (A) and internal areas division (B).





Chapter 5

*General conclusions*



Despite their economic importance, information on the molecular recognition and response of cucurbits, in particular, of zucchini plants to aphids is scarce. The previous chapters have shown that *Cucurbita pepo* is an interesting plant species to be investigated in order to understand the regulatory mechanisms of aphid induced defence response.

Our study provided a newsy overview of zucchini direct and indirect response during early stages of compatible interaction with *Aphis gossypii*. Illumina RNA-sequencing was adopted to identify main molecular players involved in zucchini-aphid recognition and interaction. In the first place, *C. pepo* transcriptome was *de novo* assembled from aphid-challenged leaf tissue (**Chapter 2**). We surveyed the poly (A)<sup>+</sup> transcriptome of zucchini at a remarkable depth and produced 71,648 assembled transcripts with 51,398 sequences successfully annotated via the Blast2GO suite (Götz et al., 2008). The assembly provided an exhaustive enough coverage to discover genes of major metabolic pathways, contributing to existing sequence resources for zucchini plant (Blanca *et al.*, 2011; Wyatt *et al.*, 2015; Xanthopoulou *et al.*, 2016). The features of the transcriptome we assembled were comparable with those of already available *C. pepo* transcriptomes. Furthermore, a sequence-based comparison revealed the presence of novel transcripts with functions associated with plant stress perception, signalling and response such as oxidative stress and Ca<sup>2+</sup>-dependent cell signalling. Identification of novel transcripts involved in stress response mechanisms provides an additional value to this resource. This study confirmed that Illumina sequencing technology could be applied as a rapid and cost-effective method for assembly and analysis of transcriptome in non-model species that lack a reference genome (e.g. Shi *et al.*, 2011).

We believe that, following publication in public repository, this transcriptome will serve as an important information platform to accelerate research on gene expression, genomics and functional genomics in *C. pepo*.

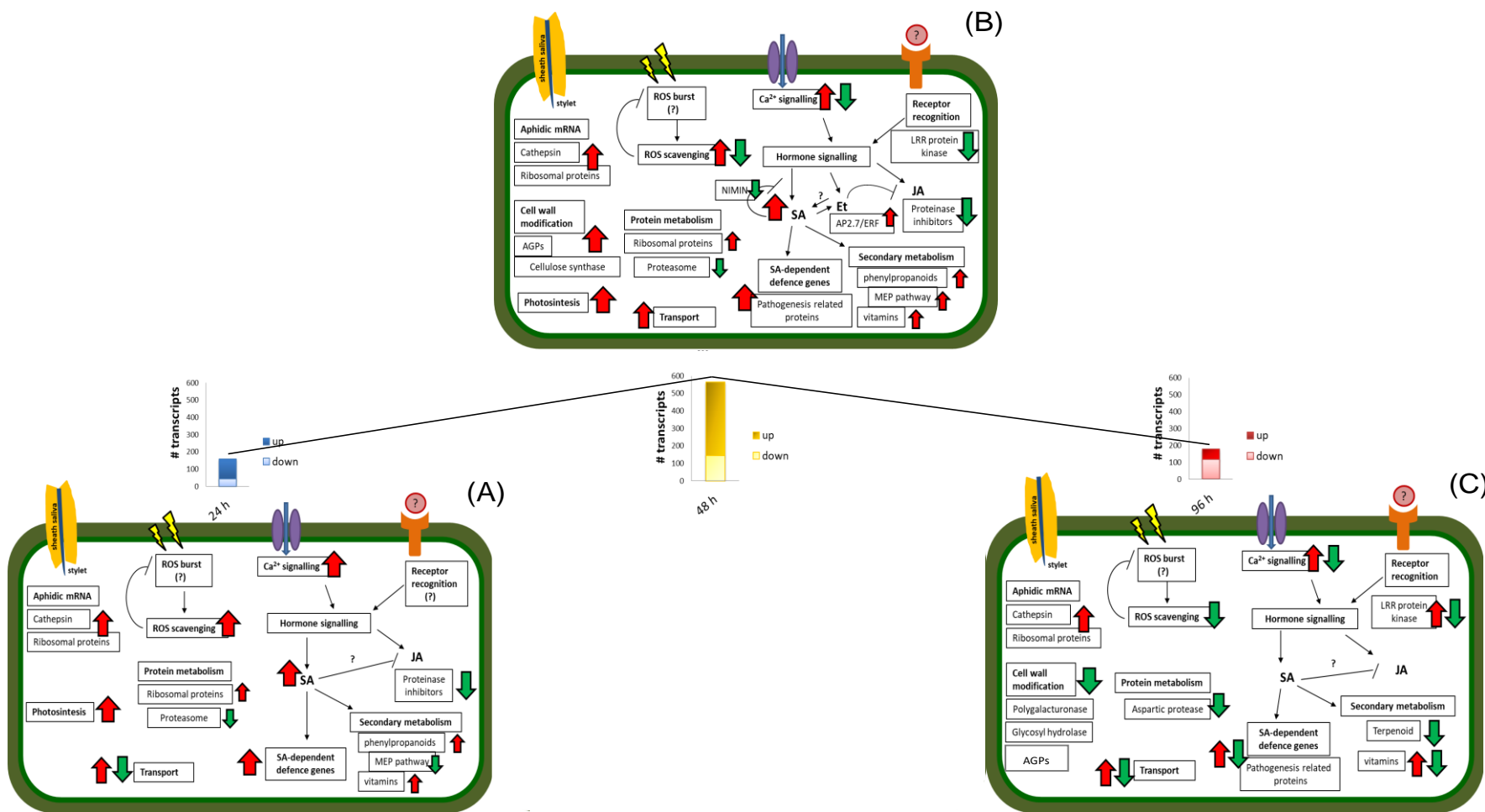
The high quality reference transcriptome for the zucchini cultivar "San Pasquale" was characterised with the aim of improving the knowledge on expressed genes modulated following aphid infestation.

Aphid feeding behaviour on zucchini plants lead to complex pattern of gene expression in which both proteins directly involved in defence (such as proteins related to pathogenesis or protease inhibitors) and proteins involved in cell wall strengthening, in oxidative stress, in the biogenesis of primary and secondary metabolites and in photosynthesis were observed (**Chapter 3**). The transcriptomic reorganization following aphid infestation at the three time points is illustrated in figure 5.1. The observed trends in gene expression may represent a model for *C. pepo* molecular responses to aphid. Indeed, we observed an increase in the number of DEGs from 24 to 48 hpi. At this time point the number of DEGs picked and then fall at 96 hpi, where more than 65% of genes was down-regulated (figure 5.1).

In our experimental conditions zucchini plant perception of aphid feeding lead to activation of Ca<sup>2+</sup>-dependent signalling and expression of ROS scavenging enzymes. At 24 hpi (figure 5.1, A), SA biosynthesis and SA-mediated defence response genes were up-regulated, whereas genes belonging to JA-mediated response (serine proteinase inhibitors) were down-regulated. Primary metabolic processes, such as photosynthesis and protein synthesis, were not negatively influenced by aphid feeding. On the other hand, secondary metabolic processes such as MEP pathway, phenylpropanoid pathway and vitamin biosynthesis were affected. Moreover, we identified aphid-derived transcripts injected in plant tissue during feeding process among overexpressed genes, since the first time point analysed.

At 48 hpi (figure 5.1, B), expression of genes related to  $\text{Ca}^{2+}$  signalling and ROS detoxification start to be also negatively influenced by aphid infestation. Moreover, genes encoding for LRR receptor-like protein kinases were found down-regulated, probably related to aphids' effort to suppress salivary effectors recognition. On the other hand, overexpression of several genes encoding for pathogenesis related proteins reinforce the role of SA in zucchini response. Furthermore, a transcript annotated as NIMIN1, a negative regulator of SA-mediated defence response (Weigel *et al.*, 2005), was down-regulated. An ethylene-responsive transcription factor (AP2-7/ERF) was found up-regulated in aphid-challenged plants and could contribute to the negative regulation of jasmonate signals (Shoji *et al.*, 2000). At 48 hpi biosynthesis and modification of cell wall components as direct defence to phloem-feeders were also activated. Interestingly, all the aphid mRNA identified were overexpressed at this time point. After 96 h (figure 5.1, C), the majority of pathway previously activated was down-regulated. Number of genes associated with ROS and  $\text{Ca}^{2+}$  signalling was reduced, while LRR receptor-like protein kinases were both up- and down-regulated. This response could indicate the adaptation of the plant to the infested conditions. At this stage, in fact key defence molecules have been produced and they try to serve as protecting agents. Genes involved in cell wall degradation were down-regulated, highlighting the importance of cell wall reinforcement in zucchini defence response to *A. gossypii*. Among down-regulated genes involved in secondary metabolism a gene encoding for a terpene synthase was strongly influenced. This result allows us to link transcriptomic data with results from the analysis performed on VOCs emitted by aphid-infested zucchini plants (**Chapter 4**). Indeed, "San Pasquale" plants infested for 48 h with 10 adult aphids showed a reduced emission of sesquiterpene (*E*)-caryophyllene compared with control plants. This compound is known to be involved in attraction of natural enemies of herbivore pest in several plant species. Among the identified VOCs, (*E*)-caryophyllene was the only one with significant changes in the amount emitted. Moreover, an increase in emission of this compound was observed in samples collected from plants infested with 300 *A. gossypii* for 96 h.

Four-arm olfactometer bioassays showed the attraction of female parasitic wasp *Aphidius colemani* to synthetic (*E*)-caryophyllene, while no impact on *A. gossypii* winged forms was detected. Thus, these results are consistent with the hypothesis of parasitoid attraction. To date, it is not yet clear the mechanism activated during aphid feeding which is responsible for the suppression of (*E*)-caryophyllene emission with low numbers of aphids, but it is evident that this strategy became ineffective with higher numbers of aphids.



**Figure 5.1.** Up- and down-regulated genes identified during the time course are reported in histograms connected by a continue black line to underline the trend in DEGs number identified. The model summarizes the main signalling and molecular events in zucchini plants at 24 (A), 48 (B) and 96 (C) hpi. Red and green arrows indicate up-regulation and down-regulation of genes involved in a specific pathway. (?) indicates lack of information on specific event.

In conclusion, *Aphis gossypii* extensively manipulate “San Pasquale” transcriptome. The analysis of differentially expressed genes showed that aphids elicit a defence response mainly regulated by SA-signalling pathway, and that the JA pathway is probably antagonised during zucchini response. Finally, overexpression of genes involved in photosynthesis and protein metabolism, as well as the modification of expression of a number of cell wall-related genes could be related to plant constant effort to balance cell homeostasis and activate an effective defence response, while aphids manipulate plant response to ensure food supply. Moreover, the strong presence of aphid-derived transcripts among DEGs could be related to a specific aphid strategy to increase the nutrient value of the phloem sap during feeding and to negatively influence complete activation of zucchini plant response. These hypotheses are supported by results obtained from the analyses of HI-VOCs emitted by zucchini plants infested with low numbers of aphids. The reduced (*E*)-caryophyllene emission is probable related to an aphid manipulation mechanism to avoid specific recognition from natural enemies, but this defence suppression could not be maintained with higher numbers of aphids. Further bioassays will be relevant to identify other biologically active compounds among the emitted VOCs.

Identification of genes involved in aphid defence represents a valuable resource for *Cucurbita pepo*. Future functional studies should be performed to elucidate the role of genes identified in defence response. The most interesting genes could be used as targets in experiments based on new plant breeding techniques (NPBT, e.g. genome-editing) to enhance endogenous resistance traits for an effective *Aphis gossypii* control.



## REFERENCES

- Aharoni, A., Giri, A. P., Deuerlein, S., Griepink, F., de Kogel, W. J., Verstappen, F. W., ... & Bouwmeester, H. J. (2003). Terpenoid metabolism in wild-type and transgenic Arabidopsis plants. *The Plant Cell*, 15(12), 2866-2884.
- Akacha, N. B., Boubaker, O., & Gargouri, M. (2005). Production of hexenol in a two-enzyme system: kinetic study and modelling. *Biotechnology letters*, 27(23), 1875-1878.
- Alagna, F., D'Agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M., Giuliano, G., Chiusano, M. L., Baldoni, L., & Perrotta, G. (2009). Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC genomics*, 10(1), 399.
- Alborn, H. T., Hansen, T. V., Jones, T. H., Bennett, D. C., Tumlinson, J. H., Schmelz, E. A., & Teal, P. E. (2007). Disulfoxy fatty acids from the American bird grasshopper *Schistocerca americana*, elicitors of plant volatiles. *Proceedings of the National Academy of Sciences*, 104(32), 12976-12981.
- Alborn, H. T., Turlings, T. C. J., Jones, T. H., Stenhagen, G., Loughrin, J. H., & Tumlinson, J. H. (1997). An elicitor of plant volatiles from beet armyworm oral secretion. *Science*, 276(5314), 945-949.
- 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.
- Anderson, J. A., & Peters, D. C. (1994). Ethylene production from wheat seedlings infested with biotypes of *Schizaphis graminum* (Homoptera: Aphididae). *Environmental entomology*, 23(4), 992-998.
- Ando, K., Carr, K. M., & Grumet, R. (2012). Transcriptome analyses of early cucumber fruit growth identifies distinct gene modules associated with phases of development. *BMC genomics*, 13(1), 518.
- Andrews, E. S., Theis, N., & Adler, L. S. (2007). Pollinator and herbivore attraction to *Cucurbita* floral volatiles. *Journal of chemical ecology*, 33(9), 1682-1691.
- Anstead, J., Samuel, P., Song, N., Wu, C., Thompson, G. A., & Goggin, F. (2010). Activation of ethylene-related genes in response to aphid feeding on resistant and susceptible melon and tomato plants. *Entomologia experimentalis et applicata*, 134(2), 170-181.
- Avci, U., Earl Petzold, H., Ismail, I. O., Beers, E. P., & Haigler, C. H. (2008). Cysteine proteases XCP1 and XCP2 aid micro-autolysis within the intact central vacuole during xylogenesis in Arabidopsis roots. *The Plant Journal*, 56(2), 303-315.
- Baldwin, I. T., Halitschke, R., Paschold, A., Von Dahl, C. C., & Preston, C. A. (2006). Volatile signaling in plant-plant interactions: "talking trees" in the genomics era. *Science*, 311(5762), 812-815.
- Bartram, S., Jux, A., Gleixner, G., & Boland, W. (2006). Dynamic pathway allocation in early terpenoid biosynthesis of stress-induced lima bean leaves. *Phytochemistry*, 67(15), 1661-1672.
- Bernasconi, M. L., Turlings, T. C., Ambrosetti, L., Bassetti, P., & Dorn, S. (1998). Herbivore-induced emissions of maize volatiles repel the corn leaf aphid, *Rhopalosiphum maidis*. *Entomologia Experimentalis et Applicata*, 87(2), 133-142.
- Bhattarai, K. K., Atamian, H. S., Kaloshian, I., & Eulgem, T. (2010). WRKY72-type transcription factors contribute to basal immunity in tomato and Arabidopsis as well as gene-for-gene resistance mediated by the tomato *R* gene *Mi-1*. *The Plant Journal*, 63(2), 229-240.
- Blanca, J., Cañizares, J., Cordero, L., Pascual, L., Diez, M. J., & Nuez, F. (2012). Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *PLoS one*, 7(10), e48198.
- Blanca, J., Cañizares, J., Roig, C., Ziarsolo, P., Nuez, F., & Picó, B. (2011). Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC genomics*, 12(1), 104.
- Boatright, J., Negre, F., Chen, X., Kish, C. M., Wood, B., Peel, G., ... & Dudareva, N. (2004). Understanding in vivo benzenoid metabolism in petunia petal tissue. *Plant Physiology*, 135(4), 1993-2011.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
- Bos, J. I., Prince, D., Pitino, M., Maffei, M. E., Win, J., & Hogenhout, S. A. (2010). A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (green peach aphid). *PLoS genetics*, 6(11), e1001216.

- Boter, M., Ruíz-Rivero, O., Abdeen, A., & Prat, S. (2004). Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. *Genes & development*, 18(13), 1577-1591.
- Botha, A. M., Swanevelder, Z. H., & Lapitan, N. L. (2010). Transcript profiling of wheat genes expressed during feeding by two different biotypes of *Diuraphis noxia*. *Environmental entomology*, 39(4), 1206-1231.
- Boyko, E. V., Smith, C. M., Thara, V. K., Bruno, J. M., Deng, Y., Starkey, S. R., & Klaahsen, D. L. (2006). Molecular basis of plant gene expression during aphid invasion: wheat *Pto*- and *Pti*-like sequences are involved in interactions between wheat and Russian wheat aphid (Homoptera: Aphididae). *Journal of economic entomology*, 99(4), 1430-1445.
- Brown, R. N., & Myers, J. R. (2002). A genetic map of squash (*Cucurbita* sp.) with randomly amplified polymorphic DNA markers and morphological markers. *Journal of the American Society for Horticultural Science*, 127(4), 568-575.
- Bruce, T. J., & Pickett, J. A. (2007). Plant defence signalling induced by biotic attacks. *Current opinion in plant biology*, 10(4), 387-392.
- Bruce, T. J., & Pickett, J. A. (2011). Perception of plant volatile blends by herbivorous insects—finding the right mix. *Phytochemistry*, 72(13), 1605-1611.
- Carolan, J. C., Caragea, D., Reardon, K. T., Mutti, N. S., Dittmer, N., Pappan, K., ... & Tagu, D. (2011). Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrtosiphon pisum*): a dual transcriptomic/proteomic approach. *Journal of proteome research*, 10(4), 1505-1518.
- Carolan, J. C., Fitzroy, C. I., Ashton, P. D., Douglas, A. E., & Wilkinson, T. L. (2009). The secreted salivary proteome of the pea aphid *Acyrtosiphon pisum* characterised by mass spectrometry. *Proteomics*, 9(9), 2457-2467.
- Casteel, C. L., Walling, L. L., & Paine, T. D. (2006). Behavior and biology of the tomato psyllid, *Bactericera cockerelli*, in response to the *Mi-1.2* gene. *Entomologia experimentalis et applicata*, 121(1), 67-72.
- Chen, F., Tholl, D., D'Auria, J. C., Farooq, A., Pichersky, E., & Gershenzon, J. (2003). Biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers. *The Plant Cell*, 15(2), 481-494.
- Cherepneva, G. N., Schmidt, K. H., Kulaeva, O. N., Oelmüller, R., & Kusnetsov, V. V. (2003). Expression of the ribosomal proteins S14, S16, L13a and L30 is regulated by cytokinin and abscisic acid: implication of the involvement of phytohormones in translational processes. *Plant science*, 165(5), 925-932.
- Cho, S. K., Jung, K. W., Jeung, J. U., Kang, K. H., Shim, K. S., You, M. K., ... & Shin, J. S. (2005). Analysis of differentially expressed transcripts from planthopper-infested wild rice (*Oryza minuta*). *Plant cell reports*, 24(1), 59-67.
- Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2009). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research*, 38(6), 1767-1771.
- Cohen, R., Hanan, A., & Paris, H. S. (2003). Single-gene resistance to powdery mildew in zucchini squash (*Cucurbita pepo*). *Euphytica*, 130(3), 433-441.
- Coppola, V., Coppola, M., Rocco, M., Digilio, M. C., D'Ambrosio, C., Renzone, G., ... & Corrado, G. (2013). Transcriptomic and proteomic analysis of a compatible tomato-aphid interaction reveals a predominant salicylic acid-dependent plant response. *Bmc Genomics*, 14(1), 515.
- Cristofolletti, P. T., Ribeiro, A. F., Deraison, C., Rahbé, Y., & Terra, W. R. (2003). Midgut adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid *Acyrtosiphon pisum*. *Journal of Insect Physiology*, 49(1), 11-24.
- D'Agostino, N., Golas, T., Van de Geest, H., Bombarely, A., Dawood, T., Zethof, J., Driedonks, N., Wijiker, E., Bargsten, J., Nap, J. P., Mariani, C., & Rieu, I. (2013). Genomic analysis of the native European Solanum species, *S. dulcamara*. *BMC genomics*, 14(1), 356.
- Dahmani-Mardas, F., Troadec, C., Boualem, A., Leveque, S., Alsadon, A. A., Aldoss, A. A., ... & Bendahmane, A. (2010). Engineering melon plants with improved fruit shelf life using the TILLING approach. *PloS one*, 5(12), e15776.
- Dalton, D. A., Boniface, C., Turner, Z., Lindahl, A., Kim, H. J., Jelinek, L., Govindarajulu, M., Finger, R. E., & Taylor, C. G. (2009). Physiological roles of glutathione S-transferases in soybean root nodules. *Plant physiology*, 150(1), 521-530.

- De Moraes, C. M., Lewis, W. J., Pare, P. W., Alborn, H. T., & Tumlinson, J. H. (1998). Herbivore-infested plants selectively attract parasitoids. *Nature*, 393(6685), 570.
- De Moraes, C. M., Mescher, M. C., & Tumlinson, J. H. (2001). Caterpillar-induced nocturnal plant volatiles repel conspecific females. *Nature*, 410(6828), 577.
- De Vos, M., Van Oosten, V. R., Van Poecke, R. M., Van Pelt, J. A., Pozo, M. J., Mueller, M. J., ... & Pieterse, C. M. (2005). Signal signature and transcriptome changes of *Arabidopsis* during pathogen and insect attack. *Molecular plant-microbe interactions*, 18(9), 923-937.
- Degen, T., Dillmann, C., Marion-Poll, F., & Turlings, T. C. (2004). High genetic variability of herbivore-induced volatile emission within a broad range of maize inbred lines. *Plant physiology*, 135(4), 1928-1938.
- Deleu, W., Esteras, C., Roig, C., González-To, M., Fernández-Silva, I., Gonzalez-Ibeas, D., ... & Monforte, A. J. (2009). A set of EST-SNPs for map saturation and cultivar identification in melon. *BMC Plant Biology*, 9(1), 90.
- Delp, G., Gradin, T., Åhman, I., & Jonsson, L. M. (2009). Microarray analysis of the interaction between the aphid *Rhopalosiphum padi* and host plants reveals both differences and similarities between susceptible and partially resistant barley lines. *Molecular Genetics and Genomics*, 281(3), 233-248.
- Dempsey, D. M. A., Vlot, A. C., Wildermuth, M. C., & Klessig, D. F. (2011). Salicylic acid biosynthesis and metabolism. *The Arabidopsis Book*, e0156.
- Denoeud, F., Aury, J. M., Da Silva, C., Noel, B., Rogier, O., Delledonne, M., ... & Jaillon, O. (2008). Annotating genomes with massive-scale RNA sequencing. *Genome biology*, 9(12), R175.
- Deraison, C., Darboux, I., Duportets, L., Gorjankina, T., Rahbé, Y., & Jouanin, L. (2004). Cloning and characterization of a gut-specific cathepsin L from the aphid *Aphis gossypii*. *Insect molecular biology*, 13(2), 165-177.
- Dewdney, J., Reuber, T. L., Wildermuth, M. C., Devoto, A., Cui, J., Stutius, L. M., ... & Ausubel, F. M. (2000). Three unique mutants of *Arabidopsis* identify eds loci required for limiting growth of a biotrophic fungal pathogen. *The Plant Journal*, 24(2), 205-218.
- Dewick, P. M. (1999). The biosynthesis of C 5–C 25 terpenoid compounds. *Natural product reports*, 16(1), 97-130.
- Diaz, A., Fergany, M., Formisano, G., Ziarsolo, P., Blanca, J., Fei, Z., ... & Oliver, M. (2011). A consensus linkage map for molecular markers and quantitative trait loci associated with economically important traits in melon (*Cucumis melo* L.). *BMC plant biology*, 11(1), 111.
- Dicke, M., & Baldwin, I. T. (2010). The evolutionary context for herbivore-induced plant volatiles: beyond the 'cry for help'. *Trends in plant science*, 15(3), 167-175.
- Dicke, M., & van Loon, J. J. (2000). Multitrophic effects of herbivore-induced plant volatiles in an evolutionary context. *Entomologia experimentalis et applicata*, 97(3), 237-249.
- Dicke, M., Van Beek, T. A., Posthumus, M. V., Dom, N. B., Van Bokhoven, H., & De Groot, A. E. (1990). Isolation and identification of volatile kairomone that affects acarine predator-prey interactions. Involvement of host plant in its production. *Journal of chemical ecology*, 16(2), 381-396.
- Digilio, M. C., Cascone, P., Iodice, L., & Guerrieri, E. (2012). Interactions between tomato volatile organic compounds and aphid behaviour. *Journal of plant interactions*, 7(4), 322-325.
- Dillwith, J. W., Berberet, R. C., Bergman, D. K., Neese, P. A., Edwards, R. M., & McNew, R. W. (1991). Plant biochemistry and aphid populations: studies on the spotted alfalfa aphid, *Therioaphis maculata*. *Archives of Insect Biochemistry and Physiology*, 17(4), 235-251.
- Divol, F., Vilaine, F., Thibivilliers, S., Amselem, J., Palauqui, J. C., Kusiak, C., & Dinant, S. (2005). Systemic response to aphid infestation by *Myzus persicae* in the phloem of *Apium graveolens*. *Plant molecular biology*, 57(4), 517.
- Dogimont, C., Chovelon, V., Pauquet, J., Boualem, A., & Bendahmane, A. (2014). The Vat locus encodes for a CC-NBS-LRR protein that confers resistance to *Aphis gossypii* infestation and *A. gossypii*-mediated virus resistance. *The Plant Journal*, 80(6), 993-1004.
- Dorschner, K. W. (1990). Aphid induced alteration of the availability and form of nitrogenous compounds in plants.
- Drukker, B., Bruin, J., & Sabelis, M. W. (2000). Anthocorid predators learn to associate herbivore-induced plant volatiles with presence or absence of prey. *Physiological Entomology*, 25(3), 260-265.

- Drzewiecka, K., Jeleń, H., Narożna, D., Rucińska-Sobkowiak, R., Kęsy, J., Floryszak-Wieczorek, J., ... & Morkunas, I. (2014). Differential induction of *Pisum sativum* defense signaling molecules in response to pea aphid infestation. *Plant Science*, 221, 1-12.
- Du, Y., Poppy, G. M., Powell, W., Pickett, J. A., Wadhams, L. J., & Woodcock, C. M. (1998). Identification of semiochemicals released during aphid feeding that attract parasitoid *Aphidius ervi*. *Journal of chemical Ecology*, 24(8), 1355-1368.
- Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., ... & Gershenzon, J. (2005). The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. *Proceedings of the National Academy of Sciences of the United States of America*, 102(3), 933-938.
- Dudareva, N., Klempien, A., Muhlemann, J. K., & Kaplan, I. (2013). Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytologist*, 198(1), 16-32.
- Dudareva, N., Negre, F., Nagegowda, D. A., & Orlova, I. (2006). Plant volatiles: recent advances and future perspectives. *Critical reviews in plant sciences*, 25(5), 417-440.
- Dudareva, N., Pichersky, E., & Gershenzon, J. (2004). Biochemistry of plant volatiles. *Plant physiology*, 135(4), 1893-1902.
- Ebert, T. A., & Cartwright, B. (1997). Biology and ecology of *Aphis gossypii* Glover (Homoptera: aphididae). *Southwestern Entomologist*, 22(1), 116-153.
- Ellis, C., Karafyllidis, I., Wasternack, C., & Turner, J. G. (2002). The Arabidopsis mutant *cev1* links cell wall signaling to jasmonate and ethylene responses. *The Plant Cell Online*, 14(7), 1557-1566.
- Elzinga, D. A., De Vos, M., & Jander, G. (2014). Suppression of plant defenses by a *Myzus persicae* (green peach aphid) salivary effector protein. *Molecular Plant-Microbe Interactions*, 27(7), 747-756.
- Esteras, C., Gómez, P., Monforte, A. J., Blanca, J., Vicente-Dólera, N., Roig, C., Nuez, F., & Picó, B. (2012). High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait loci mapping. *BMC genomics*, 13(1), 80.
- Ezura, H., & Owino, W. O. (2008). Melon, an alternative model plant for elucidating fruit ripening. *Plant Science*, 175(1), 121-129.
- FAO. (2014). FAOstat. Retrieved Feb, 2014.
- Fernandez-Silva, I., Eduardo, I., Blanca, J., Esteras, C., Pico, B., Nuez, F., ... & Monforte, A. J. (2008). Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.). *Theoretical and Applied Genetics*, 118(1), 139-150.
- Feussner, I., & Wasternack, C. (2002). The lipoxygenase pathway. *Annual review of plant biology*, 53(1), 275-297.
- Fonseca, J. P., & Dong, X. (2014). Functional characterization of a Nudix hydrolase *AtNUDX8* upon pathogen attack indicates a positive role in plant immune responses. *PloS one*, 9(12), e114119.
- Foyer, C. H., Karpinska, B., & Krupinska, K. (2014). The functions of WHIRLY1 and REDOX-RESPONSIVE TRANSCRIPTION FACTOR 1 in cross tolerance responses in plants: a hypothesis. *Phil. Trans. R. Soc. B*, 369(1640), 20130226.
- Francis, F., Guillonneau, F., Leprince, P., De Pauw, E., Haubruge, E., Jia, L., & Goggin, F. L. (2010). Tritrophic interactions among *Macrosiphum euphorbiae* aphids, their host plants and endosymbionts: investigation by a proteomic approach. *Journal of Insect Physiology*, 56(6), 575-585.
- Frenkel, O., Portillo, I., Brewer, M. T., Peros, J. P., Cadle-Davidson, L., & Milgroom, M. G. (2012). Development of microsatellite markers from the transcriptome of *Erysiphe necator* for analysing population structure in North America and Europe. *Plant pathology*, 61(1), 106-119.
- Furch, A. C., Hafke, J. B., Schulz, A., & van Bel, A. J. (2007). Ca<sup>2+</sup>-mediated remote control of reversible sieve tube occlusion in *Vicia faba*. *Journal of Experimental Botany*, 58(11), 2827-2838.
- Fürstenberg-Hägg, J., Zagrobelny, M., & Bak, S. (2013). Plant defense against insect herbivores. *International journal of molecular sciences*, 14(5), 10242-10297.
- Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V. M., ... & Alioto, T. (2012). The genome of melon (*Cucumis melo* L.). *Proceedings of the National Academy of Sciences*, 109(29), 11872-11877.
- Garcion, C., Lohmann, A., Lamodièrè, E., Catinot, J., Buchala, A., Doermann, P., & Métraux, J. P. (2008). Characterization and biological function of the ISOCHORISMATE SYNTHASE2 gene of Arabidopsis. *Plant physiology*, 147(3), 1279-1287.

- Gatehouse, J. (2011). Prospects for using proteinase inhibitors to protect transgenic plants against attack by herbivorous insects. *Current Protein and Peptide Science*, 12(5), 409-416.
- Gershenzon, J., McConkey, M. E., & Croteau, R. B. (2000). Regulation of monoterpene accumulation in leaves of peppermint. *Plant Physiology*, 122(1), 205-214.
- Gibson, R. W., & Pickett, J. A. (1983). Wild potato repels aphids by release of aphid alarm pheromone. *Nature*, 302(5909), 608-609.
- Grousse, C., Moulia, B., Silk, W., & Bonnemain, J. L. (2005). Aphid infestation causes different changes in carbon and nitrogen allocation in alfalfa stems as well as different inhibitions of longitudinal and radial expansion. *Plant Physiology*, 137(4), 1474-1484.
- Glinwood, R., Ahmed, E., Qvarfordt, E., Ninkovic, V., & Pettersson, J. (2009). Airborne interactions between undamaged plants of different cultivars affect insect herbivores and natural enemies. *Arthropod-Plant Interactions*, 3(4), 215-224.
- Goggin, F. L. (2007). Plant-aphid interactions: molecular and ecological perspectives. *Current opinion in plant biology*, 10(4), 399-408.
- Gong, L., Pachner, M., Kalai, K., & Lelley, T. (2008). SSR-based genetic linkage map of *Cucurbita moschata* and its synteny with *Cucurbita pepo*. *Genome*, 51(11), 878-887.
- Gonzalez-Ballester, D., Pootakham, W., Mus, F., Yang, W., Catalanotti, C., Magneschi, L., ... & Fernandez, E. (2011). Reverse genetics in Chlamydomonas: a platform for isolating insertional mutants. *Plant Methods*, 7(1), 24.
- Gordon, A., & Hannon, G. J. (2010). Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished) [http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit), 5.
- Gordon, K. H., & Waterhouse, P. M. (2007). RNAi for insect-proof plants. *Nature biotechnology*, 25(11), 1231-1232.
- Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Montserrat, R., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, 36(10), 3420-3435.
- Gouinguéné, S., Degen, T., & Turlings, T. C. (2001). Variability in herbivore-induced odour emissions among maize cultivars and their wild ancestors (teosinte). *Chemoecology*, 11(1), 9-16.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Chen, Z. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29(7), 644-652.
- Green, T. R., & Ryan, C. A. (1972). Wound-induced proteinase inhibitor in plant leaves: a possible defense mechanism against insects. *Science*, 175(4023), 776-777.
- Guerrieri, E., & Digilio, M. C. (2008). Aphid-plant interactions: a review. *Journal of Plant Interactions*, 3(4), 223-232.
- Guerrieri, E., Poppy, G. M., Powell, W., Tremblay, E., & Pennacchio, F. (1999). Induction and systemic release of herbivore-induced plant volatiles mediating in-flight orientation of *Aphidius ervi*. *Journal of Chemical Ecology*, 25(6), 1247-1261.
- Guo, S., Liu, J., Zheng, Y., Huang, M., Zhang, H., Gong, G., ... & Xu, Y. (2011). Characterization of transcriptome dynamics during watermelon fruit development: sequencing, assembly, annotation and gene expression profiles. *BMC genomics*, 12(1), 454.
- Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., ... & Min, J. (2013). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature genetics*, 45(1), 51-58.
- Guo, S., Zheng, Y., Joung, J. G., Liu, S., Zhang, Z., Crasta, O. R., ... & Fei, Z. (2010). Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC genomics*, 11(1), 384.
- Halitschke, R., Schittko, U., Pohnert, G., Boland, W., & Baldwin, I. T. (2001). Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. III. Fatty acid-amino acid conjugates in herbivore oral secretions are necessary and sufficient for herbivore-specific plant responses. *Plant physiology*, 125(2), 711-717.
- Han, B. Y., & Chen, Z. M. (2002). Composition of the volatiles from intact and tea aphid-damaged tea shoots and their allurements to several natural enemies of the tea aphid. *Journal of Applied Entomology*, 126(9), 497-500.

- Hardie, J., Isaacs, R., Pickett, J. A., Wadhams, L. J., & Woodcock, C. M. (1994). Methyl salicylate and (-)-(1R, 5S)-myrtenal are plant-derived repellents for black bean aphid, *Aphis fabae* Scop. (Homoptera: Aphididae). *Journal of Chemical Ecology*, 20(11), 2847-2855.
- Harmel, N., Almohamad, R., FAUCONNIER, M. L., Du Jardin, P., Verheggen, F., Marlier, M., ... & Francis, F. (2007). Role of terpenes from aphid-infested potato on searching and oviposition behavior of *Episyrphus balteatus*. *Insect Science*, 14(1), 57-63.
- Harmel, N., Létocart, E., Cherqui, A., Giordanengo, P., Mazzucchelli, G., Guillonneau, F., ... & Francis, F. (2008). Identification of aphid salivary proteins: a proteomic investigation of *Myzus persicae*. *Insect molecular biology*, 17(2), 165-174.
- Hatanaka, A. (1993). The biogenesis of green odour by green leaves. *Phytochemistry*, 34(5), 1201-1218.
- Heil, M. (2008). Indirect defence via tritrophic interactions. *New Phytologist*, 178(1), 41-61.
- Heil, M., & Silva Bueno, J. C. S. (2007). Herbivore-induced volatiles as rapid signals in systemic plant responses: how to quickly move the information?. *Plant signaling & behavior*, 2(3), 191-193.
- Hilker, M., & Meiners, T. (2002). Induction of plant responses to oviposition and feeding by herbivorous arthropods: a comparison. *Entomologia Experimentalis et Applicata*, 104(1), 181-192.
- Hogenhout, S. A., & Bos, J. I. (2011). Effector proteins that modulate plant–insect interactions. *Current opinion in plant biology*, 14(4), 422-428.
- Houseman, J. G., MacNaughton, W. K., & Downe, A. E. R. (1984). Cathepsin B and aminopeptidase activity in the posterior midgut of *Euschistus euschistoides* (Hemiptera: Pentatomidae). *The Canadian Entomologist*, 116(10), 1393-1396.
- Houseman, J. G., Morrison, P. E., & Downe, A. E. R. (1985). Cathepsin B and aminopeptidase in the posterior midgut of *Phymata wolffii* Stål (Hemiptera: Phymatidae). *Canadian Journal of Zoology*, 63(6), 1288-1291.
- Howe, G. A., & Schaller, A. (2008). Direct defenses in plants and their induction by wounding and insect herbivores. In *Induced plant resistance to herbivory* (pp. 7-29). Springer Netherlands.
- Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., ... & Ren, Y. (2009). The genome of the cucumber, *Cucumis sativus* L. *Nature genetics*, 41(12), 1275-1281
- Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome research*, 9(9), 868-877.
- Hulm, J. L., McIntosh, K. B., & Bonham-Smith, P. C. (2005). Variation in transcript abundance among the four members of the *Arabidopsis thaliana* RIBOSOMAL PROTEIN S15a gene family. *Plant science*, 169(1), 267-278.
- Iseli, C., Jongeneel, C. V., & Bucher, P. (1999, August). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. In *ISMB* (Vol. 99, pp. 138-148).
- Jamal, F., Pandey, P. K., Singh, D., & Khan, M. Y. (2013). Serine protease inhibitors in plants: nature's arsenal crafted for insect predators. *Phytochemistry reviews*, 12(1), 1-34.
- James, D. G. (2003). Field evaluation of herbivore-induced plant volatiles as attractants for beneficial insects: methyl salicylate and the green lacewing, *Chrysopa nigricornis*. *Journal of chemical ecology*, 29(7), 1601-1609.
- Jaouannet, M., Rodriguez, P. A., Thorpe, P., Lenoir, C. J., MacLeod, R., Escudero-Martinez, C., & Bos, J. I. (2014). Plant immunity in plant–aphid interactions. *Frontiers in plant science*, 5.
- Ji, R., Yu, H., Fu, Q., Chen, H., Ye, W., Li, S., & Lou, Y. (2013). Comparative transcriptome analysis of salivary glands of two populations of rice brown planthopper, *Nilaparvata lugens*, that differ in virulence. *PLoS one*, 8(11), e79612.
- Jones, J. D., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323.
- Kaloshian, I. (2004). Gene-for-gene disease resistance: bridging insect pest and pathogen defense. *Journal of chemical ecology*, 30(12), 2419-2438.
- Kaloshian, I., Kinsey, M. G., Ullman, D. E., & Williamson, V. M. (1997). The impact of *Meu1*-mediated resistance in tomato on longevity, fecundity and behavior of the potato aphid, *Macrosiphum euphorbiae*. *Entomologia experimentalis et applicata*, 83(2), 181-187.
- Kappers, I. F., Hoogerbrugge, H., Bouwmeester, H. J., & Dicke, M. (2011). Variation in herbivory-induced volatiles among cucumber (*Cucumis sativus* L.) varieties has consequences for the attraction of carnivorous natural enemies. *Journal of chemical ecology*, 37(2), 150-160.

- Kessler, A., & Baldwin, I. T. (2001). Defensive function of herbivore-induced plant volatile emissions in nature. *Science*, 291(5511), 2141-2144.
- Klingler, J., Creasy, R., Gao, L., Nair, R. M., Calix, A. S., Jacob, H. S., ... & Singh, K. B. (2005). Aphid resistance in *Medicago truncatula* involves antixenosis and phloem-specific, inducible antibiosis, and maps to a single locus flanked by NBS-LRR resistance gene analogs. *Plant physiology*, 137(4), 1445-1455.
- Knudsen, J. T., Eriksson, R., Gershenzon, J., & Ståhl, B. (2006). Diversity and distribution of floral scent. *The botanical review*, 72(1), 1-120.
- Köllner, T. G., Held, M., Lenk, C., Hiltbold, I., Turlings, T. C., Gershenzon, J., & Degenhardt, J. (2008). A maize (*E*)- $\beta$ -caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *The Plant Cell*, 20(2), 482-494.
- Korth, K. L., & Thompson, G. A. (2006). Chemical signals in plants: jasmonates and the role of insect-derived elicitors in responses to herbivores. In *Multigenic and induced systemic resistance in plants*(pp. 259-278). Springer US.
- Kuśnierczyk, A., Tran, D. H., Winge, P., Jørstad, T. S., Reese, J. C., Troczyńska, J., & Bones, A. M. (2011). Testing the importance of jasmonate signalling in induction of plant defences upon cabbage aphid (*Brevicoryne brassicae*) attack. *BMC genomics*, 12(1), 423.
- Kuśnierczyk, A., Winge, P. E. R., Jørstad, T. S., Troczyńska, J., Rossiter, J. T., & Bones, A. M. (2008). Towards global understanding of plant defence against aphids—timing and dynamics of early *Arabidopsis* defence responses to cabbage aphid (*Brevicoryne brassicae*) attack. *Plant, Cell & Environment*, 31(8), 1097-1115.
- Kuśnierczyk, A., Winge, P., Midelfart, H., Armbruster, W. S., Rossiter, J. T., & Bones, A. M. (2007). Transcriptional responses of *Arabidopsis thaliana* ecotypes with different glucosinolate profiles after attack by polyphagous *Myzus persicae* and oligophagous *Brevicoryne brassicae*. *Journal of Experimental Botany*, 58(10), 2537-2552.
- Kutsukake, M., Shibao, H., Nikoh, N., Morioka, M., Tamura, T., Hoshino, T., ... & Fukatsu, T. (2004). Venomous protease of aphid soldier for colony defense. *Proceedings of the National Academy of Sciences of the United States of America*, 101(31), 11338-11343.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357-359.
- Le Mire, G., Nguyen, M. L., Fassotte, B., du Jardin, P., Verheggen, F., Delaplace, P., & Jijakli, M. H. (2016). Implementing plant biostimulants and biocontrol strategies in the agroecological management of cultivated ecosystems. *Biotechnologie, Agronomie, Société et Environnement*, 20(S1), 299.
- Lecourieux, D., Ranjeva, R., & Pugin, A. (2006). Calcium in plant defence-signalling pathways. *New Phytologist*, 171(2), 249-269.
- Lee, Y. H., Jeon, H. J., Kim, B. D., & Hong, K. H. (1995). Use of random amplified polymorphic DNAs for linkage group analysis in interspecific hybrid F2 generation of *Cucurbita*. *Journal of The Korean Society for Horticultural Science (Korea Republic)*.
- Li, X., Schuler, M. A., & Berenbaum, M. R. (2002). Jasmonate and salicylate induce expression of herbivore cytochrome P450 genes. *Nature*, 419(6908), 712-715.
- Li, Z., Huang, S., Liu, S., Pan, J., Zhang, Z., Tao, Q., ... & Si, L. (2009). Molecular isolation of the M gene suggests that a conserved-residue conversion induces the formation of bisexual flowers in cucumber plants. *Genetics*, 182(4), 1381-1385.
- Liang, D., Liu, M., Hu, Q., He, M., Qi, X., Xu, Q., ... & Chen, X. (2015). Identification of differentially expressed genes related to aphid resistance in cucumber (*Cucumis sativus* L.). *Scientific Reports*, 5.
- Lichtenthaler, H. K., Rohmer, M., & Schwender, J. (1997). Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. *Physiologia plantarum*, 101(3), 643-652.
- Liu, X., Kim, C. N., Yang, J., Jemmerson, R., & Wang, X. (1996). Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome c. *Cell*, 86(1), 147-157.
- Lo Pinto, M., Wajnberg, E., Colazza, S., Curty, C., & Fauvergue, X. (2004). Olfactory response of two aphid parasitoids, *Lysiphlebus testaceipes* and *Aphidius colemani*, to aphid-infested plants from a distance. *Entomologia Experimentalis et Applicata*, 110(2), 159-164.
- Lu, H., Rate, D. N., Song, J. T., & Greenberg, J. T. (2003). ACD6, a novel ankyrin protein, is a regulator and an effector of salicylic acid signaling in the *Arabidopsis* defense response. *The Plant Cell*, 15(10), 2408-2420.

- Macaulay, K. M., Heath, G. A., Ciulli, A., Murphy, A. M., Abell, C., Carr, J. P., & Smith, A. G. (2017). The biochemical properties of the two *Arabidopsis thaliana* isochorismate synthases. *Biochemical Journal*, 474(10), 1579-1590.
- Maffei, M. E., Mithöfer, A., & Boland, W. (2007). Insects feeding on plants: rapid signals and responses preceding the induction of phytochemical release. *Phytochemistry*, 68(22), 2946-2959.
- Mascarell-Creus, A., Cañizares, J., Vilarrasa-Blasi, J., Mora-García, S., Blanca, J., Gonzalez-Ibeas, D., ... & López-Bigas, N. (2009). An oligo-based microarray offers novel transcriptomic approaches for the analysis of pathogen resistance and fruit quality traits in melon (*Cucumis melo* L.). *BMC genomics*, 10(1), 467.
- Mauck, K. E., De Moraes, C. M., & Mescher, M. C. (2014). Biochemical and physiological mechanisms underlying effects of Cucumber mosaic virus on host-plant traits that mediate transmission by aphid vectors. *Plant, cell & environment*, 37(6), 1427-1439.
- McConkey, M. E., Gershenzon, J., & Croteau, R. B. (2000). Developmental regulation of monoterpene biosynthesis in the glandular trichomes of peppermint. *Plant Physiology*, 122(1), 215-224.
- McDonald, R. P. (2014). *Factor analysis and related methods*. Psychology Press.
- McGregor, S. E. (1976). *Insect pollination of cultivated crop plants*(Vol. 496). Washington (DC): Agricultural Research Service, US Department of Agriculture.
- Mello, M. O., & Silva-Filho, M. C. (2002). Plant-insect interactions: an evolutionary arms race between two distinct defense mechanisms. *Brazilian Journal of Plant Physiology*, 14(2), 71-81.
- Meyer, A., Miersch, O., Büttner, C., Dathe, W., & Sembdner, G. (1984). Occurrence of the plant growth regulator jasmonic acid in plants. *Journal of Plant Growth Regulation*, 3(1), 1-8.
- Miles, P. W. (1987). Feeding process of *Aphidoidea* in relation to effects on their food plants. *Aphids: their biology, natural enemies, and control*. Edited by AK Minks and P. Harrewijn.
- Milligan, S. B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P., & Williamson, V. M. (1998). The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *The Plant Cell*, 10(8), 1307-1319.
- Mithöfer, A., & Boland, W. (2008). Recognition of herbivory-associated molecular patterns. *Plant Physiology*, 146(3), 825-831.
- Mithöfer, A., & Boland, W. (2012). Plant defense against herbivores: chemical aspects. *Annual review of plant biology*, 63, 431-450.
- Mohase, L., & van der Westhuizen, A. J. (2002). Salicylic acid is involved in resistance responses in the Russian wheat aphid-wheat interaction. *Journal of Plant Physiology*, 159(6), 585-590.
- Moin, M., Bakshi, A., Saha, A., Dutta, M., Madhav, S. M., & Kirti, P. B. (2016). Rice ribosomal protein large subunit genes and their spatio-temporal and stress regulation. *Frontiers in plant science*, 7.
- Molina, C., Rotter, B., Horres, R., Udupa, S. M., Besser, B., Bellarmino, L., ... & Winter, P. (2008). SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. *BMC genomics*, 9(1), 553.
- Montero-Pau, J., Blanca, J., Esteras, C., Martínez-Pérez, E. M., Gómez, P., Monforte, A. J., ... & Picó, B. (2017). An SNP-based saturated genetic map and QTL analysis of fruit-related traits in Zucchini using Genotyping-by-sequencing. *BMC genomics*, 18(1), 94.
- Moran, P. J., & Thompson, G. A. (2001). Molecular responses to aphid feeding in *Arabidopsis* in relation to plant defense pathways. *Plant Physiology*, 125(2), 1074-1085.
- Moran, P. J., Cheng, Y., Cassell, J. L., & Thompson, G. A. (2002). Gene expression profiling of *Arabidopsis thaliana* in compatible plant-aphid interactions. *Archives of insect biochemistry and physiology*, 51(4), 182-203.
- Morkunas, I., & Gabryś, B. (2011). Phytohormonal signaling in plant responses to aphid feeding. *Acta Physiologiae Plantarum*, 33(6), 2057-2073.
- Morozova, O., Hirst, M., & Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual review of genomics and human genetics*, 10, 135-151.
- Mur, L. A., Kenton, P., Atzorn, R., Miersch, O., & Wasternack, C. (2006). The outcomes of concentration-specific interactions between salicylate and jasmonate signaling include synergy, antagonism, and oxidative stress leading to cell death. *Plant physiology*, 140(1), 249-262.



- Musser, R. O., Cipollini, D. F., Hum-Musser, S. M., Williams, S. A., Brown, J. K., & Felton, G. W. (2005). Evidence that the caterpillar salivary enzyme glucose oxidase provides herbivore offense in solanaceous plants. *Archives of insect biochemistry and physiology*, 58(2), 128-137
- Mutti, N. S., Park, Y., Reese, J. C., & Reeck, G. R. (2006). RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrtosiphon pisum*. *Journal of insect science*, 6(38), 1-7.
- Nazzi, F. (1995). OLFA 1.0: A computer program for collecting and analysing behavioural data with the four-way olfactometer. Exeter Software. ISBN 0-925031-26-7.
- Nicholson, S. J., Hartson, S. D., & Puterka, G. J. (2012). Proteomic analysis of secreted saliva from Russian Wheat Aphid (*Diuraphis noxia* Kurd.) biotypes that differ in virulence to wheat. *Journal of Proteomics*, 75(7), 2252-2268.
- Niinemets, Ü., Kännaste, A., & Copolovici, L. (2013). Quantitative patterns between plant volatile emissions induced by biotic stresses and the degree of damage. *Frontiers in Plant Science*, 4.
- Novaes, E., Drost, D. R., Farmerie, W. G., Pappas, G. J., Grattapaglia, D., Sederoff, R. R., & Kirst, M. (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC genomics*, 9(1), 312.
- Ogawa, S., Lozach, J., Benner, C., Pascual, G., Tangirala, R. K., Westin, S., ... & Glass, C. K. (2005). Molecular determinants of crosstalk between nuclear receptors and toll-like receptors. *Cell*, 122(5), 707-721.
- Oshlack, A., Robinson, M. D., & Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome biology*, 11(12), 220.
- Palanisamy, V., & Wong, D. T. (2010). Transcriptomic analyses of saliva. *Oral Biology: Molecular Techniques and Applications*, 43-51.
- Paré, P. W., & Tumlinson, J. H. (1996). Plant volatile signals in response to herbivore feeding. *Florida Entomologist*, 93-103.
- Paré, P. W., & Tumlinson, J. H. (1997). *De novo* biosynthesis of volatiles induced by insect herbivory in cotton plants. *Plant physiology*, 114(4), 1161-1167.
- Pareja, M., Qvarfordt, E., Webster, B., Mayon, P., Pickett, J., Birkett, M., & Glinwood, R. (2012). Herbivory by a phloem-feeding insect inhibits floral volatile production. *PLoS One*, 7(2), e31971.
- Paris H.S. (2008) Summer Squash. In: Prohens J., Nuez F. (eds) Vegetables I. Handbook of Plant Breeding, vol 1. Springer, New York, NY
- Park, S. J., Huang, Y., & Ayoubi, P. (2006). Identification of expression profiles of sorghum genes in response to greenbug phloem-feeding using cDNA subtraction and microarray analysis. *Planta*, 223(5), 932-947.
- Pegadaraju, V., Knepper, C., Reese, J., & Shah, J. (2005). Premature leaf senescence modulated by the Arabidopsis PHYTOALEXIN DEFICIENT4 gene is associated with defense against the phloem-feeding green peach aphid. *Plant Physiology*, 139(4), 1927-1934.
- Pegadaraju, V., Louis, J., Singh, V., Reese, J. C., Bautor, J., Feys, B. J., ... & Shah, J. (2007). Phloem-based resistance to green peach aphid is controlled by Arabidopsis PHYTOALEXIN DEFICIENT4 without its signaling partner ENHANCED DISEASE SUSCEPTIBILITY1. *The Plant Journal*, 52(2), 332-341.
- Pichersky, E., Noel, J. P., & Dudareva, N. (2006). Biosynthesis of plant volatiles: nature's diversity and ingenuity. *Science*, 311(5762), 808-811.
- Pohnert, G., Jung, V., Haukioja, E., Lempa, K., & Boland, W. (1999). New fatty acid amides from regurgitant of lepidopteran (Noctuidae, Geometridae) caterpillars. *Tetrahedron*, 55(37), 11275-11280.
- Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. *Nature*, 475(7355), 189-195.
- Powell, G., Tosh, C. R., & Hardie, J. (2006). Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annu. Rev. Entomol.*, 51, 309-330.
- Rahbé, Y., Sauvion, N., Febvay, G., Peumans, W. J., & Gatehouse, A. M. (1995). Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrtosiphon pisum*. *Entomologia Experimentalis et Applicata*, 76(2), 143-155.
- Rao, S. A., Carolan, J. C., & Wilkinson, T. L. (2013). Proteomic profiling of cereal aphid saliva reveals both ubiquitous and adaptive secreted proteins. *PLoS One*, 8(2), e57413.
- Rashid, A. (2016). Defense responses of plant cell wall non-catalytic proteins against pathogens. *Physiological and Molecular Plant Pathology*, 94, 38-46.

- Ren, Y., Zhang, Z., Liu, J., Staub, J. E., Han, Y., Cheng, Z., ... & Xie, B. (2009). An integrated genetic and cytogenetic map of the cucumber genome. *PLoS one*, 4(6), e5795.
- Ren, Y., Zhao, H., Kou, Q., Jiang, J., Guo, S., Zhang, H., ... & Levi, A. (2012). A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. *PLoS One*, 7(1), e29453.
- Roberts, A., & Pachter, L. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature methods*, 10(1), 71-73.
- Roberts, A., Pimentel, H., Trapnell, C., & Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*, 27(17), 2325-2329.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., ... & Griffith, M. (2010). *De novo* assembly and analysis of RNA-seq data. *Nature methods*, 7(11), 909-912.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140.
- Rodriguez, P. A., & Bos, J. I. (2013). Toward understanding the role of aphid effectors in plant infestation. *Molecular Plant-Microbe Interactions*, 26(1), 25-30.
- Rodriguez, P. A., Stam, R., Warbroek, T., & Bos, J. I. (2014). Mp10 and Mp42 from the aphid species *Myzus persicae* trigger plant defenses in *Nicotiana benthamiana* through different activities. *Molecular Plant-Microbe Interactions*, 27(1), 30-39.
- Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural product reports*, 16(5), 565-574.
- Rossi, M., Goggin, F. L., Milligan, S. B., Kaloshian, I., Ullman, D. E., & Williamson, V. M. (1998). The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. *Proceedings of the National Academy of Sciences*, 95(17), 9750-9754.
- Saibo, N. J., Lourenço, T., & Oliveira, M. M. (2008). Transcription factors and regulation of photosynthetic and related metabolism under environmental stresses. *Annals of botany*, 103(4), 609-623.
- Sasso, R., Iodice, L., Cristina Digilio, M., Carretta, A., Ariati, L., & Guerrieri, E. (2007). Host-locating response by the aphid parasitoid *Aphidius ervi* to tomato plant volatiles. *Journal of Plant Interactions*, 2(3), 175-183.
- Schmelz, E. A., Carroll, M. J., LeClere, S., Phipps, S. M., Meredith, J., Chourey, P. S., ... & Teal, P. E. (2006). Fragments of ATP synthase mediate plant perception of insect attack. *Proceedings of the National Academy of Sciences*, 103(23), 8894-8899.
- Schmelz, E. A., LeClere, S., Carroll, M. J., Alborn, H. T., & Teal, P. E. (2007). Cowpea chloroplastic ATP synthase is the source of multiple plant defense elicitors during insect herbivory. *Plant Physiology*, 144(2), 793-805.
- Schulz, M. H., Zerbino, D. R., Vingron, M., & Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 28(8), 1086-1092.
- Schwartzberg, E. G., Böröczky, K., & Tumlinson, J. H. (2011). Pea aphids, *Acyrtosiphon pisum*, suppress induced plant volatiles in broad bean, *Vicia faba*. *Journal of chemical ecology*, 37(10), 1055.
- Seo, H. S., Song, J. T., Cheong, J. J., Lee, Y. H., Lee, Y. W., Hwang, I., ... & Do Choi, Y. (2001). Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proceedings of the National Academy of Sciences*, 98(8), 4788-4793.
- Shapiro, L., Moraes, C. M., Stephenson, A. G., & Mescher, M. C. (2012). Pathogen effects on vegetative and floral odours mediate vector attraction and host exposure in a complex pathosystem. *Ecology Letters*, 15(12), 1430-1438.
- Shi, C. Y., Yang, H., Wei, C. L., Yu, O., Zhang, Z. Z., Jiang, C. J., ... & Wan, X. C. (2011). Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC genomics*, 12(1), 131.
- Shiba, H., Uchida, D., Kobayashi, H., & Natori, M. (2001). Involvement of cathepsin B-and L-like proteinases in silk gland histolysis during metamorphosis of *Bombyx mori*. *Archives of biochemistry and biophysics*, 390(1), 28-34.
- Shiu, S. H., & Bleecker, A. B. (2001). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proceedings of the National Academy of Sciences*, 98(19), 10763-10768.
- Shoji, T., Nakajima, K., & Hashimoto, T. (2000). Ethylene suppresses jasmonate-induced gene expression in nicotine biosynthesis. *Plant and Cell Physiology*, 41(9), 1072-1076.

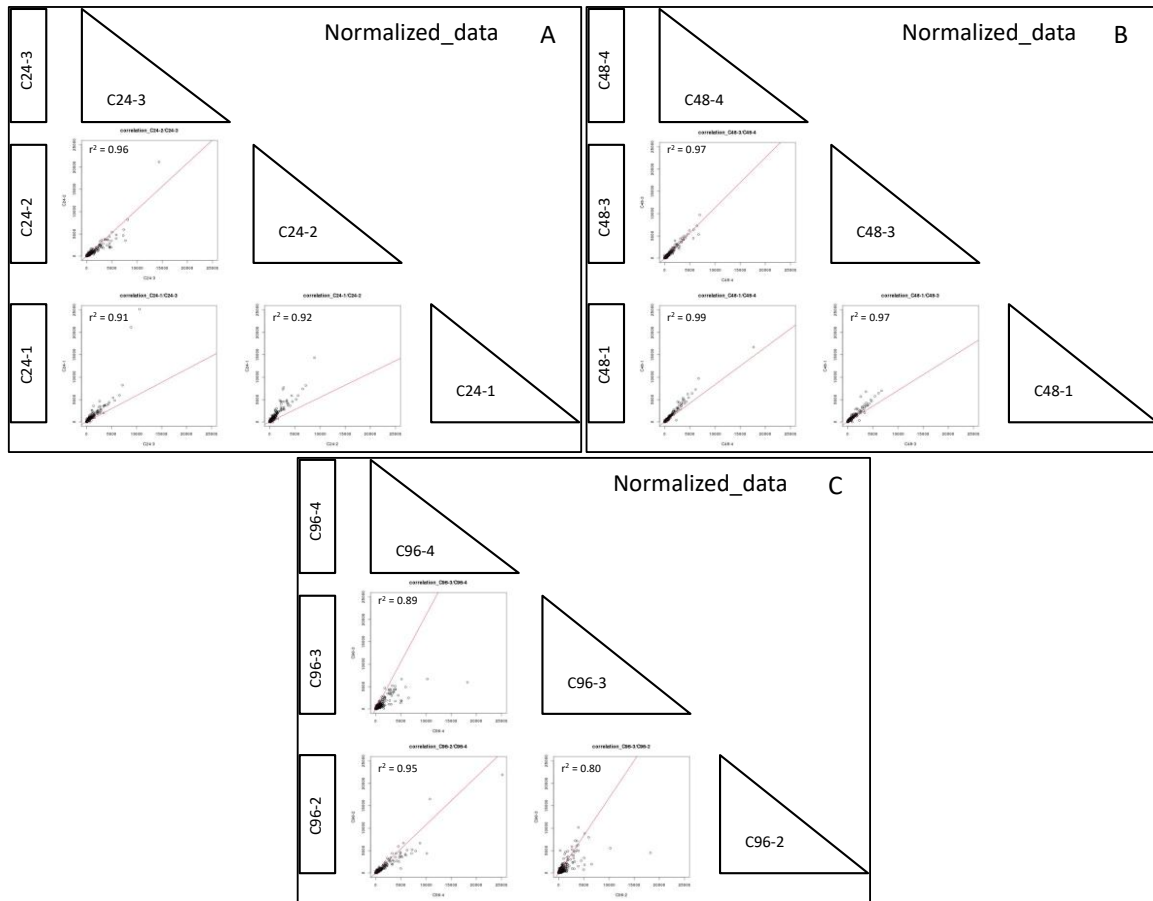
- Showalter, A. M. (1993). Structure and function of plant cell wall proteins. *The Plant Cell Online*, 5(1), 9-23.
- Singh, G., Singh, N. P., & Singh, R. (2014). Food plants of a major agricultural pest *Aphis gossypii* glover (Homoptera: Aphididae) from India: An updated checklist. *International Journal of Life Sciences Biotechnology and Pharma Research*, 3(2), 1.
- Smith, C. M., & Boyko, E. V. (2007). The molecular bases of plant resistance and defense responses to aphid feeding: current status. *Entomologia experimentalis et applicata*, 122(1), 1-16.
- Solano, R., Stepanova, A., Chao, Q., & Ecker, J. R. (1998). Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes & development*, 12(23), 3703-3714.
- Song, M. S., Kim, D. G., & Lee, S. H. (2005). Isolation and characterization of a jasmonic acid carboxyl methyltransferase gene from hot pepper (*Capsicum annuum* L.). *Journal of Plant Biology*, 48(3), 292-297.
- Spiteller, D., & Boland, W. (2003). N-(17-Acyloxy-acyl)-glutamines: novel surfactants from oral secretions of lepidopteran larvae. *The Journal of organic chemistry*, 68(23), 8743-8749.
- Strickler, S. R., Bombarely, A., & Mueller, L. A. (2012). Designing a transcriptome next-generation sequencing project for a nonmodel plant species 1. *American Journal of Botany*, 99(2), 257-266.
- Su, Y. L., Li, J. M., Li, M., Luan, J. B., Ye, X. D., Wang, X. W., & Liu, S. S. (2012). Transcriptomic analysis of the salivary glands of an invasive whitefly. *PLoS One*, 7(6), e39303.
- Sudheesh, S., Verma, P., Forster, J. W., Cogan, N. O., & Kaur, S. (2016). Generation and Characterisation of a Reference Transcriptome for Lentil (*Lens culinaris* Medik.). *International journal of molecular sciences*, 17(11), 1887.
- Tagu, D., Klingler, J. P., Moya, A., & Simon, J. C. (2008). Early progress in aphid genomics and consequences for plant-aphid interactions studies. *Molecular plant-microbe interactions*, 21(6), 701-708.
- Tamiru, A., Bruce, T. J., Woodcock, C. M., Caulfield, J. C., Midega, C. A., Ogot, C. K., ... & Khan, Z. R. (2011). Maize landraces recruit egg and larval parasitoids in response to egg deposition by a herbivore. *Ecology Letters*, 14(11), 1075-1083.
- Terra, W. R., Espinoza-Fuentes, F. P., Ribeiro, A. F., & Ferreira, C. (1988). The larval midgut of the housefly (*Musca domestica*): ultrastructure, fluid fluxes and ion secretion in relation to the organization of digestion. *Journal of Insect Physiology*, 34(6), 463-472.
- Theis, N., Kesler, K., & Adler, L. S. (2009). Leaf herbivory increases floral fragrance in male but not female *Cucurbita pepo* subsp. *texana* (Cucurbitaceae) flowers. *American journal of botany*, 96(5), 897-903.
- Tholl, D. (2006). Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Current opinion in plant biology*, 9(3), 297-304.
- Thompson, G. A., & Goggin, F. L. (2006). Transcriptomics and functional genomics of plant defence induction by phloem-feeding insects. *Journal of Experimental Botany*, 57(4), 755-766.
- Tsuda, K., Sato, M., Glazebrook, J., Cohen, J. D., & Katagiri, F. (2008). Interplay between MAMP-triggered and SA-mediated defense responses. *The Plant Journal*, 53(5), 763-775.
- Turlings, T. C., Lengwiler, U. B., Bernasconi, M. L., & Wechsler, D. (1998). Timing of induced volatile emissions in maize seedlings. *Planta*, 207(1), 146-152.
- Turlings, T. C., Tumlinson, J. H., & Lewis, W. J. (1990). Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science(Washington)*, 250(4985), 1251-1253.
- Unamba, C. I., Nag, A., & Sharma, R. K. (2015). Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. *Frontiers in plant science*, 6.
- van de Ven, W. T., LeVesque, C. S., Perring, T. M., & Walling, L. L. (2000). Local and systemic changes in squash gene expression in response to silverleaf whitefly feeding. *The Plant Cell*, 12(8), 1409-1423.
- Van Emden, H. F., & Harrington, R. (2007). *Aphids as crop pests*. CABI. Wallingford, United Kingdom.
- Van Kleeff, P. J., Galland, M., Schuurink, R. C., & Bleeker, P. M. (2016). Small RNAs from *Bemisia tabaci* are transferred to *Solanum lycopersicum* phloem during feeding. *Frontiers in plant science*, 7.
- Vancanneyt, G., Sanz, C., Farmaki, T., Paneque, M., Ortego, F., Castañera, P., & Sánchez-Serrano, J. J. (2001). Hydroperoxide lyase depletion in transgenic potato plants leads to an

- increase in aphid performance. *Proceedings of the national academy of sciences*, 98(14), 8139-8144.
- Vandermoten, S., Mescher, M. C., Francis, F., Haubruge, E., & Verheggen, F. J. (2012). Aphid alarm pheromone: an overview of current knowledge on biosynthesis and functions. *Insect biochemistry and molecular biology*, 42(3), 155-163.
  - Vassão, D. G., Gang, D. R., Koeduka, T., Jackson, B., Pichersky, E., Davin, L. B., & Lewis, N. G. (2006). Chavicol formation in sweet basil (*Ocimum basilicum*): cleavage of an esterified C9 hydroxyl group with NAD (P) H-dependent reduction. *Organic & biomolecular chemistry*, 4(14), 2733-2744.
  - Vera, J. C., Wheat, C. W., Fescemyer, H. W., Frilander, M. J., Crawford, D. L., Hanski, I., & Marden, J. H. (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular ecology*, 17(7), 1636-1647.
  - Verma, V., Ravindran, P., & Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC plant biology*, 16(1), 86.
  - Vet, L. E., & Dicke, M. (1992). Ecology of infochemical use by natural enemies in a tritrophic context. *Annual review of entomology*, 37(1), 141-172.
  - Voelckel, C., Weisser, W. W., & Baldwin, I. T. (2004). An analysis of plant-aphid interactions by different microarray hybridization strategies. *Molecular Ecology*, 13(10), 3187-3195.
  - Walling, L. L. (2000). The myriad plant responses to herbivores. *Journal of Plant Growth Regulation*, 19(2), 195-216.
  - Walling, L. L. (2008). Avoiding effective defenses: strategies employed by phloem-feeding insects. *Plant Physiology*, 146(3), 859-866.
  - Wang, C. H., Yu, J., Cai, Y. X., Zhu, P. P., Liu, C. Y., Zhao, A. C., ... & Yu, M. D. (2016). Characterization and Functional Analysis of 4-Coumarate: CoA Ligase Genes in Mulberry. *PLoS one*, 11(5), e0155814.
  - Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in plant science*, 9(5), 244-252.
  - Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, 10(1), 57-63.
  - Webster, B., Bruce, T., Dufour, S., Birkemeyer, C., Birkett, M., Hardie, J., & Pickett, J. (2008). Identification of volatile compounds used in host location by the black bean aphid, *Aphis fabae*. *Journal of chemical ecology*, 34(9), 1153-1161.
  - Webster, B., Bruce, T., Pickett, J., & Hardie, J. (2010). Volatiles functioning as host cues in a blend become nonhost cues when presented alone to the black bean aphid. *Animal Behaviour*, 79(2), 451-457.
  - Wechter, W. P., Levi, A., Harris, K. R., Davis, A. R., Fei, Z., Katzir, N., ... & Tadmor, Y. (2008). Gene expression in developing watermelon fruit. *Bmc Genomics*, 9(1), 275.
  - Wei, Q., Wang, Y., Qin, X., Zhang, Y., Zhang, Z., Wang, J., ... & Chen, J. (2014). An SNP-based saturated genetic map and QTL analysis of fruit-related traits in cucumber using specific-length amplified fragment (SLAF) sequencing. *BMC genomics*, 15(1), 1158.
  - Weigel, R. R., Bäuscher, C., Pfitzner, A. J., & Pfitzner, U. M. (2001). NIMIN-1, NIMIN-2 and NIMIN-3, members of a novel family of proteins from Arabidopsis that interact with NPR1/NIM1, a key regulator of systemic acquired resistance in plants. *Plant molecular biology*, 46(2), 143-160.
  - Weigel, R. R., Pfitzner, U. M., & Gatz, C. (2005). Interaction of NIMIN1 with NPR1 modulates PR gene expression in Arabidopsis. *The Plant Cell*, 17(4), 1279-1291.
  - Whitaker, T. W., & Robinson, R. W. (1986). Squash breeding. *Breeding vegetable crops*, 209-242.
  - Wightman, R., Chomicki, G., Kumar, M., Carr, P., & Turner, S. R. (2013). SPIRAL2 determines plant microtubule organization by modulating microtubule severing. *Current Biology*, 23(19), 1902-1907.
  - Wildermuth, M. C., Dewdney, J., Wu, G., & Ausubel, F. M. (2001). Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature*, 414(6863), 562-565.
  - Will, T., & van Bel, A. J. (2006). Physical and chemical interactions between aphids and plants. *Journal of Experimental Botany*, 57(4), 729-737.
  - Will, T., & Vilcinskas, A. (2015). The structural sheath protein of aphids is required for phloem feeding. *Insect biochemistry and molecular biology*, 57, 34-40.

- Will, T., Furch, A. C., & Zimmermann, M. R. (2013). How phloem-feeding insects face the challenge of phloem-located defenses. *Frontiers in Plant Science*, 4.
- Will, T., Kornemann, S. R., Furch, A. C., Tjallingii, W. F., & van Bel, A. J. (2009). Aphid watery saliva counteracts sieve-tube occlusion: a universal phenomenon?. *Journal of Experimental Biology*, 212(20), 3305-3312.
- Will, T., Tjallingii, W. F., Thönnissen, A., & van Bel, A. J. (2007). Molecular sabotage of plant defense by aphid saliva. *Proceedings of the National Academy of Sciences*, 104(25), 10536-10541.
- Wu, J., & Baldwin, I. T. (2010). New insights into plant responses to the attack from insect herbivores. *Annual review of genetics*, 44, 1-24.
- Wu, T., Luo, S., Wang, R., Zhong, Y., Xu, X., Lin, Y. E., ... & Huang, H. (2014). The first Illumina-based *de novo* transcriptome sequencing and analysis of pumpkin (*Cucurbita moschata* Duch.) and SSR marker development. *Molecular breeding*, 34(3), 1437-1447.
- Wyatt, L. E., Strickler, S. R., Mueller, L. A., & Mazourek, M. (2015). An acorn squash (*Cucurbita pepo* ssp. *ovifera*) fruit and seed transcriptome as a resource for the study of fruit traits in *Cucurbita*. *Horticulture Research*, 2, hortres201470.
- Xanthopoulou, A., Ganopoulos, I., Psomopoulos, F., Manioudaki, M., Moysiadis, T., Kapazoglou, A., Osathanunkul, M., Michailidou, S., Kalivas, A., Tsaftaris, A., Nianiou-Obeidat, I., & Madesis, I. (2017). De novo comparative transcriptome analysis of genes involved in fruit morphology of pumpkin cultivars with extreme size difference and development of EST-SSR markers. *Gene*, 622, 50-66.
- Xanthopoulou, A., Psomopoulos, F., Ganopoulos, I., Manioudaki, M., Tsaftaris, A., Nianiou-Obeidat, I., & Madesis, P. (2016). De novo transcriptome assembly of two contrasting pumpkin cultivars. *Genomics data*, 7, 200-201.
- Xu, Z., Peters, R. J., Weirather, J., Luo, H., Liao, B., Zhang, X., ... & Au, K. F. (2015). Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *The Plant Journal*, 82(6), 951-961.
- Yahyaa, M., Tholl, D., Cormier, G., Jensen, R., Simon, P. W., & Ibdah, M. (2015). Identification and characterization of terpene synthases potentially involved in the formation of volatile terpenes in carrot (*Daucus carota* L.) roots. *J Agric Food Chem*, 63(19), 4870-8.
- Yamamoto, Y., Zhao, X., Suzuki, A. C., & Takahashi, S. Y. (1994). Cysteine proteinase from the eggs of the silkworm, *Bombyx mori*: Site of synthesis and a suggested role in yolk protein degradation. *Journal of insect physiology*, 40(5), 447-454.
- Yang, A., Zhou, Z., Pan, Y., Jiang, J., Dong, Y., Guan, X., Sun, H., Gao, S., & Chen, Z. (2016). RNA sequencing analysis to capture the transcriptome landscape during skin ulceration syndrome progression in sea cucumber *Apostichopus japonicus*. *BMC genomics*, 17(1), 459.
- Yang, L., Xie, C., Li, W., Zhang, R., Jue, D., & Yang, Q. (2013). Expression of a wild eggplant ribosomal protein L13a in potato enhances resistance to *Verticillium dahliae*. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 115(3), 329-340.
- Yoshinaga, N., Aboshi, T., Abe, H., Nishida, R., Alborn, H. T., Tumlinson, J. H., & Mori, N. (2008). Active role of fatty acid amino acid conjugates in nitrogen metabolism in *Spodoptera litura* larvae. *Proceedings of the National Academy of Sciences*, 105(46), 18058-18063.
- Yoshinaga, N., Aboshi, T., Ishikawa, C., Fukui, M., Shimoda, M., Nishida, R., ... & Mori, N. (2007). Fatty acid amides, previously identified in caterpillars, found in the cricket *Teleogryllus taiwanemma* and fruit fly *Drosophila melanogaster* larvae. *Journal of chemical ecology*, 33(7), 1376-1381.
- Zellnig, G., Pöckl, M. H., Möstl, S., & Zechmann, B. (2014). Two and three dimensional characterization of Zucchini Yellow Mosaic Virus induced structural alterations in *Cucurbita pepo* L. plants. *Journal of structural biology*, 186(2), 245-252.
- Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, 18(5), 821-829.
- Zhang, F., Zhu, L., & He, G. (2004). Differential gene expression in response to brown planthopper feeding in rice. *Journal of plant physiology*, 161(1), 53-62.
- Zhang, R., Calixto, C. P., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., ... & Frey dit Frey, N. (2017). A high quality *Arabidopsis* transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic acids research*, 45(9), 5061-5073.

- Zhang, R., Wang, B., Ouyang, J., Li, J., & Wang, Y. (2008). Arabidopsis indole synthase, a homolog of tryptophan synthase alpha, is an enzyme involved in the trp-independent indole-containing metabolite biosynthesis. *Journal of integrative plant biology*, 50(9), 1070-1077.
- Zhu, J., & Park, K. C. (2005). Methyl salicylate, a soybean aphid-induced plant volatile attractive to the predator *Coccinella septempunctata*. *Journal of chemical ecology*, 31(8), 1733-1746.
- Zhu, J., Obrycki, J. J., Ochieng, S. A., Baker, T. C., Pickett, J. A., & Smiley, D. (2005). Attraction of two lacewing species to volatiles produced by host plants and aphid prey. *Naturwissenschaften*, 92(6), 277-281.
- Zhu, Z., Xu, F., Zhang, Y., Cheng, Y. T., Wiermer, M., Li, X., & Zhang, Y. (2010). Arabidopsis resistance protein SNC1 activates immune responses through association with a transcriptional corepressor. *Proceedings of the National Academy of Sciences*, 107(31), 13960-13965.
- Zhu-Salzman, K. E. Y. A. N., Bi, J. L., & Liu, T. X. (2005). Molecular strategies of plant defense and insect counter-defense. *Insect Science*, 12(1), 3-15.
- Zhu-Salzman, K., Salzman, R. A., Ahn, J. E., & Koiwa, H. (2004). Transcriptional regulation of sorghum defense determinants against a phloem-feeding aphid. *Plant Physiology*, 134(1), 420-431.
- Zraidi, A., Stift, G., Pachner, M., Shojaeiyan, A., Gong, L., & Lelley, T. (2007). A consensus map for *Cucurbita pepo*. *Molecular breeding*, 20(4), 375-388.

## APPENDIX



**Figure A.1.** Scatter plot and correlation coefficient ( $r^2$ ) of normalized read count values of biological replicates of control samples collected at 24 hpi (A), 48 hpi (B) and 96 hpi (C).

Sequence name ( <i>C. pepo</i> )	Sequence name ( <i>A. gossypii</i> )	Identities	Gaps	e-value	Score
CUCPM_L16634_T_1 (1151 nt)	DN11910_c0_g1_i1 (1169 nt)	1146/1146 (100%)	0/1146 (0%)	0.0	2117 bits (1146)

DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	AATTAAACACATTACACGGCTGTACTAGTAGTCTGCAAGTTGTCATTCAATTTGTTTTT ACTAGTAGTCTGCAAGTTGTCATTCAATTTGTTTTT
consensus	AATTAAACACATTACACGGCTGTACTAGTAGTCTGCAAGTTGTCATTCAATTTGTTTTT
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	AAACAAGATAAAGTTCATAATGGCTAGGGTATTAATTTTATTGCTGTTATTTTGTTCAG AAACAAGATAAAGTTCATAATGGCTAGGGTATTAATTTTATTGCTGTTATTTTGTTCAG
consensus	AAACAAGATAAAGTTCATAATGGCTAGGGTATTAATTTTATTGCTGTTATTTTGTTCAG
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	TGTCATATGACAGAACAAGCATACTTTTGGAAAAAGACTACATCAACAAAATCAATGA TGTCATATGACAGAACAAGCATACTTTTGGAAAAAGACTACATCAACAAAATCAATGA
consensus	TGTCATATGACAGAACAAGCATACTTTTGGAAAAAGACTACATCAACAAAATCAATGA
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	AAAAGCATCAACATGGACGGCTGGTTTCAATTTTCGATCCATCCACACCGAAAGAAGACAT AAAAGCATCAACATGGACGGCTGGTTTCAATTTTCGATCCATCCACACCGAAAGAAGACAT
consensus	AAAAGCATCAACATGGACGGCTGGTTTCAATTTTCGATCCATCCACACCGAAAGAAGACAT
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	TTTAAAGCTCTTAGGATCAAAGGTGTACAACTCCAAGCAAAATTAACCTCAAATGTA TTTAAAGCTCTTAGGATCAAAGGTGTACAACTCCAAGCAAAATTAACCTCAAATGTA
consensus	TTTAAAGCTCTTAGGATCAAAGGTGTACAACTCCAAGCAAAATTAACCTCAAATGTA
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	CAAATCAGAAAGATGAAAACATGATAAATCTTTTGGCAGAATTCAAAAGAAATTTGATGC CAAATCAGAAAGATGAAAACATGATAAATCTTTTGGCAGAATTCAAAAGAAATTTGATGC
consensus	CAAATCAGAAAGATGAAAACATGATAAATCTTTTGGCAGAATTCAAAAGAAATTTGATGC
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	AAGGAAAAAATGGAGACATTGTACGACTATTGGAAAAGTCCGAGACC AAGGAAATTTGTGG AAGGAAAAAATGGAGACATTGTACGACTATTGGAAAAGTCCGAGACC AAGGAAATTTGTGG
consensus	AAGGAAAAAATGGAGACATTGTACGACTATTGGAAAAGTCCGAGACC AAGGAAATTTGTGG
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	AAGTTGTTGGGCCTTATCTACAAGCTCAGCGTTTGTGACCGTCTATGTGTAGCTACCAA AAGTTGTTGGGCCTTATCTACAAGCTCAGCGTTTGTGACCGTCTATGTGTAGCTACCAA
consensus	AAGTTGTTGGGCCTTATCTACAAGCTCAGCGTTTGTGACCGTCTATGTGTAGCTACCAA
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	CGGAGATTTCAATCAATTATTGTCCGAGAAGAAGTAACTTTCTGCTGTACAAAGTGTGG CGGAGATTTCAATCAATTATTGTCCGAGAAGAAGTAACTTTCTGCTGTACAAAGTGTGG
consensus	CGGAGATTTCAATCAATTATTGTCCGAGAAGAAGTAACTTTCTGCTGTACAAAGTGTGG
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	ATATGGCTGTAATGGAGGATACCCAATAAAAAGCATGGGAACGTTTTAAGAAAACACGGTCT ATATGGCTGTAATGGAGGATACCCAATAAAAAGCATGGGAACGTTTTAAGAAAACACGGTCT
consensus	ATATGGCTGTAATGGAGGATACCCAATAAAAAGCATGGGAACGTTTTAAGAAAACACGGTCT
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	TGTCAC TGGAGGAGAGTATAAATCAGGAGAGGGCTGTGAACCATACAGAGTTCTCCCTTG TGTCAC TGGAGGAGAGTATAAATCAGGAGAGGGCTGTGAACCATACAGAGTTCTCCCTTG
consensus	TGTCAC TGGAGGAGAGTATAAATCAGGAGAGGGCTGTGAACCATACAGAGTTCTCCCTTG
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	CCCATATGATGAATATGGAACAATACTTGCTCTGGA AAAACCGATGGAACAAAATCATAG CCCATATGATGAATATGGAACAATACTTGCTCTGGA AAAACCGATGGAACAAAATCATAG
consensus	CCCATATGATGAATATGGAACAATACTTGCTCTGGA AAAACCGATGGAACAAAATCATAG
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	ATGTACAAGAATGTGTACGGAGATCAAGACCTTGATTTTGATGACGACACACAGATACAC ATGTACAAGAATGTGTACGGAGATCAAGACCTTGATTTTGATGACGACACACAGATACAC
consensus	ATGTACAAGAATGTGTACGGAGATCAAGACCTTGATTTTGATGACGACACACAGATACAC
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	AAGAGACTCGTATTACCTTACATATGGAAGTATCCAAAAGACGTTATGACTTATGGTCC AAGAGACTCGTATTACCTTACATATGGAAGTATCCAAAAGACGTTATGACTTATGGTCC
consensus	AAGAGACTCGTATTACCTTACATATGGAAGTATCCAAAAGACGTTATGACTTATGGTCC
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	CATTGAAGCATCTTTTCGACGTTTATGACGATTTCCCCAGTTACAAGTCAGGAGTTTACGT CATTGAAGCATCTTTTCGACGTTTATGACGATTTCCCCAGTTACAAGTCAGGAGTTTACGT
consensus	CATTGAAGCATCTTTTCGACGTTTATGACGATTTCCCCAGTTACAAGTCAGGAGTTTACGT
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	TGCATCGGAAAATGCTTCATATTTAGGAGGACATGCAGTAAAAATGATTGGATGGGGTGA TGCATCGGAAAATGCTTCATATTTAGGAGGACATGCAGTAAAAATGATTGGATGGGGTGA
consensus	TGCATCGGAAAATGCTTCATATTTAGGAGGACATGCAGTAAAAATGATTGGATGGGGTGA
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	AGAATACGGAACCCCATATTTGGTTGATGATGAATTCATGGAACGAGCAATGGGGTGACCA AGAATACGGAACCCCATATTTGGTTGATGATGAATTCATGGAACGAGCAATGGGGTGACCA
consensus	AGAATACGGAACCCCATATTTGGTTGATGATGAATTCATGGAACGAGCAATGGGGTGACCA
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	AGGCTTTTCAA AATTCGACGAGGCACAAATGAATGTGGAATCGATAATTCGACTACTGG AGGCTTTTCAA AATTCGACGAGGCACAAATGAATGTGGAATCGATAATTCGACTACTGG
consensus	AGGCTTTTCAA AATTCGACGAGGCACAAATGAATGTGGAATCGATAATTCGACTACTGG
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	TGGTGTACCAGTAAC TAATTAATGCATCATTAGTTCAAACAATGTTAAATATTTTATGA TGGTGTACCAGTAAC TAATTAATGCATCATTAGTTCAAACAATGTTAAATATTTTATGA
consensus	TGGTGTACCAGTAAC TAATTAATGCATCATTAGTTCAAACAATGTTAAATATTTTATGA
DN11910_c0_g1_i1+ CUCPM_L16634_T_1-	TAAAA TAAAAAAAATTTGTTATCATTCAA TAAAA TAAAAAAAATTTGTTATCATTCAA
consensus	TAAAA TAAAAAAAATTTGTTATCATTCAA

**Figure A.2.** Sequence alignment of CUCPM\_L16634\_T\_1 and DN11910\_c0\_g1\_i1, from *C. pepo* and *A. gossypii* respectively, annotated as Cathepsin B. The portion highlight in red represents the amplicon sequence.



Sequence name ( <i>C. pepo</i> )	Sequence name ( <i>A. gossypii</i> )	Identities	Gaps	e-value	Score
CUCPM_TC17109 (1182 nt)	DN11910_c0_g3_i1 (728 nt)	681/689 (99%)	0/689 (0%)	0.0	1229 bits (665)
DN11910_c0_g3_i1+ CUCPM_TC17109-	TAAAATATGCAGATCGGAATTTATTCTACTACTATTGTACATATGAGTACACAGTAATAA CATATGAGTACACAGTAATAA				
consensus	TAAAATATGCAGATCGGAATTTATTCTACTACTATTGTACATATGAGTACACAGTAATAA				
DN11910_c0_g3_i1+ CUCPM_TC17109-	AACAAATTGTGATTTGACTTTTTTAATAATAAAAACAAAATTACAAATACAGCAACAATG AACAAATTGTGATTTGACTTTTTTAATAATAAAAACAAAATTACAAATACAGCAACAATG				
consensus	AACAAATTGTGATTTGACTTTTTTAATAATAAAAACAAAATTACAAATACAGCAACAATG				
DN11910_c0_g3_i1+ CUCPM_TC17109-	GCTAGGGTATTTATGCTATTGTCTGTGATATTCATCAGTGTTTACGTGACAGAAACAAGCA GCTAGGGTATTTATGCTATTGTCTGTGATATTCATCAGTGTTTACGTGACAGAAACAAGCA				
consensus	GCTAGGGTATTTATGCTATTGTCTGTGATATTCATCAGTGTTTACGTGACAGAAACAAGCA				
DN11910_c0_g3_i1+ CUCPM_TC17109-	TACTTTTGGGAAGAAGATTACATTAACAAAATCAATGAGCAAGCAACACATGGAAAGCA TACTTTTGGGAAGAAGATTACATTAACAAAATCAATGAGCAAGCAACACATGGAAAGCA				
consensus	TACTTTTGGGAAGAAGATTACATTAACAAAATCAATGAGCAAGCAACACATGGAAAGCA				
DN11910_c0_g3_i1+ CUCPM_TC17109-	GGCGCCAAC TTCGACCCAAAATGGCTGAAGAAAATTTTCGTAAAAC TTTGGGATCCAAA GGCGCCAAC TTCGACCCAAAATGGCTGAAGAAAATTTTCGTAAAAC TTTGGGATCCAAA				
consensus	GGCGCCAAC TTCGACCCAAAATGGCTGAAGAAAATTTTCGTAAAAC TTTGGGATCCAAA				
DN11910_c0_g3_i1+ CUCPM_TC17109-	GGAGTGCAAAATTCCAAACAAGTAAATCACAAAATGTACAAGACC GAAGATGAATCTTAC GGAGTGCAAAATTCCAAACAAGTAAATCACAAAATGTACAAGACC GAAGATGAATCTTAC				
consensus	GGAGTGCAAAATTCCAAACAAGTAAATCACAAAATGTACAAGACC GAAGATGAATCTTAC				
DN11910_c0_g3_i1+ CUCPM_TC17109-	GAAAAC T TATTCGGCAAAAATTCCAAAGCATT T TGACGCTAGGAAAAAATGGAGACGTTGC GACAAC T TATTCGGCAAAAATTCCAAAGCATT T TGACGCTAGGAAAAAATGGAGACGTTGC				
consensus	GAAAAC T TATTCGGCAAAAATTCCAAAGCATT T TGACGCTAGGAAAAAATGGAGACGTTGC				
DN11910_c0_g3_i1+ CUCPM_TC17109-	AGCACGATCGGAAAAGTTCGTGACCAAGGAAATTTGGGATCCTGTTGGGCATTGGCTACG AGCACGATCGGAAAAGTTCGTGACCAAGGAAATTTGGGATCCTGTTGGGCATTGGCTACG				
consensus	AGCACGATCGGAAAAGTTCGTGACCAAGGAAATTTGGGATCCTGTTGGGCATTGGCTACG				
DN11910_c0_g3_i1+ CUCPM_TC17109-	AGCTCTGCTTTTCGCCGATCGTTTGTGTGTAGCTACAAAACGGAGATTTCAATCAATTTGTTA AGCTCTGCTTTTCGCCGATCGTTTGTGTGTAGCTACAAAACGGAGATTTCAATCAATTTGTTA				
consensus	AGCTCTGCTTTTCGCCGATCGTTTGTGTGTAGCTACAAAACGGAGATTTCAATCAATTTGTTA				
DN11910_c0_g3_i1+ CUCPM_TC17109-	TCCGCCGAAGAAGTCACTTTCTGCTGTCATACATGTGGTTTCGGATGTCACGGTGGTTAT TCCGCCGAAGAAGTCACTTTCTGCTGTCATACATGTGGTTTCGGATGTCACGGTGGTTAT				
consensus	TCCGCCGAAGAAGTCACTTTCTGCTGTCATACATGTGGTTTCGGATGTCACGGTGGTTAT				
DN11910_c0_g3_i1+ CUCPM_TC17109-	CCAATAAGAGCCTGGAAACGTTTTAAGAATCACGGTCTAGTAACCGGAGGAGATTACAAA CCAATAAGAGCCTGGAAACGTTTTAAGAATCACGGTCTAGTAACCGGAGGAGATTACAAA				
consensus	CCAATAAGAGCCTGGAAACGTTTTAAGAATCACGGTCTAGTAACCGGAGGAGATTACAAA				
DN11910_c0_g3_i1+ CUCPM_TC17109-	TCCGGAGAGGGTTGTGAACCATACAGAGTGCCACCTTGCCCTTATGACGAACAAGGAAAT TCCGGAGAGGGTTGTGAACCATACAGAGTGCCACCTTGCCCTTATGACGAACAAGGAAAT				
consensus	TCCGGAGAGGGTTGTGAACCATACAGAGTGCCACCTTGCCCTTATGACGAACAAGGAAAT				
DN11910_c0_g3_i1+ CUCPM_TC17109-	AATACATG AATACATGCGCGGGTAAACCAATGGAAAAGAATCACAGATGCACAAGAACATGTTACGGGA				
consensus	AATACATGCGCGGGTAAACCAATGGAAAAGAATCACAGATGCACAAGAACATGTTACGGGA				
CUCPM_TC17109-	GATCAAGAGCTCGATTCGACGAAGATCACAGATACACACGTGACTACTACTATCTGACT GATCAAGAGCTCGATTCGACGAAGATCACAGATACACACGTGACTACTACTATCTGACT				
consensus	GATCAAGAGCTCGATTCGACGAAGATCACAGATACACACGTGACTACTACTATCTGACT				
CUCPM_TC17109-	TACGGCAGCATCCAAAAGACGTTATGACTTACGGACCAATTGAAGCATCGTTTGTATGTG TACGGCAGCATCCAAAAGACGTTATGACTTACGGACCAATTGAAGCATCGTTTGTATGTG				
consensus	TACGGCAGCATCCAAAAGACGTTATGACTTACGGACCAATTGAAGCATCGTTTGTATGTG				
CUCPM_TC17109-	TACAGCGATTTTCCAGCTACAAGTCAAGTATTTACGAGAGAACC GAAAATGCCACATAT TACAGCGATTTTCCAGCTACAAGTCAAGTATTTACGAGAGAACC GAAAATGCCACATAT				
consensus	TACAGCGATTTTCCAGCTACAAGTCAAGTATTTACGAGAGAACC GAAAATGCCACATAT				
CUCPM_TC17109-	TTAGGAGGACACGCTGTGAAGTTAATCGGTTGGGGTGAACAATACGGAATTCATATTGG TTAGGAGGACACGCTGTGAAGTTAATCGGTTGGGGTGAACAATACGGAATTCATATTGG				
consensus	TTAGGAGGACACGCTGTGAAGTTAATCGGTTGGGGTGAACAATACGGAATTCATATTGG				
CUCPM_TC17109-	TTGATGGTCAACTCATGGAACGAAGGTTGGGGTGACAACGGTTTATTCAAATTCGACGA TTGATGGTCAACTCATGGAACGAAGGTTGGGGTGACAACGGTTTATTCAAATTCGACGA				
consensus	TTGATGGTCAACTCATGGAACGAAGGTTGGGGTGACAACGGTTTATTCAAATTCGACGA				
CUCPM_TC17109-	GGCACAACGAATGCGGAGTCGATAATTCTACAAC T GCTGGTGTACCAAGTTACCAACTAA GGCACAACGAATGCGGAGTCGATAATTCTACAAC T GCTGGTGTACCAAGTTACCAACTAA				
consensus	GGCACAACGAATGCGGAGTCGATAATTCTACAAC T GCTGGTGTACCAAGTTACCAACTAA				
CUCPM_TC17109-	CAAAATTTGGTATTTGTCTAAGGATAACAATGTCATTTTTTTCTGTTTGAATAAAAATTA CAAAATTTGGTATTTGTCTAAGGATAACAATGTCATTTTTTTCTGTTTGAATAAAAATTA				
consensus	CAAAATTTGGTATTTGTCTAAGGATAACAATGTCATTTTTTTCTGTTTGAATAAAAATTA				
CUCPM_TC17109-	AGTTTCTTCTTAAAAAATAA AGTTTCTTCTTAAAAAATAA				
consensus	AGTTTCTTCTTAAAAAATAA				

**Figure A.3.** Sequence alignment of CUCPM\_TC17109 and DN11910\_c0\_g3\_i1, from *C. pepo* and *A. gossypii* respectively, annotated as Cathepsin B. The portion highlight in red represents the amplicon sequence.

Sequence name ( <i>C. pepo</i> )	Sequence name ( <i>A. gossypii</i> )	Identities	Gaps	e-value	Score
CUCPM_L16529_T_1 (551 nt)	DN2214_c0_g1_i1 (683 nt)	550/551 (99%)	0/551 (0%)	0.0	1013 bits (548)

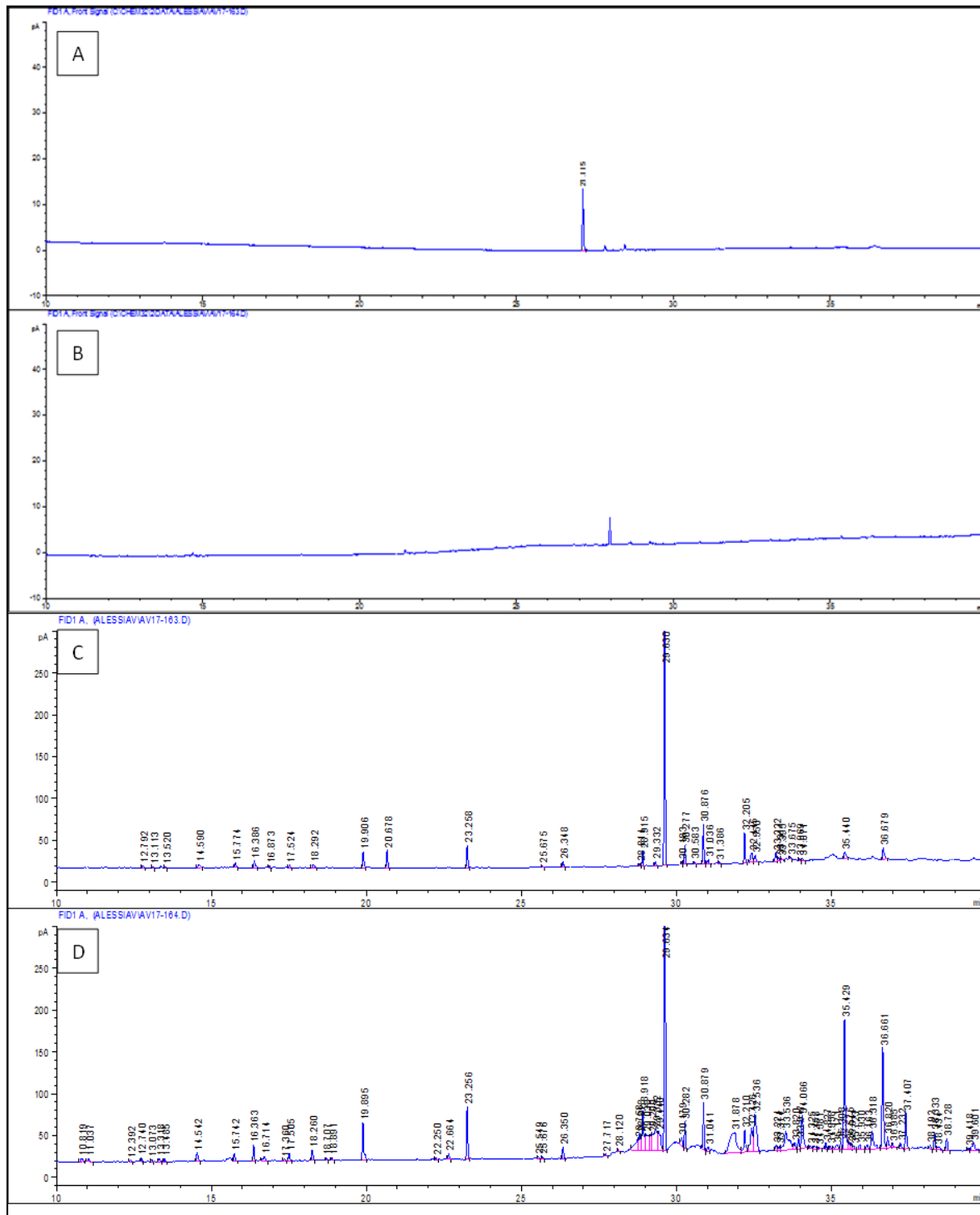
  

DN2214_c0_g1_i1-	CCATTAGTAAAGGCTATTATATATAAAAATCAATATTCATTA AAAATGTATAAGGATAAAA
consensus	CCATTAGTAAAGGCTATTATATATAAAAATCAATATTCATTA AAAATGTATAAGGATAAAA
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	TGGAAAATTA AAAAAAAAAACAATATGAGATTTGCATAAAAATGATCACTTTTATTAACCT AAATGATCACTTTTATTAACCT
consensus	TGGAAAATTA AAAAAAAAAACAATATGAGATTTGCATAAAAATGATCACTTTTATTAACCT
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	ATTTCTTCTTAGATACACCAACAGTTCTTCCTCTACGTCCAGTAGTTTTGGTATGCTGAC ATTTCTTCTTAGATACACCAACAGTTCTTCCTCTACGTCCAGTAGTTTTGGTATGCTGAC
consensus	ATTTCTTCTTAGATACACCAACAGTTCTTCCTCTACGTCCAGTAGTTTTGGTATGCTGAC
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	CACGTACTCTCAAACCCCAATAATGACGTAGACCACGGTGAGCCTTAATTTTCTTCAACC CACGTACTCTCAAACCCCAATAATGACGTAGACCACGGTGAGCCTTAATTTTCTTCAACC
consensus	CACGTACTCTCAAACCCCAATAATGACGTAGACCACGGTGAGCCTTAATTTTCTTCAACC
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	TTTCCAAATCTTCACGCAATTTGCTGTCAAGGGTACTAGAGGTAAGTTGAGAGAATTTTC TTTCCAAATCTTCACGCAATTTGCTGTCAAGGGTACTAGAGGTAAGTTGAGAGAATTTTC
consensus	TTTCCAAATCTTCACGCAATTTGCTGTCAAGGGTACTAGAGGTAAGTTGAGAGAATTTTC
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	CATCAACAACATCCTTTTGTCTATTCAAAAACAGTCAGGAATTTTATATTGACGAGGAT CATCAACAACATCCTTTTGTCTATTCAAAAACAGTCAGGAATTTTATATTGACGAGGAT
consensus	CATCAACAACATCCTTTTGTCTATTCAAAAACAGTCAGGAATTTTATATTGACGAGGAT
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	TTTGATAAATGTGATAATTTTTCAACCTCTTCGTCGGTACATTCTCCAGCTCTTTTGT TTTGATAAATGTGATAATTTTTCAACCTCTTCGTCGGTACATTCTCCAGCTCTTTTGT
consensus	TTTGATAAATGTGATAATTTTTCAACCTCTTCGTCGGTACATTCTCCAGCTCTTTTGT
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	TAAGGTCAACATCGGCCTTTTCAATACAATGTTGGAAAACCGTCTACCAATACCTTTGA TAAGGTCAACATCGGCCTTTTCAATACAATGTTGGAAAACCGTCTACCAATACCTTTGA
consensus	TAAGGTCAACATCGGCCTTTTCAATACAATGTTGGAAAACCGTCTACCAATACCTTTGA
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	TGGCGGTGATGGCGAACATGACTTTACGTTTGCCATCAATGTTGGTACTGAGGATACGCA TGGCGGTGATGGCGAACATGACTTTACGTTTGCCATCAATGTTGGTACTGAGGATACGCA
consensus	TGGCGGTGATGGCGAACATGACTTTACGTTTGCCATCAATGTTGGTACTGAGGATACGCA
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	AGATGTTGAAACTTTTCTGGGATGACCAAGACATTTTGGTAATGTTGTCGTCACAA AGATGTTGAAACTTTTCTGGGATGACCAAGACATTTTGGTAATGTTGTCGTCACAA
consensus	AGATGTTGAAACTTTTCTGGGATGACCAAGACATTTTGGTAATGTTGTCGTCACAA
DN2214_c0_g1_i1- CUCPM_L16529_T_1+	GCACGGGTGAAAAAGATTACAATTTGTACAACAGAGTCCTCGGTAAGAACGTGAAACG GCACGGGTGAAAAAGATTACAATTTGTACAACAGAGTCCTCGGTAAGAACGTGAAACG
consensus	GCACGGGTGAAAAAGATTACAATTTGTACAACAGAGTCCTCGGTAAGAACGTGAAACG
DN2214_c0_g1_i1-	CTATAAAACGTACCGGGTTATGG
consensus	CTATAAAACGTACCGGGTTATGG

**Figure A.4.** Sequence alignment of CUCPM\_L16529\_T\_1 and DN2214\_c0\_g1\_i1, from *C. pepo* and *A. gossypii* respectively, annotated as Ribosomal protein S18. The portion highlight in red represents the amplicon sequence.

Sequence name ( <i>C. pepo</i> )	Sequence name ( <i>A. gossypii</i> )	Identities	Gaps	e-value	Score
CUCPM_L16501_T_1 (386 nt)	DN18663_c0_g1_i1 (447 nt)	386/386 (100%)	0/386 (0%)	0.0	713 bits (386)
DN18663_c0_g1_i1- CUCPM_L16501_T_1+	AAATGCTAATTTTTTTAATGTCCTTATATTAGATAATGATCAAATTTTTGGAAATTACA CTTATATTAGATAATGATCAAATTTTTGGAAATTACA				
consensus	AAATGCTAATTTTTTTAATGTCCTTATATTAGATAATGATCAAATTTTTGGAAATTACA				
DN18663_c0_g1_i1- CUCPM_L16501_T_1+	AGTCATTCACTTCCCTCAAACGCTGATTGCGGATCTTACAGTAGTTGCAGCGGTGCTAG AGTCATTCACTTCCCTCAAACGCTGATTGCGGATCTTACAGTAGTTGCAGCGGTGCTAG				
consensus	AGTCATTCACTTCCCTCAAACGCTGATTGCGGATCTTACAGTAGTTGCAGCGGTGCTAG				
DN18663_c0_g1_i1- CUCPM_L16501_T_1+	AATATACGTAAGCTCCACCTGCTACAGTCTTCTACACCGTTTGCATGACCAAATTCCTA AATATACGTAAGCTCCACCTGCTACAGTCTTCTACACCGTTTGCATGACCAAATTCCTA				
consensus	AATATACGTAAGCTCCACCTGCTACAGTCTTCTACACCGTTTGCATGACCAAATTCCTA				
DN18663_c0_g1_i1- CUCPM_L16501_T_1+	CACAAGCTCTTTCATGGATTTCCTTCCGCAGAATGAACAGGTATATTTGCTGTGTTGGG CACAAGCTCTTTCATGGATTTCCTTCCGCAGAATGAACAGGTATATTTGCTGTGTTGGG				
consensus	CACAAGCTCTTTCATGGATTTCCTTCCGCAGAATGAACAGGTATATTTGCTGTGTTGGG				
DN18663_c0_g1_i1- CUCPM_L16501_T_1+	TTATTTCAATTTTCTTGACCATCTTACGCAGCGAGGCACCATATCGGGTACCATATTTAC TTATTTCAATTTTCTTGACCATCTTACGCAGCGAGGCACCATATCGGGTACCATATTTAC				
consensus	TTATTTCAATTTTCTTGACCATCTTACGCAGCGAGGCACCATATCGGGTACCATATTTAC				
DN18663_c0_g1_i1- CUCPM_L16501_T_1+	CTACAATTCCTACTTTCTTCGTACGTTTGCCATTTTGACGGCGGTGAAATCCTTTGGAC CTACAATTCCTACTTTCTTCGTACGTTTGCCATTTTGACGGCGGTGAAATCCTTTGGAC				
consensus	CTACAATTCCTACTTTCTTCGTACGTTTGCCATTTTGACGGCGGTGAAATCCTTTGGAC				
DN18663_c0_g1_i1- CUCPM_L16501_T_1+	TACTTCAATTAACACAAGTAAATAAAACGAGTGTCACTAACAAGAAGAGTTAGGAAAGGA TACTTCAATTAACACAAGTAAATAAAACGAGTGTCACTAACAAGAAGAG				
consensus	TACTTCAATTAACACAAGTAAATAAAACGAGTGTCACTAACAAGAAGAGTTAGGAAAGGA				
DN18663_c0_g1_i1-	AAAAAATATTCAGTGATTATACATAAT				
consensus	AAAAAATATTCAGTGATTATACATAAT				

**Figure A.5.** Sequence alignment of CUCPM\_L16501\_T\_1 and DN18663\_c0\_g1\_i1, from *C. pepo* and *A. gossypii* respectively, annotated as Ribosomal protein L37. The portion highlight in red represents the amplicon sequence.



**Figure A.6.** GC profile of headspace volatiles collected for 24 h from zucchini plants (B and D) infested with 10 *A. gossypii* for 48 h and (A and C) un-infested. Results obtained for molecules collected using Porapak Q (A and B) and Tenax TA (C and D) polymers are compared.

**Table A.1.** Number of reads generated from sequencing (*raw data*) and after quality filtering and adapter trimming (*clean data*) for each sample.

File Name	Raw data		Clean data
	# read	# paired read	# single read
A24_1_ATCACG_L007_R1_001.fastq	35,344,346	23,674,612	7,451,885
A24_1_ATCACG_L007_R2_001.fastq	35,344,346	23,674,612	
A24_2_CGATGT_L007_R1_001.fastq	35,230,308	23,732,387	7,382,828
A24_2_CGATGT_L007_R2_001.fastq	35,230,308	23,732,387	
A24_3_TTAGGC_L007_R1_001.fastq	34,366,050	22,964,925	7,204,783
A24_3_TTAGGC_L007_R2_001.fastq	34,366,050	22,964,925	
A48_2_TGACCA_L007_R1_001.fastq	37,211,641	25,261,715	7,571,915
A48_2_TGACCA_L007_R2_001.fastq	37,211,641	25,261,715	
A48_3_ACAGTG_L007_R1_001.fastq	36,623,056	24,901,053	7,442,903
A48_3_ACAGTG_L007_R2_001.fastq	36,623,056	24,901,053	
A48_4_GCCAAT_L007_R1_001.fastq	37,996,974	25,666,812	7,820,577
A48_4_GCCAAT_L007_R2_001.fastq	37,996,974	25,666,812	
A96_1_CAGATC_L006_R1_001.fastq	38,622,935	27,596,629	5,791,605
A96_1_CAGATC_L006_R2_001.fastq	38,622,935	27,596,629	
A96_2_ACTTGA_L006_R1_001.fastq	34,353,147	24,532,466	5,227,154
A96_2_ACTTGA_L006_R2_001.fastq	34,353,147	24,532,466	
A96_3_GATCAG_L006_R1_001.fastq	29,037,507	20,856,920	4,406,645
A96_3_GATCAG_L006_R2_001.fastq	29,037,507	20,856,920	
C24_1_TAGCTT_L006_R1_001.fastq	31,430,108	22,032,822	5,256,322
C24_1_TAGCTT_L006_R2_001.fastq	31,430,108	22,032,822	
C24_2_GGCTAC_L007_R1_001.fastq	28,740,043	21,066,607	4,270,278
C24_2_GGCTAC_L007_R2_001.fastq	28,740,043	21,066,607	
C24_3_CTTGTA_L007_R1_001.fastq	33,677,909	25,020,509	4,812,235
C24_3_CTTGTA_L007_R2_001.fastq	33,677,909	25,020,509	
C48_1_AGTCAA_L008_R1_001.fastq	36,265,357	28,423,670	4,540,249
C48_1_AGTCAA_L008_R2_001.fastq	36,265,357	28,423,670	
C48_3_AGTTCC_L008_R1_001.fastq	35,144,118	27,591,869	4,394,338
C48_3_AGTTCC_L008_R2_001.fastq	35,144,118	27,591,869	
C48_4_ATGTCA_L008_R1_001.fastq	37,518,763	28,808,248	5,107,225
C48_4_ATGTCA_L008_R2_001.fastq	37,518,763	28,808,248	
C96_2_CCGTCC_L008_R1_001.fastq	33,527,557	25,873,268	4,394,873
C96_2_CCGTCC_L008_R2_001.fastq	33,527,557	25,873,268	
C96_3_GTCCGC_L008_R1_001.fastq	31,060,525	24,146,031	4,067,956
C96_3_GTCCGC_L008_R2_001.fastq	31,060,525	24,146,031	
C96_4_GTGAAA_L008_R1_001.fastq	35,937,098	28,576,994	4,591,206
C96_4_GTGAAA_L008_R2_001.fastq	35,937,098	28,576,994	

**Table A.2\_A.** List of differentially expressed genes identified in zucchini plants at 24 h following infestation with *Aphis gossypii*.

ID transcript	logFC	PValue	FDR	Description
CUCPM_L10047_T_1	4.43960237	4.91E-08	7.45E-06	lumazine-binding family protein
CUCPM_L10539_T_10	4.22023204	0.000293	0.007905	ubiquitin-specific protease 16
CUCPM_L108_T_5	2.04389325	1.64E-09	4.47E-07	glutamate-1-semialdehyde 2,1-aminomutase 2
CUCPM_L1141_T_2	2.0478705	4.13E-10	1.44E-07	ribosomal protein L23AA
CUCPM_L11491_T_1	3.32864899	4.44E-05	0.00187	-
CUCPM_L11963_T_25	3.56864849	6.23E-06	0.000381	Inositol monophosphatase family protein
CUCPM_L12655_T_1	6.33086732	9.15E-06	0.00053	-
CUCPM_L13000_T_2	2.18969859	3.88E-05	0.001703	chlororespiratory reduction 7
CUCPM_L13419_T_1	3.28767151	0.002928	0.041528	-
CUCPM_L13518_T_1	4.17618266	7.39E-07	6.81E-05	heavy metal atpase 4
CUCPM_L14332_T_1	6.73036596	2.58E-10	1.01E-07	Tetratricopeptide repeat (TPR)-like superfamily protein
CUCPM_L14512_T_4	2.2639033	1.57E-13	3.66E-10	60S acidic ribosomal protein family
CUCPM_L14690_T_1	-2.7529835	4.28E-10	1.46E-07	unknown protein\x3b Has 416500
CUCPM_L15055_T_1	5.12649416	0.000494	0.011572	-
CUCPM_L15356_T_1	5.44650418	0.003117	0.043407	ATP synthase epsilon chain, mitochondrial
CUCPM_L15434_T_1	-5.6534475	0.000342	0.008682	Pentatricopeptide repeat (PPR) superfamily protein
CUCPM_L15699_T_1	6.60359443	1.55E-05	0.000819	Enolase
CUCPM_L15847_T_1	3.19928644	0.001628	0.027499	-
CUCPM_L1594_T_1	-4.4191992	6.94E-07	6.45E-05	unknown protein\x3b FUNCTIONS IN\x3a molecular_function unknown
CUCPM_L16076_T_1	5.00627743	0.000629	0.013836	-
CUCPM_L16114_T_1	-2.4073843	0.000578	0.013127	-
CUCPM_L1631_T_3	2.05219882	1.22E-10	5.68E-08	unknown protein\x3b FUNCTIONS IN\x3a molecular_function unknown\
CUCPM_L16538_T_1	5.34865229	0.001121	0.021321	Ribosomal L29e protein family
CUCPM_L16873_T_1	6.56097099	7.67E-06	0.000457	-
CUCPM_L17006_T_1	6.04659588	0.000611	0.013568	-
CUCPM_L17078_T_1	-2.6120256	1.64E-05	0.000842	flavodoxin family protein / radical SAM domain-containing protein
CUCPM_L17130_T_2	-4.5638272	0.000683	0.014544	Plastid-lipid associated protein PAP / fibrillin family protein
CUCPM_L1772_T_6	4.05399636	7.94E-05	0.00291	kinesin like protein for actin based chloroplast movement 2
CUCPM_L17768_T_6	2.88441072	3.05E-07	3.38E-05	transcription regulators
CUCPM_L1863_T_1	2.1656561	3.44E-10	1.23E-07	VIRB2-interacting protein 2
CUCPM_L18753_T_6	2.43965251	6.44E-12	5.99E-09	Ribosomal protein L10 family protein
CUCPM_L190_T_15	3.13466359	3.19E-11	1.86E-08	protochlorophyllide oxidoreductase B
CUCPM_L19052_T_1	2.23466818	4.28E-05	0.001828	nascent polypeptide-associated complex subunit alpha-like protein 2
CUCPM_L20074_T_1	2.70494844	9.06E-08	1.22E-05	-
CUCPM_L20195_T_5	2.25612023	5.17E-05	0.002109	ATP citrate lyase (ACL) family protein
CUCPM_L21191_T_10	-2.2247065	1.4E-09	3.91E-07	carbonic anhydrase 2
CUCPM_L21217_T_3	2.23758546	0.00177	0.029439	unknown protein\x3b FUNCTIONS IN\x3a molecular_function unknown
CUCPM_L21276_T_1	-5.0403369	3.88E-06	0.000266	-
CUCPM_L21314_T_3	2.49140147	0.000537	0.012455	AWPM-19-like family protein
CUCPM_L21651_T_1	2.38064679	0.000551	0.012714	-
CUCPM_L22399_T_1	-3.0673551	0.001536	0.026563	Plant protein of unknown function (DUF247)
CUCPM_L22592_T_1	-4.8116074	2.66E-08	4.37E-06	-
CUCPM_L22795_T_1	-2.4945306	0.003696	0.048611	-
CUCPM_L23200_T_1	-5.2618872	0.000634	0.01391	-
CUCPM_L23327_T_1	3.07758077	0.000584	0.013216	Yippee family putative zinc-binding protein
CUCPM_L23391_T_1	-5.7192325	0.00094	0.018584	-

CUCPM_L23412_T_4	2.0824176	8.69E-08	1.19E-05	unknown protein\x3b FUNCTIONS IN\x3a molecular_function unknown
CUCPM_L23544_T_1	-5.9368997	7.28E-05	0.00274	-
CUCPM_L23634_T_1	-7.3283526	2.68E-09	6.51E-07	-
CUCPM_L2377_T_3	2.22180864	2.82E-05	0.001321	Mitochondrial substrate carrier family protein
CUCPM_L23846_T_1	-6.0627621	0.002817	0.040788	Lactate/malate dehydrogenase family protein
CUCPM_L23931_T_1	-6.7780876	1.1E-09	3.33E-07	-
CUCPM_L24167_T_1	-2.0734166	0.002891	0.041279	-
CUCPM_L24223_T_1	-7.4878459	3.21E-07	3.5E-05	-
CUCPM_L24542_T_1	5.98366118	0.001063	0.020462	-
CUCPM_L2466_T_17	2.08857861	0.000875	0.017756	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
CUCPM_L24777_T_1	-6.2264576	0.000426	0.010368	-
CUCPM_L2479_T_2	2.00712835	4.18E-11	2.25E-08	Ribosomal protein PSRP-3/Ycf65
CUCPM_L24933_T_1	-2.8097542	8.21E-05	0.002975	-
CUCPM_L25732_T_1	6.17130804	0.000241	0.006903	-
CUCPM_L2599_T_5	2.31324824	0.001742	0.02911	63 kDa inner membrane family protein
CUCPM_L26217_T_1	-3.026101	0.001864	0.030427	double-stranded RNA-binding domain (DsRBD)-containing protein
CUCPM_L26546_T_1	2.54016496	0.000681	0.014535	-
CUCPM_L26579_T_1	-6.0450545	0.003407	0.046163	-
CUCPM_L2673_T_4	2.04973037	0.000175	0.005319	sucrose phosphate synthase 3F
CUCPM_L26954_T_1	-2.7124441	0.000274	0.007577	-
CUCPM_L27064_T_1	2.31468222	4.19E-05	0.001801	-
CUCPM_L27208_T_1	2.21317789	5.88E-06	0.000368	-
CUCPM_L27391_T_1	-5.9766213	2.04E-05	0.001019	Serine protease inhibitor, potato inhibitor I-type family protein
CUCPM_L27424_T_1	3.12673264	1.44E-07	1.78E-05	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
CUCPM_L277_T_2	2.1561521	1.23E-09	3.58E-07	60S acidic ribosomal protein family
CUCPM_L281_T_16	2.13087858	4.09E-09	9.21E-07	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein
CUCPM_L3230_T_2	2.0972592	1.21E-05	0.000677	RAD-like 1
CUCPM_L3630_T_4	-3.9053892	0.002834	0.040947	gamma tonoplast intrinsic protein
CUCPM_L3725_T_3	2.17407897	8.81E-05	0.003146	protochlorophyllide oxidoreductase C
CUCPM_L395_T_22	2.11326422	0.001507	0.026363	mitochondrial HSO70 2
CUCPM_L403_T_1	2.10043286	0.000279	0.007672	Cupredoxin superfamily protein
CUCPM_L4093_T_6	3.76114749	2.94E-05	0.001366	-
CUCPM_L536_T_4	-2.2133078	2.36E-05	0.001153	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
CUCPM_L540_T_7	2.03008277	9.44E-09	1.88E-06	Pollen Ole e 1 allergen and extensin family protein
CUCPM_L558_T_7	4.39171046	0.000166	0.005148	ribosomal protein L18
CUCPM_L6339_T_5	2.03882169	1.09E-10	5.46E-08	chloroplast heat shock protein 70-1
CUCPM_L6428_T_1	2.58265734	8.14E-05	0.00296	-
CUCPM_L666_T_6	2.00662795	1.22E-09	3.58E-07	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein
CUCPM_L7163_T_1	5.92293912	0.000179	0.005436	-
CUCPM_L7410_T_1	5.53781451	0.001047	0.020207	Zinc-binding ribosomal protein family protein
CUCPM_L7427_T_1	2.1187294	0.00191	0.030858	unknown protein
CUCPM_L7611_T_24	4.33796136	0.000859	0.017587	Mitochondrial transcription termination factor family protein
CUCPM_L8083_T_4	2.31083121	7.42E-07	6.81E-05	Ribosomal protein L39 family protein
CUCPM_L8937_T_1	-2.0122792	0.002379	0.036242	Serine protease inhibitor, potato inhibitor I-type family protein
CUCPM_L8960_T_3	2.7825882	5.22E-05	0.002119	Gibberellin-regulated family protein
CUCPM_L9010_T_14	2.77837845	0.000912	0.018177	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein
CUCPM_L9517_T_1	2.55506975	0.000898	0.018013	-
CUCPM_L960_T_4	2.04854799	5.78E-09	1.2E-06	photosystem I subunit G
CUCPM_L9932_T_7	2.47861792	0.003215	0.04425	-

CUCPM_TC10336	-2.2331126	0.000346	0.008694	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
CUCPM_TC10648	2.25279138	1.48E-05	0.000785	phototropin 1
CUCPM_TC10660	2.13762214	2.08E-06	0.000158	phototropin 1
CUCPM_TC11181	2.48298682	0.001542	0.026572	early nodulin-like protein 9
CUCPM_TC11462	2.23479541	0.000663	0.014313	Protein kinase superfamily protein
CUCPM_TC1215	2.04036704	4.04E-06	0.000273	Ribosomal L38e protein family
CUCPM_TC12186	2.16982478	0.000283	0.007691	C-terminal cysteine residue is changed to a serine 1
CUCPM_TC12250	2.20711958	0.00308	0.043202	Major facilitator superfamily protein
CUCPM_TC12321	2.56113572	0.000247	0.007001	EF hand calcium-binding protein family
CUCPM_TC12356	2.89371596	1.03E-09	3.18E-07	Thioredoxin superfamily protein
CUCPM_TC12656	2.02575987	3.08E-07	3.39E-05	myo-inositol oxygenase 4
CUCPM_TC1306	2.08879365	3.46E-12	4.11E-09	Translation protein SH3-like family protein
CUCPM_TC13473	2.21529063	0.00023	0.006654	Inositol monophosphatase family protein
CUCPM_TC13474	2.4750682	6.05E-08	8.7E-06	Inositol monophosphatase family protein
CUCPM_TC1380	2.08765826	0.002296	0.03531	Isochorismate synthase 1, chloroplastic
CUCPM_TC14009	2.04202201	0.000279	0.007672	Thioesterase superfamily protein
CUCPM_TC14437	-2.4600413	0.000246	0.006986	NAD(P)-binding Rossmann-fold superfamily protein
CUCPM_TC14576	2.08221988	0.00038	0.009426	NFU domain protein 3
CUCPM_TC14673	-2.9968405	0.003208	0.04425	Acyl-CoA N-acyltransferases (NAT) superfamily protein
CUCPM_TC15553	2.0325203	0.000113	0.003835	2Fe-2S ferredoxin-like superfamily protein
CUCPM_TC15918	-2.0450039	0.000411	0.010061	phytosulfokine 4 precursor
CUCPM_TC16054	2.41866379	5.29E-05	0.002126	Chaperone DnaJ-domain superfamily protein
CUCPM_TC16148	-2.9672269	5.94E-12	5.93E-09	unknown protein
CUCPM_TC166	3.17349078	0.000721	0.015193	thiaminC
CUCPM_TC1696	-2.0634391	3.26E-06	0.000232	CHY-type/CTCHY-type/RING-type Zinc finger protein
CUCPM_TC17084	3.21596074	0.00028	0.00768	glycine-rich protein 23
CUCPM_TC17141	5.65920883	2.28E-05	0.001118	-
CUCPM_TC17968	2.13059473	3.33E-06	0.000236	nudix hydrolase homolog 8
CUCPM_TC18604	-2.0034905	7.94E-05	0.00291	secretory carrier 3
CUCPM_TC18610	2.34730659	5.53E-05	0.002204	-
CUCPM_TC18931	2.06895702	2.38E-11	1.51E-08	acyl carrier protein 4
CUCPM_TC19781	2.01855905	4.14E-06	0.000275	-
CUCPM_TC19819	2.80483506	1.78E-09	4.68E-07	arabinogalactan protein 9
CUCPM_TC20042	2.63935973	1.4E-05	0.000752	4-coumarate\3aCoA ligase 1
CUCPM_TC20082	5.34865534	0.000896	0.018013	-
CUCPM_TC20338	2.01840752	0.003565	0.047657	Tetratricopeptide repeat (TPR)-like superfamily protein
CUCPM_TC2043	-3.0436223	0.000569	0.013002	proteasome alpha subunit F1
CUCPM_TC21759	-2.1755497	0.002456	0.037	BTB and TAZ domain protein 1
CUCPM_TC21874	-2.1698078	0.002691	0.039426	-
CUCPM_TC21896	-2.3281272	0.002162	0.033862	-
CUCPM_TC21956	3.25826756	7.47E-07	6.81E-05	Plant protein of unknown function (DUF247)
CUCPM_TC21987	3.59219089	9.25E-05	0.00326	-
CUCPM_TC22413	2.05823912	0.001186	0.02236	Ribulose bisphosphate carboxylase (small chain) family protein
CUCPM_TC2364	2.28403251	1.67E-09	4.47E-07	Nuclear transport factor 2 (NTF2) family protein with RNA binding (RRM-RBD-RNP motifs) domain
CUCPM_TC249	2.28932822	4.73E-08	7.26E-06	prochlorophyllide oxidoreductase A
CUCPM_TC2509	2.50577727	0.003761	0.049198	Glycosyl transferase, family 35
CUCPM_TC251	6.86558815	1.02E-20	7.09E-17	prochlorophyllide oxidoreductase A
CUCPM_TC3039	-2.5633857	0.001387	0.024886	unknown protein
CUCPM_TC3247	2.43798175	0.000679	0.014504	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein



CUCPM_TC3324	-2.618746	5.04E-08	7.46E-06	Aldolase-type TIM barrel family protein
CUCPM_TC376	2.38878474	4.19E-11	2.25E-08	peroxisomal NAD-malate dehydrogenase 2
CUCPM_TC3792	2.3771882	0.000207	0.006188	ATP synthase protein I -related
CUCPM_TC4107	2.31312233	0.000293	0.007902	germin 3
CUCPM_TC4178	2.07825192	0.000147	0.004666	adenosine kinase 2
CUCPM_TC420	3.72781946	5.49E-11	2.84E-08	Polyketide cyclase/dehydrase and lipid transport superfamily protein
CUCPM_TC443	4.12076626	8.77E-07	7.84E-05	actin depolymerizing factor 1
CUCPM_TC455	2.14682081	0.003758	0.049197	plasma membrane intrinsic protein 1x3b4
CUCPM_TC53	2.42277332	0.000215	0.006339	catalase 2
CUCPM_TC5303	-2.3181879	0.002787	0.040469	-
CUCPM_TC6405	2.53043399	0.000474	0.011302	Ankyrin repeat family protein
CUCPM_TC783	2.78631899	0.000301	0.008016	-
CUCPM_TC9465	-2.0313224	0.001653	0.027857	annexin 5

**Table A.2\_B.** List of differentially expressed genes identified in zucchini plants at 48 h following infestation with *Aphis gossypii*.

ID transcript	logFC	PValue	FDR	Description
CUCPM_L10047_T_1	2.56731	0.000513	0.005596	lumazine-binding family protein
CUCPM_L10444_T_1	3.080666	2.6E-06	7.58E-05	-
CUCPM_L1063_T_12	-2.16329	4.99E-06	0.000135	alpha-L-arabinofuranosidase 1
CUCPM_L10663_T_1	4.106108	1.72E-06	5.39E-05	unknown protein
CUCPM_L10727_T_1	6.564251	0.000572	0.006071	cytochrome P45, family 82, subfamily C, polypeptide 3
CUCPM_L108_T_5	2.66585	1.29E-16	4.74E-14	glutamate-1-semialdehyde 2,1-aminomutase 2
CUCPM_L10930_T_1	2.312127	2.09E-09	1.47E-07	SPIRAL1-like1
CUCPM_L11075_T_1	-2.29395	4.16E-05	0.000756	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
CUCPM_L1141_T_2	2.577452	2.72E-16	9.27E-14	ribosomal protein L23AA
CUCPM_L11426_T_17	2.922024	0.003968	0.026172	unknown protein
CUCPM_L11500_T_2	2.249296	1.88E-09	1.33E-07	antitermination NusB domain-containing protein
CUCPM_L11519_T_1	-2.26355	2.74E-06	7.93E-05	RING/U-box superfamily protein
CUCPM_L11717_T_1	2.597651	1.61E-06	5.06E-05	arabinogalactan protein 18
CUCPM_L11848_T_1	-2.07359	1.05E-05	0.000248	NIM1-interacting 1
CUCPM_L11960_T_1	2.817458	2.03E-05	0.000418	3-ketoacyl-CoA synthase 19
CUCPM_L1197_T_1	2.025749	1.46E-06	4.71E-05	CAX-interacting protein 2
CUCPM_L12014_T_3	2.12879	2.48E-08	1.36E-06	thylakoid lumen 15. kDa protein
CUCPM_L12025_T_5	-2.09424	0.00018	0.002399	threonine aldolase 1
CUCPM_L12046_T_2	2.15385	5.81E-06	0.000154	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
CUCPM_L1215_T_1	2.487432	9.12E-07	3.13E-05	Zinc-binding ribosomal protein family protein
CUCPM_L12207_T_1	3.097899	4.78E-06	0.00013	Lactate/malate dehydrogenase family protein
CUCPM_L12332_T_10	2.479975	2.52E-05	0.000502	rotamase CYP 7
CUCPM_L12399_T_7	2.303522	3.12E-10	2.42E-08	UDP-D-apiose/UDP-D-xylose synthase 2
CUCPM_L12439_T_12	2.448434	1.96E-05	0.000409	unknown protein
CUCPM_L12648_T_4	3.127979	0.004728	0.03012	CYCLIN D3x3b1
CUCPM_L12655_T_1	7.525256	5.06E-08	2.59E-06	-
CUCPM_L130_T_16	2.480732	2.1E-20	2.93E-17	fructose-bisphosphate aldolase 2
CUCPM_L13000_T_2	2.737063	4.26E-08	2.23E-06	chlororespiratory reduction 7
CUCPM_L13037_T_3	2.085912	0.002833	0.020429	Pollen Ole e 1 allergen and extensin family protein
CUCPM_L1319_T_1	2.331065	4.46E-23	1.56E-19	magnesium chelatase i2
CUCPM_L13300_T_1	5.363406	0.003082	0.021797	ABC-2 type transporter family protein
CUCPM_L13384_T_2	2.357384	0.000272	0.003316	HVA22-like protein F

CUCPM_L13548_T_2	2.180857	2.45E-10	1.96E-08	Defender against death (DAD family) protein
CUCPM_L13988_T_1	3.1265	2.25E-06	6.67E-05	unknown protein
CUCPM_L13999_T_1	6.376146	3.22E-05	0.000614	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
CUCPM_L1400_T_1	2.292945	2.82E-13	4.82E-11	Ribosomal protein L1 family protein
CUCPM_L14025_T_1	-3.64269	1.17E-05	0.00027	Protein of unknown function (DUF 3339)
CUCPM_L14050_T_1	2.11704	0.003926	0.026003	pleiotropic drug resistance 5
CUCPM_L14082_T_1	-2.68098	0.003459	0.023594	phytosulfokin receptor 1
CUCPM_L1409_T_12	2.623497	5.4E-08	2.71E-06	ATPase, F/V complex, subunit C protein
CUCPM_L14108_T_1	2.410835	0.000208	0.00269	Adenine nucleotide alpha hydrolases-like superfamily protein
CUCPM_L14233_T_1	-2.31837	0.004235	0.027566	-
CUCPM_L14269_T_1	3.315038	0.000386	0.004448	-
CUCPM_L14332_T_1	2.497562	4.82E-06	0.000131	Pentatricopeptide repeat-containing protein
CUCPM_L1447_T_2	2.173701	5E-16	1.58E-13	plastid-specific ribosomal protein 4
CUCPM_L14512_T_4	2.800056	8.06E-22	1.61E-18	6S acidic ribosomal protein family
CUCPM_L14578_T_1	2.60633	2.72E-13	4.79E-11	Ribosomal protein S3 family protein
CUCPM_L14585_T_1	2.235749	5.51E-08	2.75E-06	SAUR-like auxin-responsive protein family
CUCPM_L14893_T_1	3.056304	3.03E-11	2.98E-09	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
CUCPM_L1491_T_2	2.764957	0.000116	0.001689	-
CUCPM_L15151_T_1	3.644401	0.008451	0.046544	Pentatricopeptide repeat (PPR) superfamily protein
CUCPM_L15312_T_1	2.288835	9.17E-08	4.26E-06	-
CUCPM_L15356_T_1	6.893761	0.000238	0.002974	ATP synthase epsilon chain, mitochondrial
CUCPM_L1540_T_10	-2.41873	1.77E-08	1E-06	BTB and TAZ domain protein 1
CUCPM_L15459_T_1	-3.0327	0.000691	0.007063	vacuolar ATP synthase G3
CUCPM_L15528_T_1	2.326538	0.001741	0.014138	homeobox protein 25
CUCPM_L15699_T_1	7.25358	1.71E-06	5.36E-05	Enolase
CUCPM_L15744_T_1	5.690612	0.000747	0.007491	-
CUCPM_L15894_T_1	3.924406	0.00154	0.012946	-
CUCPM_L15916_T_1	4.41412	0.000249	0.003078	-
CUCPM_L1594_T_1	-3.21331	5.72E-05	0.000972	unknown protein
CUCPM_L15970_T_1	3.179272	0.003839	0.025572	-
CUCPM_L16055_T_2	-2.48375	7.39E-07	2.64E-05	unknown protein
CUCPM_L16073_T_1	2.055668	4.22E-05	0.000766	mitochondrial import receptor subunit TOM5 homolog
CUCPM_L16096_T_1	7.145129	9.24E-06	0.000223	Ribosomal protein S27a / Ubiquitin family protein
CUCPM_L161_T_11	2.756394	4.74E-12	5.51E-10	Ribosomal protein S4
CUCPM_L1631_T_3	2.100784	4.2E-12	4.97E-10	unknown protein
CUCPM_L1638_T_5	2.573424	0.000145	0.002017	Translation initiation factor 3 protein
CUCPM_L16501_T_1	6.782404	0.000101	0.001536	Zinc-binding ribosomal protein family protein
CUCPM_L16529_T_1	6.959961	4.95E-05	0.000866	Ribosomal protein S13/S18 family
CUCPM_L16538_T_1	7.116209	1E-05	0.000239	Ribosomal L29e protein family
CUCPM_L16571_T_2	3.630664	6.12E-06	0.000161	cytochrome P45, family 89, subfamily A, polypeptide 5
CUCPM_L16572_T_5	-4.36901	0.000114	0.001671	Modifier of rudimentary (Mod(r)) protein
CUCPM_L16634_T_1	6.171519	0.00026	0.003197	Cysteine proteinases superfamily protein
CUCPM_L16719_T_1	6.125979	0.007394	0.041879	-
CUCPM_L16823_T_1	2.346148	5.39E-05	0.000921	unknown protein
CUCPM_L16824_T_1	3.186778	0.000926	0.00882	-
CUCPM_L16859_T_1	2.357994	2.28E-05	0.000459	unknown protein
CUCPM_L16873_T_1	10.06164	2.94E-12	3.87E-10	-
CUCPM_L16896_T_1	6.784702	0.00063	0.006558	ATPase, V1 complex, subunit B protein
CUCPM_L1696_T_2	3.025867	3.85E-17	1.73E-14	RuBisCO large subunit-binding protein subunit alpha

CUCPM_L17224_T_1	2.61603	1.31E-12	1.87E-10	Arabidopsis thaliana protein match is\x3a sequence-specific DNA binding transcription factors\x3btranscription regulators
CUCPM_L17253_T_1	2.031531	0.003028	0.021527	Calcium-binding EF hand family protein
CUCPM_L17280_T_5	2.140064	0.004404	0.028426	unknown protein
CUCPM_L1749_T_3	2.288119	4.29E-06	0.000118	proline iminopeptidase
CUCPM_L17539_T_4	2.337289	0.000384	0.004427	O-Glycosyl hydrolases family 17 protein
CUCPM_L1761_T_14	2.387984	0.00022	0.002788	hydroxyproline-rich glycoprotein family protein
CUCPM_L17764_T_4	-2.3058	3.07E-05	0.000589	peroxidase 2
CUCPM_L17768_T_6	3.207842	6.73E-10	5.02E-08	BEST Arabidopsis thaliana protein match is\x3a transcription regulators
CUCPM_L17789_T_1	-2.54624	3.79E-08	2.02E-06	nudix hydrolase homolog 17
CUCPM_L1786_T_1	2.032598	3.34E-09	2.29E-07	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein
CUCPM_L17883_T_2	2.156252	4.43E-09	2.92E-07	Glutaredoxin family protein
CUCPM_L1805_T_1	2.663344	3.1E-06	8.86E-05	malate dehydrogenase
CUCPM_L18136_T_5	2.066032	3.45E-06	9.72E-05	tubulin alpha-2 chain
CUCPM_L18279_T_5	2.30425	8.31E-06	0.000206	aspartate aminotransferase 5
CUCPM_L18301_T_1	2.116495	2.95E-07	1.21E-05	emp24/gp25L/p24 family/GOLD family protein
CUCPM_L18317_T_5	2.127102	0.000895	0.008605	Thioredoxin superfamily protein
CUCPM_L18354_T_1	2.200134	1.87E-08	1.05E-06	RNA-binding CRS1 / YhbY (CRM) domain protein
CUCPM_L18386_T_1	3.121863	2.64E-17	1.31E-14	rotamase CYP 4
CUCPM_L18480_T_1	2.083648	2.3E-08	1.28E-06	FASCICLIN-like arabinogalactan 2
CUCPM_L1863_T_1	2.508653	7.99E-18	4.64E-15	Reticulon-like protein B2
CUCPM_L18753_T_6	3.106564	1.32E-19	1.59E-16	Ribosomal protein L1 family protein
CUCPM_L18756_T_4	2.285645	2.28E-15	6.24E-13	cytosolic ribosomal protein S15
CUCPM_L18787_T_2	2.89956	3.68E-08	1.97E-06	histone H2A 12
CUCPM_L18848_T_1	2.424259	5.46E-13	8.37E-11	Ribosomal protein S24e family protein
CUCPM_L18858_T_1	2.504989	6.94E-19	6.45E-16	ribosomal protein S9
CUCPM_L18870_T_1	2.735621	2.59E-06	7.56E-05	-
CUCPM_L190_T_15	3.313734	3.39E-12	4.26E-10	protochlorophyllide oxidoreductase B
CUCPM_L19052_T_1	2.174195	4.96E-05	0.000866	nascent polypeptide-associated complex subunit alpha-like protein 2
CUCPM_L1926_T_2	-2.61179	0.006579	0.038401	Cystathionine beta-synthase (CBS) family protein
CUCPM_L1932_T_4	2.757224	1.3E-11	1.34E-09	Nucleic acid-binding, OB-fold-like protein
CUCPM_L19409_T_2	2.806435	4.62E-05	0.000822	Reticulon family protein
CUCPM_L1947_T_4	2.12849	0.000144	0.002006	aldehyde dehydrogenase 5F1
CUCPM_L19631_T_1	6.743078	0.000122	0.00177	Ribosomal L29 family protein
CUCPM_L19667_T_1	3.266594	2.45E-06	7.21E-05	monodehydroascorbate reductase 1
CUCPM_L19671_T_1	-2.43142	0.001634	0.013507	-
CUCPM_L19740_T_1	2.0985	0.000161	0.002189	-
CUCPM_L20195_T_5	2.162473	0.002859	0.020597	ATP citrate lyase (ACL) family protein
CUCPM_L20235_T_5	-2.24098	2.7E-05	0.000532	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
CUCPM_L20277_T_2	-2.88499	9.87E-05	0.001514	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
CUCPM_L20388_T_1	-2.28714	0.004021	0.02642	Calcium-binding EF-hand family protein
CUCPM_L20452_T_2	3.309079	9.02E-14	1.7E-11	Ribosomal protein S25 family protein
CUCPM_L2054_T_6	2.005036	1.93E-15	5.49E-13	Ribosomal protein L14
CUCPM_L20565_T_1	3.386637	4.51E-05	0.000806	HPT phosphotransmitter 4
CUCPM_L20692_T_1	2.172734	4.33E-09	2.87E-07	unknown protein\x3b Plants - 537\x3b Viruses - \x3b Other Eukaryotes - 2996 (source\x3a NCBI BLink).
CUCPM_L20747_T_1	2.444294	1.88E-06	5.79E-05	HVA22-like protein F
CUCPM_L20750_T_1	-4.57987	1.48E-08	8.7E-07	calmodulin-like 11
CUCPM_L2096_T_2	-2.0367	1.31E-05	0.000294	cytochrome P45, family 72, subfamily A, polypeptide 7
CUCPM_L20978_T_1	2.684098	2.91E-14	5.97E-12	cell wall / vacuolar inhibitor of fructosidase 2

CUCPM_L21027_T_1	-2.17212	0.005119	0.032036	unknown protein
CUCPM_L21314_T_3	3.887528	3.27E-07	1.31E-05	AWPM-19-like family protein
CUCPM_L216_T_10	2.003244	1.01E-14	2.28E-12	phosphoribulokinase
CUCPM_L21622_T_1	5.455338	0.005707	0.034631	-
CUCPM_L21902_T_1	-3.29053	0.000968	0.009139	-
CUCPM_L22251_T_1	2.25923	4.98E-17	2.11E-14	YGGT family protein
CUCPM_L22592_T_1	-7.759	1.48E-10	1.27E-08	-
CUCPM_L22670_T_1	3.837848	1.43E-06	4.62E-05	-
CUCPM_L2270_T_1	2.53726	1.17E-09	8.57E-08	Aldolase superfamily protein
CUCPM_L22922_T_1	2.882594	4.15E-05	0.000755	-
CUCPM_L23044_T_1	-4.86057	0.006526	0.038252	-
CUCPM_L23657_T_1	2.509003	1.27E-08	7.63E-07	mitochondrial import receptor subunit TOM5 homolog
CUCPM_L2377_T_3	2.485557	1.2E-07	5.47E-06	Mitochondrial substrate carrier family protein
CUCPM_L23805_T_1	5.768031	0.002383	0.017765	-
CUCPM_L23877_T_1	-2.41773	0.000607	0.006362	-
CUCPM_L23931_T_1	-5.43839	2.16E-05	0.000439	-
CUCPM_L24116_T_1	-4.07024	6.21E-05	0.001037	unknown protein
CUCPM_L24168_T_1	-6.24015	0.00111	0.01016	-
CUCPM_L24185_T_1	-5.2818	0.001964	0.015408	-
CUCPM_L24296_T_1	-6.5897	3.91E-07	1.54E-05	-
CUCPM_L24322_T_1	-4.95185	0.001667	0.013685	-
CUCPM_L24393_T_1	-5.98178	0.004012	0.026395	-
CUCPM_L24420_T_1	2.442653	1.85E-16	6.63E-14	cytochrome-c oxidases\x3belectron carriers
CUCPM_L24488_T_1	-2.06032	0.001214	0.010844	-
CUCPM_L24516_T_2	-2.61086	0.000109	0.001615	Chaperonin-like RbcX protein
CUCPM_L24558_T_1	-6.51821	3.09E-05	0.00059	-
CUCPM_L24595_T_1	-6.25116	8.07E-05	0.001276	-
CUCPM_L2479_T_2	2.739062	6.41E-19	6.39E-16	Ribosomal protein PSRP-3/Ycf65
CUCPM_L24836_T_1	-2.02947	0.001947	0.015352	-
CUCPM_L2485_T_1	2.15999	1.78E-10	1.47E-08	histone H2A 1
CUCPM_L24963_T_1	-6.57772	0.001159	0.010515	proline-rich family protein
CUCPM_L25109_T_1	-2.4137	0.006265	0.037128	-
CUCPM_L25122_T_1	-6.08899	0.001688	0.013787	-
CUCPM_L25672_T_1	-5.9797	0.006762	0.039241	-
CUCPM_L25739_T_1	-2.34968	0.007308	0.041475	-
CUCPM_L25815_T_1	-5.98336	0.002827	0.020421	-
CUCPM_L2589_T_13	-2.25069	0.000234	0.002931	eif4a-2
CUCPM_L25972_T_1	2.293083	6.35E-17	2.45E-14	Translation elongation factor EF1B/ribosomal protein S6 family protein
CUCPM_L25986_T_1	2.05583	9.81E-05	0.001505	CONTAINS InterPro DOMAIN/slx3a Mediator complex subunit Med28
CUCPM_L2599_T_5	2.071054	0.003131	0.021995	63 kDa inner membrane family protein
CUCPM_L2605_T_2	2.012655	0.000184	0.002448	mitochondrial processing peptidase alpha subunit
CUCPM_L26054_T_1	2.094523	0.000229	0.002882	ribosomal protein S13A
CUCPM_L26102_T_1	2.344454	0.000794	0.007848	proline-rich protein 4
CUCPM_L26175_T_1	-2.47924	1.1E-05	0.000257	ROTUNDIFOLIA like 12
CUCPM_L26246_T_1	-3.55033	0.000101	0.001539	GDSL esterase/lipase 1
CUCPM_L26379_T_1	2.277162	0.000359	0.004174	-
CUCPM_L2643_T_11	2.049153	1.52E-11	1.55E-09	SPIRAL1-like1
CUCPM_L26560_T_1	-3.06015	0.00132	0.011569	glutathione S-transferase TAU 8
CUCPM_L26685_T_1	-5.92768	0.000285	0.003456	-

CUCPM_L26702_T_1	-2.85333	0.002124	0.016317	-
CUCPM_L26818_T_1	6.722238	1.95E-05	0.000407	-
CUCPM_L26841_T_1	-2.98969	1.3E-08	7.81E-07	-
CUCPM_L26906_T_1	-2.32342	0.000154	0.002115	Polynucleotidyl transferase, ribonuclease H-like superfamily protein
CUCPM_L27157_T_1	2.148273	1.76E-08	1E-06	ATP sulfurylase 2
CUCPM_L27208_T_1	2.61584	1.17E-06	3.9E-05	-
CUCPM_L27240_T_1	-2.62717	3.13E-07	1.27E-05	-
CUCPM_L27283_T_2	-2.25194	1.28E-05	0.000288	unknown protein
CUCPM_L27391_T_1	-6.11406	5.75E-06	0.000153	Serine protease inhibitor, potato inhibitor I-type family protein
CUCPM_L27424_T_1	4.306723	1.01E-11	1.04E-09	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
CUCPM_L27427_T_1	5.73849	0.001299	0.011456	proline-rich family protein
CUCPM_L277_T_2	2.957245	2.02E-17	1.04E-14	6S acidic ribosomal protein family
CUCPM_L281_T_16	2.687012	4.65E-16	1.51E-13	Ribosomal protein L7Ae/L3e/S12e/Gadd45 family protein
CUCPM_L293_T_2	2.069595	1.67E-10	1.39E-08	DnaJ/Hsp4 cysteine-rich domain superfamily protein
CUCPM_L3008_T_2	2.301221	5.83E-22	1.35E-18	ribosomal protein S1
CUCPM_L3012_T_6	2.076792	9.67E-06	0.000232	BEST Arabidopsis thaliana protein match is\x3a dentin sialophosphoprotein-related
CUCPM_L3092_T_14	-3.78864	0.005856	0.035288	low-molecular-weight cysteine-rich 68
CUCPM_L320_T_3	2.083975	3.26E-10	2.5E-08	heat shock cognate protein 7-1
CUCPM_L3318_T_1	2.409484	0.000178	0.002371	Glucan endo-1,3-beta-glucosidase 7
CUCPM_L348_T_23	-2.51549	6.99E-05	0.001148	Putative leucine-rich repeat receptor-like protein kinase
CUCPM_L3529_T_1	-3.40873	0.005184	0.032315	senescence-associated gene 21
CUCPM_L3665_T_2	-3.0325	0.006737	0.03911	-
CUCPM_L3690_T_1	2.843865	2.03E-06	6.19E-05	HVA22-like protein F
CUCPM_L3731_T_1	-6.4608	4.13E-08	2.18E-06	-
CUCPM_L3773_T_3	2.111836	9.06E-07	3.11E-05	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
CUCPM_L395_T_22	2.522258	0.000186	0.002462	mitochondrial HSO7 2
CUCPM_L403_T_1	2.830812	1.25E-06	4.13E-05	Cupredoxin superfamily protein
CUCPM_L4101_T_1	2.939171	3.69E-05	0.000685	plasmodesmata-located protein 2
CUCPM_L4128_T_1	2.009364	4.06E-08	2.15E-06	-
CUCPM_L4537_T_1	3.807594	0.001335	0.011673	-
CUCPM_L475_T_3	2.035739	4.03E-07	1.58E-05	-
CUCPM_L497_T_1	2.385492	4.18E-12	4.97E-10	cytosolic ribosomal protein S15
CUCPM_L517_T_1	2.381706	8.25E-06	0.000205	Histone superfamily protein Histone H4
CUCPM_L540_T_7	2.514898	3.08E-13	5.11E-11	Pollen Ole e 1 allergen and extensin family protein
CUCPM_L5430_T_1	-2.88258	0.000741	0.007454	-
CUCPM_L545_T_1	2.162099	1.05E-21	1.84E-18	elongation factor family protein
CUCPM_L5462_T_7	2.50384	2.94E-07	1.21E-05	Ribosomal protein L3/L7 family protein
CUCPM_L5498_T_1	2.688401	1.72E-08	9.88E-07	-
CUCPM_L556_T_3	2.512426	6.6E-12	7.26E-10	Ribosomal protein L41 family
CUCPM_L5625_T_1	2.889826	0.002066	0.015981	-
CUCPM_L566_T_3	2.321082	5.87E-17	2.34E-14	Ribosomal protein L16p/L1e family protein
CUCPM_L568_T_4	2.610798	9.16E-15	2.1E-12	chaperonin 2
CUCPM_L5845_T_8	2.627201	6.53E-05	0.001081	Ribosomal protein L34e superfamily protein
CUCPM_L5912_T_4	2.442062	2.01E-18	1.47E-15	Ribosomal protein L34
CUCPM_L6030_T_7	-2.5469	1.33E-07	5.95E-06	unknown protein
CUCPM_L6110_T_1	2.441057	6.51E-12	7.21E-10	Heavy metal transport/detoxification superfamily protein
CUCPM_L6339_T_5	2.693755	2.96E-18	1.88E-15	chloroplast heat shock protein 7-1
CUCPM_L638_T_2	2.716105	1.27E-30	8.85E-27	DnaJ/Hsp4 cysteine-rich domain superfamily protein
CUCPM_L652_T_4	-3.10528	7.02E-11	6.49E-09	-

CUCPM_L6532_T_3	2.062386	2.71E-07	1.12E-05	mitochondrial acyl carrier protein 2
CUCPM_L657_T_2	2.123373	1.01E-07	4.62E-06	Uncharacterized protein family (UPF16)
CUCPM_L6635_T_12	-2.5114	0.005521	0.033825	PATATIN-like protein 6
CUCPM_L666_T_6	2.848128	9.08E-18	5.07E-15	Ribosomal protein L7Ae/L3e/S12e/Gadd45 family protein
CUCPM_L67_T_12	2.324906	2.83E-13	4.82E-11	Ribosomal protein S3 family protein
CUCPM_L673_T_12	-2.20171	3.23E-05	0.000614	cytochrome P45, family 77, subfamily A, polypeptide 1
CUCPM_L6857_T_1	2.731	0.005676	0.034463	O-fucosyltransferase family protein
CUCPM_L7034_T_4	2.470098	0.001507	0.012731	unknown protein
CUCPM_L7127_T_1	6.040256	3.42E-05	0.000644	-
CUCPM_L728_T_3	-3.82424	0.000249	0.003083	Peroxidase superfamily protein
CUCPM_L739_T_1	2.401324	2.22E-08	1.25E-06	aldehyde dehydrogenase 2B4
CUCPM_L7410_T_1	6.927107	3.79E-05	0.0007	Zinc-binding ribosomal protein family protein
CUCPM_L7427_T_1	2.951844	2.13E-05	0.000436	unknown protein
CUCPM_L7455_T_2	2.066735	7.64E-05	0.001227	unknown protein
CUCPM_L7463_T_2	2.084409	5.37E-07	2.02E-05	tRNA synthetase class I (I, L, M and V) family protein
CUCPM_L7647_T_1	-3.22196	2.84E-07	1.17E-05	-
CUCPM_L7766_T_1	5.545169	0.001461	0.012434	-
CUCPM_L7845_T_3	2.602237	2.75E-13	4.79E-11	glyceraldehyde-3-phosphate dehydrogenase C subunit 1
CUCPM_L787_T_1	2.967444	0.000107	0.001606	Ribosomal protein L14p/L23e family protein
CUCPM_L793_T_1	2.444386	2.99E-12	3.9E-10	Ribosomal protein S3Ae
CUCPM_L8088_T_4	2.009474	2.76E-05	0.00054	Nucleotide-diphospho-sugar transferases superfamily protein
CUCPM_L8406_T_1	-2.48691	3.29E-05	0.000624	stress enhanced protein 2
CUCPM_L849_T_2	2.308975	9.77E-12	1.03E-09	Ribosomal protein L31
CUCPM_L8549_T_1	2.031211	0.000249	0.003078	-
CUCPM_L8682_T_5	2.011772	0.000224	0.002832	SGS domain-containing protein
CUCPM_L8841_T_3	2.113284	2.92E-16	9.71E-14	Ribosomal protein S21e
CUCPM_L8864_T_2	2.810523	3.15E-08	1.69E-06	Translation protein SH3-like family protein
CUCPM_L8889_T_2	2.391817	7.86E-12	8.44E-10	Ribosomal protein S11 family protein
CUCPM_L8937_T_1	-2.48947	0.001589	0.01324	Serine protease inhibitor, potato inhibitor I-type family protein
CUCPM_L8938_T_3	3.07181	9.82E-19	8.57E-16	RNA-binding (RRM/RBD/RNP motifs) family protein
CUCPM_L8960_T_3	2.559065	1.08E-05	0.000255	Gibberellin-regulated family protein
CUCPM_L8967_T_2	2.176154	1.59E-09	1.15E-07	Ribosomal protein L13 family protein
CUCPM_L9096_T_1	-2.03026	1.14E-07	5.21E-06	unknown protein
CUCPM_L9153_T_1	3.906069	0.000199	0.002596	-
CUCPM_L92_T_2	2.168296	6.18E-07	2.28E-05	ribosomal protein 1
CUCPM_L9238_T_1	2.958018	0.000884	0.00854	-
CUCPM_L928_T_12	2.290284	5.27E-12	6.08E-10	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
CUCPM_L9302_T_1	2.864215	7.32E-07	2.62E-05	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
CUCPM_L935_T_7	2.247825	1.37E-19	1.59E-16	fatty acid desaturase 8
CUCPM_L9358_T_1	-2.67794	8.17E-07	2.85E-05	-
CUCPM_L9364_T_1	6.744145	9.73E-05	0.001498	GTP binding Elongation factor Tu family protein
CUCPM_L947_T_1	2.597621	4.04E-09	2.7E-07	Histone superfamily protein Histone H3.2
CUCPM_L9485_T_1	2.379707	0.000425	0.004803	NADH/x3acytochrome B5 reductase 1
CUCPM_L9559_T_11	2.23872	4.13E-09	2.74E-07	beta-D-xylosidase 4
CUCPM_L9583_T_1	2.333539	0.001831	0.014727	Copper transport protein family
CUCPM_L96_T_2	2.454916	5.72E-15	1.39E-12	2 iron, 2 sulfur cluster binding
CUCPM_L960_T_4	2.547013	1.69E-12	2.38E-10	photosystem I subunit G
CUCPM_L9625_T_3	2.160942	0.000703	0.007155	Cation efflux family protein
CUCPM_L9692_T_1	-5.5246	0.004811	0.030505	-

CUCPM_L9727_T_1	2.175624	0.002598	0.019104	Seven transmembrane MLO family protein
CUCPM_TC10069	2.092605	0.000101	0.001537	unknown protein
CUCPM_TC10202	2.277856	1.48E-06	4.71E-05	sodium/x3ahydrogen antiporter 1
CUCPM_TC10254	2.193184	0.001218	0.010867	Small nuclear ribonucleoprotein family protein
CUCPM_TC10255	-5.70044	1.32E-05	0.000294	Small nuclear ribonucleoprotein family protein
CUCPM_TC10273	2.458815	3.65E-12	4.46E-10	Ribosomal protein L35Ae family protein
CUCPM_TC10274	3.268402	8.68E-15	2.02E-12	Ribosomal protein L35Ae family protein
CUCPM_TC10280	2.071666	8.12E-08	3.83E-06	Ribosomal protein S21e
CUCPM_TC10336	-2.34756	0.000194	0.002547	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
CUCPM_TC10339	-2.17883	0.002232	0.016925	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
CUCPM_TC10369	2.075101	1.73E-08	9.91E-07	Zinc-binding ribosomal protein family protein
CUCPM_TC10382	2.590707	6.17E-16	1.91E-13	Nucleic acid-binding, OB-fold-like protein
CUCPM_TC10391	2.494082	1.16E-10	1.01E-08	6S acidic ribosomal protein family
CUCPM_TC10421	2.855957	1.52E-23	7.08E-20	Thioredoxin superfamily protein
CUCPM_TC10463	3.447766	2.23E-08	1.25E-06	unknown protein
CUCPM_TC10497	2.308753	1.38E-05	0.000306	O-Glycosyl hydrolases family 17 protein
CUCPM_TC10542	-2.47085	0.007819	0.043858	-
CUCPM_TC10572	2.650185	4.56E-08	2.36E-06	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein
CUCPM_TC10630	2.864387	6.73E-06	0.000173	-
CUCPM_TC10663	2.384388	6.69E-09	4.19E-07	Rhodanese/Cell cycle control phosphatase superfamily protein
CUCPM_TC10879	2.325221	0.000845	0.008237	-
CUCPM_TC10986	2.320026	2.99E-06	8.58E-05	ribosomal protein L5 B
CUCPM_TC110	-2.82345	0.00398	0.026223	ribosomal protein 1
CUCPM_TC11018	-2.06587	4.64E-05	0.000824	ubiquitin-conjugating enzyme 2
CUCPM_TC11090	2.419439	1.32E-15	3.85E-13	Ribosomal L29e protein family
CUCPM_TC11181	3.226336	1.83E-05	0.000388	early nodulin-like protein 9
CUCPM_TC11219	2.566178	0.001353	0.011767	unknown protein
CUCPM_TC11333	2.45772	6.65E-09	4.18E-07	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein
CUCPM_TC11352	2.111032	4.58E-15	1.14E-12	Enolase
CUCPM_TC11355	-2.3984	0.000549	0.005875	Plant protein 1589 of unknown function
CUCPM_TC11397	2.675736	1.69E-08	9.77E-07	peroxiredoxin IIF
CUCPM_TC11460	2.075684	9.17E-06	0.000222	Protein kinase superfamily protein
CUCPM_TC1151	2.306979	8.37E-08	3.92E-06	Ribosomal protein S3Ae
CUCPM_TC11560	2.229723	0.003557	0.024078	unknown protein
CUCPM_TC1168	2.072426	5.66E-06	0.000151	Phototropic-responsive NPH3 family protein
CUCPM_TC11702	-4.13198	0.000105	0.001583	-
CUCPM_TC11792	4.274808	5.94E-06	0.000157	photosystem I subunit K
CUCPM_TC1184	3.597049	9.95E-21	1.54E-17	ribosomal protein L5
CUCPM_TC12123	2.091117	0.000175	0.002348	PsbP domain-containing protein 5, chloroplastic
CUCPM_TC1215	2.706993	1.61E-09	1.16E-07	Ribosomal L38e protein family
CUCPM_TC1218	3.708033	0.000411	0.004679	chloroplastic drought-induced stress protein of 32 kD
CUCPM_TC1222	2.245282	5.85E-07	2.18E-05	cellulose synthase-like A2
CUCPM_TC12321	2.910996	3.08E-05	0.00059	EF hand calcium-binding protein family
CUCPM_TC12356	3.004066	3.54E-15	9.16E-13	Thioredoxin superfamily protein
CUCPM_TC12357	2.576114	4.66E-05	0.000828	Thioredoxin superfamily protein
CUCPM_TC12542	2.089062	2.32E-09	1.62E-07	branched-chain aminotransferase 3
CUCPM_TC12656	2.351657	2.82E-10	2.23E-08	myo-inositol oxygenase 4
CUCPM_TC12665	2.128651	0.000292	0.003523	prohibitin 3
CUCPM_TC12669	-2.09776	2.15E-05	0.000438	NAC domain containing protein 1

CUCPM_TC12683	2.456277	2.51E-15	6.73E-13	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
CUCPM_TC12785	-3.28397	5.73E-09	3.67E-07	DNase I-like superfamily protein
CUCPM_TC12799	2.137677	1.92E-05	0.000403	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase, chloroplastic
CUCPM_TC12901	2.01201	1.94E-06	5.93E-05	plasma membrane intrinsic protein 3
CUCPM_TC12961	2.452381	1.41E-05	0.000311	Cytochrome b561/ferric reductase transmembrane protein family
CUCPM_TC13004	3.990338	0.000146	0.002034	Family of unknown function (DUF662)
CUCPM_TC13050	2.272119	5.22E-09	3.39E-07	Beta-carotene hydroxylase 2, chloroplastic
CUCPM_TC13052	2.113271	7.51E-15	1.78E-12	Ribosomal protein S5 family protein
CUCPM_TC1306	2.993469	3.54E-22	9.89E-19	Translation protein SH3-like family protein
CUCPM_TC1321	-2.89422	1.05E-05	0.000248	AAA-ATPase 1
CUCPM_TC13218	2.602861	3.69E-06	0.000103	Sec14p-like phosphatidylinositol transfer family protein
CUCPM_TC13244	2.42794	4.57E-08	2.36E-06	Protein of unknown function (DUF3353)
CUCPM_TC13415	-2.93133	7.12E-11	6.54E-09	Chaperone DnaJ-domain superfamily protein
CUCPM_TC13465	2.709689	2.91E-13	4.89E-11	Protein of unknown function (DUF1218)
CUCPM_TC13474	2.919799	1.48E-05	0.000322	Inositol monophosphatase family protein
CUCPM_TC13568	-2.17245	0.001574	0.013163	Peroxidase superfamily protein
CUCPM_TC1380	2.53077	0.000131	0.001864	Isochorismate synthase 1, chloroplastic
CUCPM_TC14009	2.029356	8.95E-05	0.001394	Thioesterase superfamily protein
CUCPM_TC1401	2.141234	5.93E-06	0.000157	ribosomal protein L23AA
CUCPM_TC1413	2.605425	1.13E-15	3.35E-13	Ribosomal protein L14p/L23e family protein
CUCPM_TC14419	2.230216	0.000778	0.00772	Pathogenesis-related thaumatin superfamily protein
CUCPM_TC14487	2.208767	5.26E-08	2.66E-06	Phosphoglucomutase/phosphomannomutase family protein
CUCPM_TC14515	-2.4091	1.79E-06	5.59E-05	endoribonuclease L-PSP family protein
CUCPM_TC14576	2.17676	3.5E-05	0.000656	NFU domain protein 3
CUCPM_TC1459	2.149572	6.99E-07	2.53E-05	CAX-interacting protein 2
CUCPM_TC15023	3.249639	0.000175	0.002342	GDSL-like Lipase/Acylhydrolase superfamily protein
CUCPM_TC15155	2.338844	1.4E-05	0.000311	Ribonuclease III family protein
CUCPM_TC15419	-2.50184	4E-09	2.68E-07	unknown protein
CUCPM_TC15559	2.180496	0.000274	0.003334	-
CUCPM_TC15731	2.228687	1.37E-06	4.43E-05	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
CUCPM_TC1599	2.089008	7.65E-13	1.14E-10	Magnesium-chelatase subunit ChII, chloroplastic
CUCPM_TC16077	2.014653	1.77E-11	1.78E-09	Ribosomal protein S3 family protein
CUCPM_TC16148	-2.11392	1.49E-06	4.73E-05	unknown protein
CUCPM_TC16162	2.594545	3.95E-13	6.41E-11	Ribosomal protein L39 family protein
CUCPM_TC1625	2.000937	7.51E-09	4.68E-07	cellulose synthase A2
CUCPM_TC1626	2.122015	0.000513	0.005596	cellulose synthase A2
CUCPM_TC1631	2.078321	1.48E-10	1.27E-08	Nascent polypeptide-associated complex (NAC), alpha subunit family protein
CUCPM_TC1632	2.175931	8.17E-13	1.2E-10	Nascent polypeptide-associated complex (NAC), alpha subunit family protein
CUCPM_TC16389	-2.58006	0.000685	0.00702	heat shock protein 9.1
CUCPM_TC16497	-2.0668	1.59E-07	6.88E-06	unknown protein
CUCPM_TC16518	2.097366	0.001338	0.011684	Aquaporin PIP2-2
CUCPM_TC16568	2.995488	1.81E-06	5.61E-05	Histone superfamily protein Histone H2A.6
CUCPM_TC16680	-2.4095	3.94E-05	0.000723	myb domain protein 94
CUCPM_TC16696	2.045755	5.37E-08	2.71E-06	Aquaporin PIP2-2
CUCPM_TC16752	2.61476	2.08E-06	6.29E-05	BCL-2-associated athanogene 1
CUCPM_TC16828	-3.55273	8.21E-05	0.001296	-
CUCPM_TC16881	2.10326	0.001181	0.010647	-
CUCPM_TC16900	-2.07306	6.26E-06	0.000164	Dormancy/auxin associated family protein
CUCPM_TC16926	2.665193	0.00186	0.014913	plantacyanin



CUCPM_TC16931	-2.93108	2.72E-06	7.89E-05	LORELEI-LIKE-GPI-ANCHORED PROTEIN 1
CUCPM_TC1696	-2.3727	1.85E-07	7.93E-06	CHY-type/CTCHY-type/RING-type Zinc finger protein
CUCPM_TC16970	2.473688	0.002205	0.016752	early nodulin-like protein 9
CUCPM_TC17062	10.86539	2.06E-10	1.69E-08	-
CUCPM_TC17084	6.372018	2.67E-10	2.12E-08	glycine-rich protein 23
CUCPM_TC17091	2.481944	3.62E-06	0.000101	unknown protein
CUCPM_TC17092	2.893089	5.25E-06	0.000142	-
CUCPM_TC17094	3.295936	1.32E-06	4.31E-05	-
CUCPM_TC1710	2.422951	3.75E-11	3.63E-09	Zinc-binding ribosomal protein family protein
CUCPM_TC17105	2.565182	0.000244	0.003033	Extensin-3
CUCPM_TC17106	2.273083	0.001002	0.009374	Leucine-rich repeat extensin-like protein 3
CUCPM_TC17109	4.527006	0.002014	0.015674	Cysteine proteinases superfamily protein
CUCPM_TC17141	10.89015	1.58E-11	1.6E-09	-
CUCPM_TC17148	4.368013	0.006416	0.037783	magnesium (Mg) transporter 1
CUCPM_TC1720	2.047899	2.89E-12	3.86E-10	ATPase, F/V complex, subunit C protein
CUCPM_TC1721	2.165671	1.53E-07	6.68E-06	ATPase, F/V complex, subunit C protein
CUCPM_TC17242	2.149155	0.009171	0.049302	translocon at the outer envelope membrane of chloroplasts 34
CUCPM_TC17334	-2.25028	0.000848	0.008263	formate dehydrogenase
CUCPM_TC17415	2.639497	2.28E-13	4.07E-11	copper/zinc superoxide dismutase 2
CUCPM_TC17696	2.438422	3.86E-05	0.000711	flavanone 3-hydroxylase
CUCPM_TC17726	-2.36891	0.000509	0.005569	-
CUCPM_TC17728	-2.78221	0.006146	0.036589	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
CUCPM_TC17749	2.429038	0.000398	0.004565	serine carboxypeptidase-like 7
CUCPM_TC17817	2.630003	4E-13	6.41E-11	Zinc-binding ribosomal protein family protein
CUCPM_TC17820	-2.59969	0.003328	0.022969	myb-like HTH transcriptional regulator family protein
CUCPM_TC17960	2.561362	1.24E-05	0.000284	Elongation of fatty acids protein
CUCPM_TC1799	2.004496	0.007171	0.040934	Transcription factor bHLH6
CUCPM_TC18025	2.58533	4.82E-08	2.47E-06	Transcription factor bHLH61
CUCPM_TC1808	2.014652	8.84E-09	5.39E-07	Glucose-methanol-choline (GMC) oxidoreductase family protein
CUCPM_TC18536	2.005209	1.8E-07	7.71E-06	Transducin/WD4 repeat-like superfamily protein
CUCPM_TC18555	2.473061	5.48E-11	5.2E-09	Thioredoxin superfamily protein
CUCPM_TC18638	-2.77913	0.002358	0.017626	Peroxidase superfamily protein
CUCPM_TC18780	2.349072	9.13E-16	2.77E-13	-
CUCPM_TC1881	-2.19182	1.25E-05	0.000285	BTB and TAZ domain protein 1
CUCPM_TC18912	-2.04889	7.46E-07	2.65E-05	Thioredoxin superfamily protein
CUCPM_TC18931	2.673851	2.22E-16	7.75E-14	acyl carrier protein 4
CUCPM_TC18932	2.633554	6.01E-06	0.000158	acyl carrier protein 2
CUCPM_TC18933	2.866887	9.8E-10	7.25E-08	acyl carrier protein 4
CUCPM_TC18969	-2.81388	3.61E-05	0.000672	glutathione S-transferase TAU 8
CUCPM_TC19037	2.089293	0.000198	0.002584	phytochrome-associated protein 2
CUCPM_TC19057	2.222548	1.37E-06	4.43E-05	-
CUCPM_TC19058	2.212423	0.00094	0.008932	Ribosomal protein L1 family protein
CUCPM_TC19064	2.259383	2.64E-05	0.000521	uracil phosphoribosyltransferase
CUCPM_TC19106	-2.32662	1.45E-05	0.000318	Thioredoxin superfamily protein
CUCPM_TC19115	2.146373	5.49E-07	2.06E-05	AIG2-like (avirulence induced gene) family protein
CUCPM_TC19175	2.288334	1.67E-05	0.000358	fibrillarlin 2
CUCPM_TC19296	2.195776	1.14E-06	3.81E-05	rotamase cyclophilin 2
CUCPM_TC19313	-3.00274	1.57E-10	1.33E-08	NAC domain containing protein 83
CUCPM_TC19797	-2.24817	1.4E-08	8.33E-07	NAC domain containing protein 83

CUCPM_TC19819	2.735132	1.67E-10	1.39E-08	arabinogalactan protein 9
CUCPM_TC19904	3.376038	0.003668	0.024669	-
CUCPM_TC19910	-4.30123	0.003885	0.025833	-
CUCPM_TC19914	-6.87457	3.56E-05	0.000667	-
CUCPM_TC19981	-2.07911	0.000805	0.007935	-
CUCPM_TC19985	2.276073	0.001028	0.009572	unknown protein
CUCPM_TC20042	3.628606	5.44E-17	2.23E-14	4-coumarate\3aCoA ligase 1
CUCPM_TC20082	5.167985	0.000102	0.001553	-
CUCPM_TC20116	2.070478	2.34E-07	9.83E-06	-
CUCPM_TC20130	2.406839	6.47E-06	0.000167	Gibberellin-regulated family protein
CUCPM_TC20198	-2.39577	2.97E-05	0.000573	Plant protein 1589 of unknown function
CUCPM_TC2023	2.120641	7.51E-08	3.59E-06	unknown protein
CUCPM_TC20236	-2.3961	2.02E-05	0.000417	F-box family protein
CUCPM_TC20256	2.247452	1.63E-06	5.14E-05	hydroxyproline-rich glycoprotein family protein
CUCPM_TC20270	3.846751	8.04E-07	2.82E-05	AWPM-19-like family protein
CUCPM_TC20338	2.363596	0.001708	0.013924	Pentatricopeptide repeat-containing protein
CUCPM_TC20529	-2.15147	0.002139	0.016398	Calcium-binding EF-hand family protein
CUCPM_TC20594	2.222718	0.003934	0.026018	Plant protein of unknown function (DUF247)
CUCPM_TC20624	2.013161	0.008066	0.044957	-
CUCPM_TC20679	-2.52754	0.003062	0.021702	nudix hydrolase 1
CUCPM_TC20781	2.018013	0.000144	0.002006	Polynucleotidyl transferase, ribonuclease H-like superfamily protein
CUCPM_TC20783	2.348906	7.46E-06	0.000189	Nucleic acid-binding, OB-fold-like protein
CUCPM_TC20877	2.484151	7.04E-08	3.39E-06	Hyaluronan / mRNA binding family
CUCPM_TC20954	2.321026	2.43E-10	1.96E-08	Leucine-rich repeat extensin-like protein 3
CUCPM_TC20998	2.130342	7.82E-09	4.85E-07	Mitochondrial substrate carrier family protein
CUCPM_TC2105	2.91097	4.53E-15	1.14E-12	RuBisCO large subunit-binding protein subunit alpha
CUCPM_TC21085	-2.05106	2.38E-05	0.000477	myb-like transcription factor family protein
CUCPM_TC21320	-5.7223	0.000107	0.001602	Serine protease inhibitor, potato inhibitor I-type family protein
CUCPM_TC21344	-2.2366	7.15E-06	0.000182	response regulator 9
CUCPM_TC21595	-2.36448	5.59E-08	2.78E-06	Thioredoxin-like protein
CUCPM_TC21667	2.69004	4.97E-05	0.000866	Phosphate-responsive 1 family protein
CUCPM_TC21759	-4.59497	5.9E-08	2.9E-06	BTB and TAZ domain protein 1
CUCPM_TC21761	-3.2048	0.000822	0.008046	BTB and TAZ domain protein 1
CUCPM_TC21816	2.362553	0.002249	0.017006	-
CUCPM_TC21817	2.269263	0.000107	0.001602	Protein of unknown function, DUF538
CUCPM_TC21839	-2.97505	5.91E-07	2.19E-05	Duplicated homeodomain-like superfamily protein
CUCPM_TC21956	2.182488	7.8E-05	0.001247	Plant protein of unknown function (DUF247)
CUCPM_TC220	2.453867	5.23E-13	8.12E-11	Ribosomal protein S4
CUCPM_TC22032	-4.14942	1.74E-09	1.25E-07	Heavy metal transport/detoxification superfamily protein
CUCPM_TC221	2.161225	5.46E-12	6.24E-10	Ribosomal protein S4
CUCPM_TC22142	-2.25866	2.05E-07	8.69E-06	Protein of unknown function (DUF567)
CUCPM_TC22161	2.312761	1.55E-05	0.000335	Abscisic acid-responsive (TB2/DP1, HVA22) family protein
CUCPM_TC222	2.225319	8.87E-14	1.7E-11	Ribosomal protein S4
CUCPM_TC22200	2.149118	4.06E-08	2.15E-06	Chlorophyll A-B binding family protein
CUCPM_TC22205	-2.20082	0.00351	0.023813	unknown protein
CUCPM_TC223	2.559998	1.68E-18	1.3E-15	Ribosomal protein L31e family protein
CUCPM_TC22331	2.491743	9.98E-12	1.04E-09	Ribosomal protein L18e/L15 superfamily protein
CUCPM_TC22339	2.57313	6.48E-17	2.45E-14	Ribosomal L29 family protein
CUCPM_TC2241	2.32596	1.3E-13	2.38E-11	Ribosomal protein S1p/S2e family protein

CUCPM_TC22434	2.445002	0.000265	0.003252	Proline-rich receptor-like protein kinase
CUCPM_TC22452	2.415634	3.24E-07	1.3E-05	FASCIKLIN-like arabinogalactan-protein 1
CUCPM_TC228	2.406584	0.000844	0.00823	Ethylene-responsive transcription factor RAP2-7
CUCPM_TC2364	2.876961	7.72E-12	8.35E-10	Nuclear transport factor 2 (NTF2) family protein with RNA binding (RRM-RBD-RNP motifs) domain
CUCPM_TC238	2.540275	6.11E-12	6.87E-10	Xanthine/uracil permease family protein
CUCPM_TC2386	2.213555	9.15E-12	9.74E-10	Nucleic acid-binding, OB-fold-like protein
CUCPM_TC248	2.125529	3.35E-12	4.25E-10	zeaxanthin epoxidase (ZEP) (ABA1)
CUCPM_TC249	2.448522	6.33E-09	4.03E-07	protochlorophyllide oxidoreductase A
CUCPM_TC250	7.337826	1.78E-08	1.01E-06	protochlorophyllide oxidoreductase A
CUCPM_TC251	4.293942	8.08E-11	7.33E-09	protochlorophyllide oxidoreductase A
CUCPM_TC2531	2.469616	0.001368	0.011842	subtilase 1.3
CUCPM_TC2590	2.627626	5.43E-14	1.08E-11	Rubredoxin-like superfamily protein
CUCPM_TC2830	2.045199	1.43E-07	6.26E-06	HMG-box (high mobility group) DNA-binding family protein
CUCPM_TC2939	2.381429	0.000219	0.002788	beta-galactosidase 1
CUCPM_TC316	2.597688	1.81E-14	3.95E-12	nucleoside diphosphate kinase 2
CUCPM_TC3222	2.219532	8.82E-05	0.001378	EXORDIUM like 5
CUCPM_TC3243	2.033771	2.51E-10	2E-08	RAN GTPase 3
CUCPM_TC3247	2.466684	0.000981	0.009217	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
CUCPM_TC3280	2.132935	1.07E-12	1.55E-10	plastid-specific 5S ribosomal protein 6
CUCPM_TC3317	2.286804	2.27E-10	1.84E-08	phosphate transporter 3\3b1
CUCPM_TC3324	-2.4624	7.64E-07	2.7E-05	Aldolase-type TIM barrel family protein
CUCPM_TC3327	2.682925	0.000568	0.006042	Protein kinase protein with adenine nucleotide alpha hydrolases-like domain
CUCPM_TC3339	-2.27562	2.99E-06	8.58E-05	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
CUCPM_TC3384	2.034612	2.45E-11	2.42E-09	ATP binding
CUCPM_TC3635	-2.30967	3.79E-07	1.5E-05	Putative pentatricopeptide repeat-containing protein
CUCPM_TC373	2.002608	1.67E-08	9.74E-07	triosephosphate isomerase
CUCPM_TC376	2.047122	7.78E-08	3.69E-06	peroxisomal NAD-malate dehydrogenase 2
CUCPM_TC3790	2.902019	0.002794	0.020253	ATP synthase protein I -related
CUCPM_TC3791	2.461713	0.001792	0.014493	ATP synthase protein I -related
CUCPM_TC3792	2.626616	0.001802	0.014565	ATP synthase protein I -related
CUCPM_TC3864	2.157774	5.25E-09	3.39E-07	PsbP-like protein 1
CUCPM_TC3933	2.438346	9.54E-11	8.48E-09	Histone superfamily protein Histone H2B
CUCPM_TC4036	2.344667	4.77E-09	3.13E-07	Ribosomal L28e protein family
CUCPM_TC4107	3.563655	1.34E-07	5.98E-06	germin 3
CUCPM_TC4114	2.099314	1.34E-07	5.98E-06	Beta-carotene isomerase D27, chloroplastic
CUCPM_TC4155	2.181786	0.002308	0.017384	FKBP-like peptidyl-prolyl cis-trans isomerase family protein
CUCPM_TC4167	-2.03357	0.004792	0.030426	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
CUCPM_TC419	2.213153	1.11E-05	0.00026	Polyketide cyclase/dehydrase and lipid transport superfamily protein
CUCPM_TC420	4.039901	2.85E-12	3.86E-10	Polyketide cyclase/dehydrase and lipid transport superfamily protein
CUCPM_TC421	2.4982	0.000192	0.002529	Polyketide cyclase/dehydrase and lipid transport superfamily protein
CUCPM_TC4268	2.383558	5.77E-15	1.39E-12	Ribosomal protein L16p/L1e family protein
CUCPM_TC443	3.495965	1.99E-05	0.000414	actin depolymerizing factor 1
CUCPM_TC4482	-2.33086	4.24E-05	0.000768	SEUSS-like 2
CUCPM_TC452	2.140588	5.4E-10	4.07E-08	Probable aquaporin PIP1-4
CUCPM_TC456	2.249082	0.000172	0.00232	Probable aquaporin PIP1-4
CUCPM_TC4611	-2.03607	0.000781	0.007721	cellulose synthase like E1
CUCPM_TC464	2.104129	8.96E-09	5.44E-07	Cupredoxin superfamily protein
CUCPM_TC476	-2.05804	5.54E-11	5.22E-09	-
CUCPM_TC4793	-2.03865	5.18E-08	2.63E-06	unknown protein

CUCPM_TC49	-2.43926	6.54E-05	0.001082	catalase 2
CUCPM_TC5080	2.228233	1.09E-05	0.000256	HVA22-like protein F
CUCPM_TC5083	2.307399	1E-13	1.86E-11	Ribosomal protein L19 family protein
CUCPM_TC5181	2.332678	5.25E-11	5.02E-09	ascorbate peroxidase 4
CUCPM_TC5303	-3.18569	0.000105	0.001582	-
CUCPM_TC5304	-2.49823	6.33E-06	0.000165	-
CUCPM_TC555	2.015003	1.05E-05	0.000248	trigger factor type chaperone family protein
CUCPM_TC558	2.202563	6.35E-13	9.53E-11	Ribosomal protein S7e family protein
CUCPM_TC569	2.432869	3.88E-12	4.71E-10	Ribosomal protein S3 family protein
CUCPM_TC570	2.280557	5.47E-09	3.52E-07	Ribosomal protein S3 family protein
CUCPM_TC5704	2.212273	7.72E-14	1.5E-11	ubiquitin 6
CUCPM_TC5771	2.873151	0.002959	0.021166	sucrose synthase 6
CUCPM_TC5827	2.414778	9.62E-05	0.001484	L-ascorbate oxidase homolog
CUCPM_TC5969	2.586246	1.39E-17	7.45E-15	Ribosomal protein L17 family protein
CUCPM_TC6021	-2.0907	2.63E-05	0.000521	Auxin-responsive family protein
CUCPM_TC6255	2.264527	0.001205	0.010806	Protein of unknown function, DUF547
CUCPM_TC635	2.235403	3.25E-11	3.17E-09	Pollen Ole e 1 allergen and extensin family protein
CUCPM_TC6398	-2.44041	0.001088	0.010018	Serine protease inhibitor, potato inhibitor I-type family protein
CUCPM_TC6405	2.488194	0.0003	0.003606	Ankyrin repeat family protein
CUCPM_TC7026	2.308378	3.13E-13	5.13E-11	Ribosomal protein S19e family protein
CUCPM_TC7027	2.558763	6.27E-12	7E-10	Ribosomal protein S19e family protein
CUCPM_TC742	2.212039	4.4E-07	1.71E-05	thylakoid rhodanese-like
CUCPM_TC7427	2.179778	0.000347	0.004065	FASCICLIN-like arabinogalactan 7
CUCPM_TC7491	2.225816	0.000333	0.003947	eukaryotic translation initiation factor 3K
CUCPM_TC75	-2.47297	8.76E-06	0.000214	senescence-associated gene 21
CUCPM_TC780	2.684335	3.76E-05	0.000696	-
CUCPM_TC783	3.15138	1.99E-05	0.000414	-
CUCPM_TC83	-2.94679	1.02E-06	3.47E-05	Protein kinase superfamily protein
CUCPM_TC8506	2.625663	4.58E-18	2.78E-15	ribosomal protein L23AA
CUCPM_TC859	-2.39116	1.94E-05	0.000407	Peroxidase superfamily protein
CUCPM_TC8718	-2.26625	7.69E-06	0.000193	thylakoid processing peptide
CUCPM_TC8780	3.162716	2.82E-17	1.36E-14	NDH-dependent cyclic electron flow 1
CUCPM_TC8928	2.026686	9.38E-06	0.000226	Isochorismatase family protein
CUCPM_TC9335	-2.36071	0.003652	0.024573	Formin-like protein 5
CUCPM_TC9431	-2.98533	2.07E-06	6.29E-05	RmIC-like cupins superfamily protein
CUCPM_TC9435	-2.20457	0.00023	0.002885	Protein with RING/U-box and TRAF-like domains
CUCPM_TC9452	2.496521	0.000128	0.001832	Peroxidase superfamily protein
CUCPM_TC9502	2.753788	1.75E-19	1.88E-16	glyceraldehyde-3-phosphate dehydrogenase C subunit 1
CUCPM_TC9512	-2.29767	0.00047	0.005219	-
CUCPM_TC9558	2.479247	2.14E-14	4.59E-12	Ribosomal protein L11 family protein
CUCPM_TC963	2.119717	2.59E-06	7.56E-05	arabinogalactan protein 9
CUCPM_TC9646	3.475537	1.07E-10	9.44E-09	Pathogenesis-related thaumatin superfamily protein
CUCPM_TC9818	4.47054	8.33E-06	0.000206	protodermal factor 1
CUCPM_TC9872	2.048868	1.41E-07	6.22E-06	unknown protein

**Table A.2 C.** List of differentially expressed genes identified in zucchini plants at 96 h following infestation with *Aphis gossypii*.

ID transcript	logFC	PValue	FDR	Description
CUCPM_L11078_T_2	4.345333	7.01E-07	0.000289	-
CUCPM_L11418_T_3	-2.23009	2.83E-05	0.003108	beta-galactosidase 3
CUCPM_L11538_T_1	2.210285	0.000144	0.008869	Auxin efflux carrier family protein
CUCPM_L11834_T_1	-3.44842	2.13E-05	0.00266	unknown protein
CUCPM_L119_T_2	-2.2477	2.85E-05	0.003112	-
CUCPM_L12331_T_1	2.576773	3.46E-07	0.000179	HSP2-like chaperones superfamily protein
CUCPM_L12482_T_1	-2.32332	0.000815	0.027479	Peroxidase superfamily protein
CUCPM_L1257_T_11	-2.23157	0.001139	0.033674	evolutionarily conserved C-terminal region 7
CUCPM_L12655_T_1	4.948106	0.001139	0.033674	-
CUCPM_L12932_T_2	-2.32959	6.43E-07	0.00028	ROP-interactive CRIB motif-containing protein 1
CUCPM_L13266_T_1	-2.04296	8.36E-05	0.006241	Leucine-rich repeat (LRR) family protein
CUCPM_L13274_T_2	-3.02593	1.02E-06	0.000357	-
CUCPM_L13384_T_2	-3.07003	0.000125	0.008207	HVA22-like protein F
CUCPM_L13518_T_1	3.171136	0.000672	0.024435	Putative cadmium/zinc-transporting ATPase HMA4
CUCPM_L13988_T_1	-2.2777	0.001921	0.047199	unknown protein
CUCPM_L14005_T_1	-6.13651	3.46E-06	0.000862	peptidoglycan-binding LysM domain-containing protein
CUCPM_L14108_T_1	-2.64001	0.001073	0.032854	Adenine nucleotide alpha hydrolases-like superfamily protein
CUCPM_L14306_T_1	-3.50622	1.10E-05	0.001714	-
CUCPM_L14332_T_1	2.547169	0.000442	0.018054	Pentatricopeptide repeat-containing protein
CUCPM_L14506_T_1	2.328205	0.001648	0.043153	-
CUCPM_L1464_T_4	3.09675	2.96E-05	0.003162	Probable LRR receptor-like serine/threonine-protein kinase
CUCPM_L1491_T_2	-3.04609	9.78E-05	0.00689	CONTAINS InterPro DOMAIN/s/x3a Mediator complex subunit Med28
CUCPM_L15348_T_1	-5.53093	0.000214	0.011217	UDP-Glycosyltransferase superfamily protein
CUCPM_L15699_T_1	5.697487	0.000257	0.012651	Enolase
CUCPM_L1594_T_1	5.0035	2.52E-08	2.82E-05	unknown protein
CUCPM_L16096_T_1	5.723279	0.000447	0.018115	Ribosomal protein S27a / Ubiquitin family protein
CUCPM_L16201_T_1	6.820458	1.25E-06	0.000417	Leucine-rich repeat (LRR) family protein
CUCPM_L16368_T_1	2.750291	0.000169	0.009795	-
CUCPM_L16501_T_1	5.481973	0.001627	0.04293	Zinc-binding ribosomal protein family protein
CUCPM_L16538_T_1	6.319498	9.33E-05	0.006749	Ribosomal L29e protein family
CUCPM_L16612_T_1	6.929242	0.000135	0.008506	-
CUCPM_L16634_T_1	7.069561	2.91E-05	0.003144	Cysteine proteinases superfamily protein
CUCPM_L16823_T_1	-5.45207	2.17E-05	0.002676	unknown protein
CUCPM_L16873_T_1	5.137678	0.000857	0.028265	-
CUCPM_L18098_T_1	2.594455	0.001707	0.043951	-
CUCPM_L18223_T_2	-2.18201	3.90E-05	0.00375	Domain of unknown function (DUF23)
CUCPM_L18268_T_4	-3.15904	0.001657	0.043237	glutathione peroxidase 4
CUCPM_L18529_T_14	-2.20594	7.99E-06	0.001461	unknown protein
CUCPM_L18935_T_3	2.026507	0.001781	0.045098	Thioesterase superfamily protein
CUCPM_L19031_T_1	-2.69805	0.000827	0.027735	FASCICLIN-like arabinogalactan-protein 11
CUCPM_L19052_T_1	-2.08319	0.000758	0.026383	nascent polypeptide-associated complex subunit alpha-like protein 2
CUCPM_L19090_T_1	-2.70824	8.55E-05	0.006314	CLAVATA3/ESR-RELATED 16
CUCPM_L19182_T_1	-2.50699	1.12E-05	0.001714	xylem cysteine peptidase 1
CUCPM_L19367_T_1	-2.0044	0.000183	0.01038	Peroxidase superfamily protein
CUCPM_L19409_T_2	-2.2433	0.001828	0.045857	Reticulon family protein
CUCPM_L19631_T_1	5.558934	0.00148	0.040344	Ribosomal L29 family protein

CUCPM_L19740_T_1	-2.00498	0.000635	0.023564	-
CUCPM_L20524_T_1	-4.23243	2.34E-05	0.002796	unknown protein
CUCPM_L20565_T_1	3.792392	1.92E-06	0.000582	HPT phosphotransmitter 4
CUCPM_L20750_T_1	2.611997	0.000273	0.013246	calmodulin-like 11
CUCPM_L21011_T_1	-2.52385	4.48E-06	0.000998	-
CUCPM_L21276_T_1	9.28702	3.15E-13	1.10E-09	-
CUCPM_L21318_T_1	-3.54545	0.001995	0.048335	-
CUCPM_L21581_T_1	4.917526	0.000704	0.025267	Protein of unknown function (DUF761)
CUCPM_L22337_T_1	-2.17728	0.000328	0.015074	glycine-rich protein 5
CUCPM_L22510_T_1	-2.45084	0.000411	0.017362	-
CUCPM_L22772_T_1	-2.67686	0.002054	0.04943	-
CUCPM_L22918_T_1	-6.88051	2.15E-08	2.73E-05	-
CUCPM_L23087_T_1	-5.94186	0.000132	0.008345	-
CUCPM_L23391_T_1	-5.52619	0.001411	0.039027	-
CUCPM_L23508_T_1	-2.1999	6.07E-05	0.005131	-
CUCPM_L23634_T_1	8.816086	6.49E-14	3.02E-10	-
CUCPM_L23672_T_1	-3.01133	0.000404	0.017174	RNA-binding (RRM/RBD/RNP motifs) family protein
CUCPM_L23846_T_1	7.008985	0.000757	0.026383	Lactate/malate dehydrogenase family protein
CUCPM_L2389_T_5	2.807571	1.84E-05	0.002402	alpha/beta-Hydrolases superfamily protein
CUCPM_L23916_T_1	5.546622	0.001309	0.037174	-
CUCPM_L24223_T_1	8.472946	6.21E-09	9.63E-06	-
CUCPM_L24296_T_1	3.318257	0.000315	0.014637	-
CUCPM_L24869_T_1	5.759859	1.65E-05	0.002187	-
CUCPM_L25625_T_1	-3.12146	8.59E-06	0.001537	-
CUCPM_L25839_T_1	-2.88266	8.36E-07	0.000315	-
CUCPM_L25896_T_1	-6.72988	3.65E-06	0.000893	terpene synthase 21
CUCPM_L26102_T_1	-3.32449	0.000325	0.014954	proline-rich protein 4
CUCPM_L26379_T_1	-4.03979	4.56E-06	0.000998	-
CUCPM_L26546_T_1	-6.12787	2.42E-06	0.000649	-
CUCPM_L26663_T_1	6.352955	3.84E-05	0.00375	-
CUCPM_L26841_T_1	-2.24104	1.50E-06	0.000487	-
CUCPM_L27181_T_1	-2.57216	0.000453	0.01831	-
CUCPM_L27223_T_1	-5.62911	0.001384	0.038797	-
CUCPM_L27250_T_1	-2.72861	5.14E-05	0.004525	-
CUCPM_L27288_T_1	-2.22806	4.99E-06	0.001054	-
CUCPM_L3263_T_1	2.090392	0.000311	0.014502	nicotianamine synthase 2
CUCPM_L3344_T_4	-2.92085	7.14E-07	0.000289	glycerol-3-phosphate acyltransferase 6
CUCPM_L3742_T_1	2.905638	0.0002	0.010896	RAD-like 1
CUCPM_L4093_T_6	5.356626	8.61E-10	2.00E-06	-
CUCPM_L4778_T_1	2.774569	0.000851	0.028265	-
CUCPM_L5536_T_1	2.000046	3.31E-05	0.003352	-
CUCPM_L6219_T_1	2.34411	0.001078	0.032854	Yippee family putative zinc-binding protein
CUCPM_L6509_T_2	2.471078	5.25E-05	0.004564	alpha/beta-Hydrolases superfamily protein
CUCPM_L652_T_4	-2.45961	2.63E-08	2.82E-05	-
CUCPM_L674_T_3	-2.4751	1.25E-06	0.000417	Microtubule-associated protein TORTIFOLIA1
CUCPM_L7034_T_4	-3.78144	0.00069	0.024955	unknown protein
CUCPM_L7191_T_3	-2.07306	6.18E-07	0.000278	S-norococlaurine synthase
CUCPM_L7410_T_1	5.792775	0.000587	0.022265	Zinc-binding ribosomal protein family protein
CUCPM_L750_T_2	-2.31531	5.48E-06	0.001142	cytochrome P45, family 77, subfamily B, polypeptide 1

CUCPM_L7720_T_1	3.03134	3.88E-06	0.000919	MLP-like protein 34
CUCPM_L7955_T_1	-2.23601	1.58E-05	0.002175	unknown protein
CUCPM_L7970_T_1	-6.71461	5.91E-07	0.000275	-
CUCPM_L8119_T_1	-6.54713	5.04E-05	0.00451	GATA transcription factor 12
CUCPM_L8326_T_1	-2.29226	0.000309	0.014492	Family of unknown function (DUF662)
CUCPM_L9238_T_1	4.076919	1.20E-05	0.001789	-
CUCPM_L9559_T_11	-2.84866	6.99E-06	0.001393	beta-D-xylosidase 4
CUCPM_L9625_T_3	2.442526	0.000127	0.00826	Cation efflux family protein
CUCPM_TC10255	-4.56644	0.000384	0.016802	Small nuclear ribonucleoprotein family protein
CUCPM_TC10326	-2.06856	0.000209	0.011107	beta-6 tubulin
CUCPM_TC10389	2.554878	0.001563	0.041678	-
CUCPM_TC10408	-2.87597	3.88E-05	0.00375	polygalacturonase 2
CUCPM_TC10605	-2.11668	0.001988	0.048256	-
CUCPM_TC10630	-2.7754	0.000463	0.018628	-
CUCPM_TC10884	2.385041	1.14E-05	0.001714	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
CUCPM_TC11388	-2.09256	2.97E-05	0.003162	Family of unknown function (DUF566)
CUCPM_TC11437	-2.21878	6.29E-05	0.005255	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
CUCPM_TC12077	-2.79472	0.000136	0.00854	unknown protein
CUCPM_TC12138	-2.58052	2.28E-06	0.000629	L-ascorbate oxidase homolog
CUCPM_TC12186	-2.79631	0.00012	0.00806	Thioredoxin-like protein
CUCPM_TC12285	-2.03309	0.000542	0.021016	MLP-like protein 423
CUCPM_TC12321	2.544886	0.000234	0.011858	EF hand calcium-binding protein family
CUCPM_TC12735	-5.90293	3.05E-14	2.13E-10	Putative membrane lipoprotein
CUCPM_TC12886	2.217019	5.71E-09	9.63E-06	-
CUCPM_TC12961	-2.12188	0.001323	0.03746	Cytochrome b561/ferric reductase transmembrane protein family
CUCPM_TC13516	-2.28126	2.56E-05	0.002887	Acid phosphatase 1
CUCPM_TC137	2.09688	1.06E-09	2.12E-06	thiazole biosynthetic enzyme, chloroplast (ARA6) (THI1) (THI4)
CUCPM_TC13842	-2.40684	6.59E-05	0.005359	Isoflavone reductase-like protein
CUCPM_TC14673	-3.69701	0.000353	0.015742	Acyl-CoA N-acyltransferases (NAT) superfamily protein
CUCPM_TC14893	-2.96709	2.57E-05	0.002887	laccase 17
CUCPM_TC14975	-2.13988	1.14E-05	0.001714	unknown protein
CUCPM_TC15507	-2.21906	0.000244	0.012262	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
CUCPM_TC15559	-2.88233	0.000205	0.011064	-
CUCPM_TC16374	-2.26569	0.000152	0.00913	plasmodesmata callose-binding protein 3
CUCPM_TC16500	-2.07776	3.20E-07	0.000172	unknown protein
CUCPM_TC16752	-2.36603	0.00072	0.025692	BCL-2-associated athanogene 1
CUCPM_TC17062	7.216548	9.73E-06	0.001635	-
CUCPM_TC17109	6.445693	0.000158	0.009359	Cysteine proteinases superfamily protein
CUCPM_TC17141	7.697226	8.21E-07	0.000315	-
CUCPM_TC17148	-5.33434	0.001565	0.041678	magnesium (Mg) transporter 1
CUCPM_TC17159	2.575905	0.001891	0.046865	Aldo-keto reductase family 4 member C9
CUCPM_TC17500	-2.51659	0.000284	0.013684	Probable receptor-like serine/threonine-protein kinase
CUCPM_TC1752	2.140072	2.23E-07	0.000141	Leucine-rich repeat family protein
CUCPM_TC18025	-3.31643	5.99E-06	0.001211	Transcription factor bHLH61
CUCPM_TC18166	-2.47999	2.36E-05	0.002796	Beta-glucosidase
CUCPM_TC18221	5.181084	3.18E-06	0.000823	CONTAINS InterPro DOMAIN/slx3a Vacuolar import
CUCPM_TC18484	-2.33663	0.000438	0.017959	Plant L-ascorbate oxidase
CUCPM_TC18545	-3.1796	3.73E-06	0.000898	Pollen Ole e 1 allergen and extensin family protein
CUCPM_TC1940	3.571647	1.97E-05	0.002519	pseudo-response regulator 7

CUCPM_TC19413	-2.01824	8.06E-06	0.001461	Receptor-like serine/threonine-protein kinase
CUCPM_TC20082	6.84393	1.30E-05	0.001868	-
CUCPM_TC20292	-2.32935	0.000783	0.026905	Blue copper protein
CUCPM_TC20529	2.272154	0.000535	0.020862	Calcium-binding EF-hand family protein
CUCPM_TC20592	-2.60493	4.70E-05	0.004345	FASCICLIN-like arabinogalactan 1
CUCPM_TC20607	-2.32611	0.000608	0.022862	Zinc finger (CCCH-type) family protein / RNA recognition motif (RRM)-containing protein
CUCPM_TC20622	-2.07799	0.000162	0.009464	Aspartic proteinase nepenthesin-2
CUCPM_TC21230	2.268644	1.13E-05	0.001714	unknown protein
CUCPM_TC21296	-2.67774	9.83E-05	0.00689	tracheary element differentiation-related 7
CUCPM_TC21545	-2.12364	2.47E-05	0.002844	novel plant snare 11
CUCPM_TC21618	-2.49399	4.58E-06	0.000998	Eukaryotic aspartyl protease family protein
CUCPM_TC22045	-3.94137	1.98E-07	0.000138	-
CUCPM_TC22105	-2.09145	0.000981	0.030777	ascorbate peroxidase 3
CUCPM_TC22161	-2.49682	0.000307	0.014458	Abscisic acid-responsive (TB2/DP1, HVA22) family protein
CUCPM_TC22434	-2.44799	0.001238	0.03575	Proline-rich receptor-like protein kinase
CUCPM_TC22482	-2.86005	7.92E-06	0.001461	ribonuclease 3
CUCPM_TC2458	-2.37026	0.000101	0.00705	Cobalamin-independent synthase family protein
CUCPM_TC2819	2.600084	4.19E-05	0.003956	global transcription factor group E7
CUCPM_TC2897	-3.00478	2.51E-07	0.000152	xylem cysteine peptidase 2
CUCPM_TC294	-3.38494	0.000619	0.023152	NAD(P)-binding Rossmann-fold superfamily protein
CUCPM_TC420	2.721851	7.24E-07	0.000289	Polyketide cyclase/dehydrase and lipid transport superfamily protein
CUCPM_TC443	3.047414	0.000197	0.010867	actin depolymerizing factor 1
CUCPM_TC4668	-2.47138	8.72E-05	0.006408	glycerol-3-phosphate acyltransferase 6
CUCPM_TC5031	-4.97765	7.89E-09	1.10E-05	gamma tonoplast intrinsic protein
CUCPM_TC606	-2.37735	2.82E-06	0.000743	glycosyl hydrolase 9C2
CUCPM_TC6210	-2.98358	1.24E-07	9.10E-05	Plant-specific GATA-type zinc finger transcription factor family protein
CUCPM_TC6255	-2.4888	0.001819	0.045736	Protein of unknown function, DUF547
CUCPM_TC7	-2.13967	2.36E-05	0.002796	ribulose biphosphate carboxylase small chain 1A
CUCPM_TC7172	-2.86669	5.60E-06	0.001149	Basic-leucine zipper (bZIP) transcription factor family protein
CUCPM_TC759	-2.76661	4.15E-07	0.000207	NAD-dependent glycerol-3-phosphate dehydrogenase family protein
CUCPM_TC768	-2.21126	9.21E-06	0.001587	unknown protein
CUCPM_TC83	-2.04315	8.83E-06	0.001541	Protein kinase superfamily protein
CUCPM_TC9010	-2.3099	2.41E-05	0.00282	serine carboxypeptidase-like 48
CUCPM_TC9026	-3.38575	0.001441	0.039757	Heavy metal-associated isoprenylated plant protein 26
CUCPM_TC9818	-4.51508	6.53E-05	0.005359	protodermal factor 1



**STAGE**

Five months stage (01/04/2015 - 30/09/2015) at CREA Council for Agricultural Research and Economics-Research Centre for Vegetable and Ornamental Crops, Pontecagnano Faiano (SA), Italy; group of Dr Nunzio D'Agostino to perform bioinformatics analyses.

Five months stage (10/02/2017 – 17/07/2017) Rothamsted Research, Harpenden, Herts, UK, Chemical Ecology Group, Department of Biological Chemistry and Crop Protection, in collaboration with Prof Toby J. A. Bruce to perform collection and identification of volatile compounds.



**PUBLICATIONS**

Submitted paper: Mariangela Coppola, Elena Manco, Alessia Vitiello, Massimo Giorgini, Rosa Rao, Francesco Pennacchio, Maria Cristina Digilio (2017). Plant response to aphid attack promotes their dispersal behavior. *Entomologia experimentalis et applicata*

Paper accepted for publication: Mariangela Coppola, Pasquale Cascone, Valentina Madonna, Ilaria Di Lelio, Francesco Esposito, Concetta Avitabile, Alessandra Romanelli, Emilio Guerrieri, Alessia Vitiello, Francesco Pennacchio, Rosa Rao, Giandomenico Corrado (2017). Plant-to-plant communication triggered by systemin primes anti-herbivore resistance in tomato. *Scientific reports*

Paper: Filomena Grasso, Mariangela Coppola, Fabrizio Carbone, Luciana Baldoni, Fiammetta Alagna, Gaetano Perrotta, Antonio J. Pe rez-Pulido, Antonio Garonna, Paolo Facella, Loretta Daddiego, Loredana Lopez, Alessia Vitiello, Rosa Rao, Giandomenico Corrado (2017). The transcriptional response to the olive fruit fly (*Bactrocera oleae*) reveals extended differences between tolerant and susceptible olive (*Olea europaea* L.) varieties. *PloS one*, 12(8), e0183050

Poster communication: Alessia Vitiello, Donata Molisso, Nunzio D'Agostino, Maria Cristina Digilio, Francesco Pennacchio, Giandomenico Corrado Rosa Rao (2016). Transcriptional reprogramming of zucchini plant during aphid infestation. Proceedings of the LX SIGA Annual Congress Catania, Italy – 13/16 September, 2016 ISBN 978-88-904570-6-7

Oral communication: Alessia Vitiello, Daria Scarano, Nunzio D'Agostino, Maria Cristina Digilio, Francesco Pennacchio, Giandomenico Corrado Rosa Rao (2015). Unraveling zucchini transcriptome response to aphids. BBCC2015 Dec.4th,2015 CNR-ISA, Avellino, Italy. *Peer journal* preprint, <https://doi.org/10.7287/peerj.preprints.1635v1>

Poster communication: Alessia Vitiello, Daria Scarano, Nunzio D'Agostino, Maria Cristina Digilio, Francesco Pennacchio, Giandomenico Corrado Rosa Rao (2015). *De novo* transcriptome assembly in zucchini: a high-quality reference for digital gene expression profiling aiming at the identification of key genes involved in response to aphids. Proceedings of the Joint Congress SIBV-SIGA Milano, Italy – 8/11 September, 2015 ISBN 978-88-904570-5-0

Proceedings of the Joint Congress SIBV-SIGA  
 Milano, Italy – 8/11 September, 2015  
 ISBN 978-88-904570-5-0

Poster Communication Abstract – B.14

**DE NOVO TRANSCRIPTOME ASSEMBLY IN ZUCCHINI: A HIGH-QUALITY REFERENCE FOR DIGITAL GENE EXPRESSION PROFILING AIMING AT THE IDENTIFICATION OF KEY GENES INVOLVED IN RESPONSE TO APHIDS**

VITIELLO A. \*, SCARANO D. \*, D'AGOSTINO N. \*\*, DIGILIO M.C. \*, PENNACCHIO F. \*, CORRADO G. \*, RAO R. \*

\*) Department of Agriculture, University of Naples "Federico II", Via Università 100, 80055 Portici (Italy)

\*\*) Centro per la ricerca in agricoltura e l'analisi dell'economia agraria, Centro di Ricerca per l'Orticoltura (CRA-ORT), Via Cavallegeri 25, 84098 Pontecagnano (Italy)

*Cucurbita pepo*, RNA-seq, *Aphis gossypii*, de novo transcriptome assembly, zucchini protection

*Cucurbita pepo* belongs to the Cucurbitaceae family, the second- most large horticultural family in terms of economic importance after Solanaceae. One major issue related to zucchini cultivation, both in greenhouse and open-field, is the damages imposed by aphids such as *Aphis gossypii* (Homoptera: Aphididae). This insect represents one of the major pests of cucurbits and is a well known vector of several viruses. In this study we aim at the identification of candidate genes involved in zucchini plant response to the aphid *Aphis gossypii*. As plant material, we selected the zucchini cultivar "San Pasquale", which is extensively cultivated in the Campania region and has showed to be highly susceptible to aphid attack.

In order to monitor the effect of zucchini-aphid interaction at transcriptome level, we designed a time course experiment. Zucchini plants were grown in controlled conditions in presence or absence of *Aphis gossypii*, and leaf material was collected at three different time points (24, 48, 96 h) after infestation. RNA was extracted from individual plants and sequenced using the Illumina HiSeq 2500 platform. The sequencing generated ~34 billion of paired-end reads of 100 nucleotides in length per sample. Raw reads were pre-processed to remove sequences with low quality bases and adapter contaminations. High quality reads were *de novo* assembled into 71,648 transcripts by Velvet/Oases and CAP3 tools with an average size of 1064 nucleotides. BLAST similarity-based searches were performed against i) *Cucumis melo* protein complement (v 3.5.1); ii) *Arabidopsis thaliana* (v 10) proteins; iii) the publically available *C. pepo* transcriptome and the UniProtKb/SwissProt database. Only 821 transcripts, corresponding to 1.14% of the assembled transcriptome, have no correspondence in the four databases queried. Functional annotation of sequence data will be corroborated by protein domain identification and Gene Ontology classification.

The assembled transcriptome will be used as reference for read mapping and digital gene expression profiling analysis. The identification of differentially expressed genes will be useful for understanding the molecular response of infested plants. This knowledge represents a fundamental prerequisite for the development of novel tools for the control *A. gossypii*.

*This work was carried out in the frame of the "GenHORT - OR1 "Qualità e sostenibilità delle produzioni mediante strumenti di genomica strutturale e funzionale" project (PON. 02 00395 3215002).*

## Unraveling zucchini transcriptome response to aphids

Vitiello Alessia\*, Scarano Daria\*, D'Agostino Nunzio\*\*, Digilio Maria Cristina\*, Pennacchio Francesco\*, Corrado Giandomenico\*, Rao Rosa\*

\*) Department of Agriculture, University of Naples Federico II, Via Università 100, 80055 Portici (Italy)

\*\*\*) Centro per la ricerca in agricoltura e l'analisi dell'economia agraria, Centro di ricerca per l'orticoltura (CREA-ORT), Via Cavallotti 25, 84098 Pontecagnano (Italy)

[alessia.vitiello@unina.it](mailto:alessia.vitiello@unina.it); [rao@unina.it](mailto:rao@unina.it)

*Cucurbita pepo* belongs to the Cucurbitaceae, the second-most large horticultural family of economic importance after Solanaceae. One major issue related to zucchini cultivation is the damage caused by aphids such as *Aphis gossypii* (Homoptera: Aphididae). The aim of this study is the identification of candidate genes involved in zucchini plant response to *A. gossypii*. In order to monitor the effect of zucchini-aphid interaction at transcriptomic level, zucchini plants (cv "San Pasquale") were grown in controlled conditions in presence or absence of *A. gossypii*. Leaf material was collected at 24, 48 and 96 hours after aphid infestation. RNA extracted was sequenced using the Illumina HiSeq 2500 platform. The sequencing generated ~34 million of paired-end reads of 100 nucleotides in length per sample. High quality reads were *de novo* assembled into 71,648 transcripts. About 94% of the assembled transcripts contain coding sequences that could be translated into proteins. Over 60% of the transcripts were functionally annotated and assigned to one or more InterPro domains and Gene Ontology terms. A subset of 42,517 sequences of the *C. pepo* transcriptome was used for read mapping and differentially expressed genes (DEG) identification. Largest number of DEG were observed after 48 hours from aphid infestation. The transcriptome represents a high-quality reference for read alignment and DEG call. The understanding of the molecular response of infested plants will be essential to develop new tools for *A. gossypii* control.

Proceedings of the LX SIGA Annual Congress  
 Catania, Italy – 13/16 September, 2016  
 ISBN 978-88-904570-6-7

Poster Communication Abstract – 1.37

## TRANSCRIPTIONAL REPROGRAMMING OF ZUCCHINI PLANT DURING APHID INFESTATION

VITIELLO A.\*, MOLISSO D.\*, D'AGOSTINO N.\*\*\*, DIGILIO M.C.\*, PENNACCHIO F.\*,  
 CORRADO G.\*, RAO R.\*

\*) Department of Agricultural Sciences, University of Naples Federico II, via Università 100, 80055 Portici (Italy)

\*\*) Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Centro di Ricerca per l'Orticoltura (CREA-ORT), Via dei Cavallegeri 25, 84098 Pontecagnano Faiano (Italy)

*Cucurbita pepo*, RNA-seq, *Aphis gossypii*, zucchini protection, digital gene expression profiling

*Cucurbita pepo* is an economically important species within the genus *Cucurbita*. Zucchini is widely cultivated in temperate regions and it is one of the most variable species in terms of fruit shape. In addition to immature fruits, also flowers and seeds are consumed and are considered an important sources of nutrients. The melon aphid *Aphis gossypii* (Homoptera: Aphididae) is a serious pest of zucchini. Its feeding behaviour causes leaf curling and chlorosis, hindering plant photosynthetic capacity. *A. gossypii* is also effectively able to transmit several plant viruses, resulting in a significant yield loss. The aim of this study is to investigate, through a time-course transcriptomic analysis based on RNA-seq, the transcriptional response of zucchini plants (cv. "San Pasquale") during *Aphis gossypii* infestation. Zucchini plants were grown in controlled conditions until the 3<sup>rd</sup> leaf stage. Ten adult aphids were transferred on the 1<sup>st</sup> and 2<sup>nd</sup> leaves, which were collected after 24, 48 and 96 hours post infestation (pi). The same leaves were collected from uninfested plants and used as controls. Total RNA was extracted from each sample and sequenced using the Illumina HiSeq 2500 platform. The sequencing resulted in ~34 million of paired-end reads per sample. After quality assessment and processing, high quality reads were *de novo* assembled into 71,648 transcripts with an average length of 1331 nucleotides. About 94% of the assembled transcripts contains coding sequences that could be translated into proteins. Over 70% of the transcripts was functionally annotated using Blast2GO. It was possible to assign one or more Gene Ontology (GO) terms to 51,398 transcripts. The transcriptome was used as high quality reference for read alignment. To ensure that each locus was represented only once in the dataset, we filtered out the longest transcript for each gene locus. Then, we employed this dataset, which includes 42,517 sequences, for read mapping and differentially expressed genes (DEG) identification. The filtering criteria used for DEG call were: a logFold change in expression greater than 2 and a FDR < 0.05. Considering the three time points, 766 transcripts were differentially expressed. After 24 hours pi, 158 transcripts (115 up and 45 down) were influenced by aphid infestation. The number of affected transcripts increased to 565 after 48 hours (420 up and 145 down), and declined to 179 transcripts (62 up and 117 down) after 96 hours from the infestation. Major categories involved in zucchini response are "photosynthesis", "response to stress", "primary metabolism" and "protein metabolism". Transcripts upregulated at 24 hours pi include serine/threonine protein kinases, involved in signalling response, 2-alkenal reductase proteins, involved in detoxification mechanism of toxic aldehydes and cysteine proteinases, involved in plant-insect pest interaction. In-depth analysis of transcriptional reprogramming will help elucidating, for the first time, the zucchini response to aphid infestations.

This work was carried out in the frame of the "GenHORT - OR1 "Qualità e sostenibilità delle produzioni mediante strumenti di genomica strutturale e funzionale" project (PON. 02 00395 3215002).


## RESEARCH ARTICLE

# The transcriptional response to the olive fruit fly (*Bactrocera oleae*) reveals extended differences between tolerant and susceptible olive (*Olea europaea* L.) varieties

Filomena Grasso<sup>1</sup>, Mariangela Coppola<sup>1</sup>, Fabrizio Carbone<sup>2</sup>, Luciana Baldoni<sup>3</sup>, Fiammetta Alagna<sup>3</sup>, Gaetano Perrotta<sup>4</sup>, Antonio J. Pérez-Pulido<sup>5</sup>, Antonio Garonna<sup>1</sup>, Paolo Facella<sup>4</sup>, Loretta Daddiego<sup>4</sup>, Loredana Lopez<sup>4</sup>, Alessia Vitello<sup>1</sup>, Rosa Rao<sup>1\*</sup>, Giandomenico Corrado<sup>1\*</sup>

**1** Dipartimento di Agraria, Università degli Studi di Napoli "Federico II", Portici (NA), Italy, **2** Centro di Ricerca per l'Olivicoltura e l'Industria Olearia, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), Rende (CS), Italy, **3** Institute of Biosciences and Bioresources (IBBR), CNR, Perugia, Italy, **4** Trisaia Research Center, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Rotondella (MT), Italy, **5** Departamento Biología Molecular e Ingeniería Bioquímica, Universidad Pablo de Olavide, Sevilla, Spain

\* [rao@unina.it](mailto:rao@unina.it) (RR); [giandomenico.corrado@unina.it](mailto:giandomenico.corrado@unina.it) (CG)


 OPEN ACCESS

**Citation:** Grasso F, Coppola M, Carbone F, Baldoni L, Alagna F, Perrotta G, et al. (2017) The transcriptional response to the olive fruit fly (*Bactrocera oleae*) reveals extended differences between tolerant and susceptible olive (*Olea europaea* L.) varieties. PLoS ONE 12(8): e0183050. <https://doi.org/10.1371/journal.pone.0183050>

**Editor:** John Vontas, University of Crete, GREECE

**Received:** May 16, 2017

**Accepted:** July 30, 2017

**Published:** August 10, 2017

**Copyright:** © 2017 Grasso et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The reads were archived in the NCBI Sequence Read Archive (SRA) database (Acc. Num. SRP074450). The microarray design was deposited to the NCBI public repository Gene Expression Omnibus (GEO) (Acc. Num. GPL21855). Microarray experiments are archived at the GEO data repository (Acc. Num. GSE81296).

**Funding:** This work was partially supported by the "Progetto Olea - Genomica e miglioramento genetico dell'olivo", Ministero per le Politiche

## Abstract

The olive fruit fly *Bactrocera oleae* (Diptera: Tephritidae) is the most devastating pest of cultivated olive (*Olea europaea* L.). Intraspecific variation in plant resistance to *B. oleae* has been described only at phenotypic level. In this work, we used a transcriptomic approach to study the molecular response to the olive fruit fly in two olive cultivars with contrasting level of susceptibility. Using next-generation pyrosequencing, we first generated a catalogue of more than 80,000 sequences expressed in drupes from approximately 700k reads. The assembled sequences were used to develop a microarray layout with over 60,000 olive-specific probes. The differential gene expression analysis between infested (i.e. with II or III instar larvae) and control drupes indicated a significant intraspecific variation between the more tolerant and susceptible cultivar. Around 2500 genes were differentially regulated in infested drupes of the tolerant variety. The GO annotation of the differentially expressed genes implies that the inducible resistance to the olive fruit fly involves a number of biological functions, cellular processes and metabolic pathways, including those with a known role in defence, oxidative stress responses, cellular structure, hormone signalling, and primary and secondary metabolism. The difference in the induced transcriptional changes between the cultivars suggests a strong genetic role in the olive inducible defence, which can ultimately lead to the discovery of factors associated with a higher level of tolerance to *B. oleae*.

Agricole Alimentarie Forestali (DM. 27011/7643/10).

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

The olive fruit fly, *Bactrocera oleae* (Diptera: Tephritidae), is arguably the single largest threat of cultivated olive (*Olea europaea* L.). Since olive domestication, *B. oleae* is the most devastating pest in the Mediterranean basin and the recent invasion of California suggests that the fruit fly follows its host [1, 2]. Phytophagous larvae feed exclusively on olive fruits [3]. Depending on climate conditions, the fruit fly can produce severe crop damage and significant economic loss in the whole olive sector [4].

Adult females pierce the olive and lay eggs under the exocarp. Hatched larvae progressively consume the olive pulp, causing tissue loss, premature fruit drop and reduction of oil yield. Moreover, fly infestation increases olive oil acidity and peroxide value, as well as musty and earthy off-flavours, extensively reducing oil quality (e.g. downgrading extra virgin olive oil to less valuable categories). Indirect effects are mainly due the presence of necrotic areas and microorganisms in feeding tunnels [5–9]. The conventional *B. oleae* management relies on chemical insecticides and/or traps [4]. The olive fly, as many other pests, can acquire resistance to pesticides [10], increasing the need for more effective biological or integrated control methods [11–13].

Considering the unrivalled olive oil-health benefits, research on the olive-fruit fly interaction has the long-term potential to influence not only the olive oil production but also the health-promoting properties of the olive oil. The economic impact of the *B. oleae* does not match the research efforts aimed to examine the interaction between the olive and its key enemy. For instance, compared to other biotic stresses, very few studies shed light on the mechanisms underlying olive defence and resistance at molecular level [14, 15].

Olive varieties can have different levels of tolerance to the olive fruit fly [4, 16]. In a comparison of different cultivars, the percentage of infestation ranged from less than 10% to up to 31% [16]. This difference is significant because a percentage of exit holes lower than 10% is compatible with a production of high quality olive oil in absence of chemical control methods, if adequate harvesting and storing procedures are followed [7]. A different tolerance to the olive fruit fly is evident not only comparing cultivars of different origins but also analysing regional germplasm [16, 17]. For instance, among the twenty varieties that constitute the olive germplasm of the Campania region (Southern Italy) [18], the ‘Ortice’ and ‘Ruveia’ are reported to be highly susceptible and tolerant, respectively [19].

The basis of the different tolerance to the olive fruit fly is expected to be complex [20]. It may rely on mechanical barriers (e.g. aliphatic waxes), chemical factors (e.g. oleuropein, cyanidine), morphological characteristics (e.g. fruit size) and their combination. Similarly, the relative prominence and the contribution of these factors are yet to be fully clarified [21–25]. Moreover, the mechanism underlying the different levels of tolerance to the *B. oleae* has never been studied at molecular level. Finally, it is not known whether those features are constitutively expressed or induced by the fruit fly feeding.

To address these points, we studied the molecular response of the drupe in two olive cultivars with different levels of tolerance to the fruit fly. To this aim, we first generated a catalogue of more than 80,000 unigenes by next-generation pyrosequencing and then developed a microarray layout as affordable tool for functional genomics in olive. Our study provides insights into the molecular reaction of the drupe to larva feeding and illustrates the complexity and the differences of the drupe defence response of varieties with different levels of fruit fly susceptibility.



## Material and methods

### Plant material

Two olive varieties were studied because of their contrasting tolerance to *B. oleae*, 'Ortice' (susceptible to the olive fruit fly) and 'Ruveia' (tolerant) [19]. Olives were harvested from field plants at the Azienda Agricola Regionale Sperimentale "Improsta" (Eboli, Salerno) when at the Jaen Ripening Index (JRI) 2 [26]. The percentage of olive fly attack was calculated on 300 drupes per variety (10 replicated groups of 30 randomly collected drupes). Statistical differences were evaluated with a Student *t*-test. For molecular studies, olives with visible symptoms of pathogen attack as well as those with fly exit holes were discarded. Olives at different attack stages (punctured, with II or III instar larva, or pupae) were pooled and used with undamaged olives for massive parallel sequencing, in order to create a comprehensive repertoire of genes modulated by *B. oleae*. For the study of differential gene expression by microarray analysis, we used for each cultivar drupes with feeding II or III instar larvae and undamaged controls. For massive sequencing and for microarray analysis, olives were sliced under a light microscope to remove the larva. At this stage, drupes with sterile stings and dead/inactive larvae were also discarded. Slices were frozen in liquid nitrogen and stored at -80°C until use.

### 454 sequencing

Total RNA was isolated from 200 mg of drupe with the RNeasy Plant Mini Kit (Qiagen Hilden, Germany) and treated with DNase I (Ambion, Austin, TX, USA). Three replicates, each consisting of five olives, were used for each biological condition (punctured olives, olives with feeding II or III instar larva, olives with pupae and undamaged olives). cDNAs were synthesized using the SMART PCR cDNA Synthesis kit (Clontech, Palo Alto, CA, USA). First strand synthesis was performed using eight µg of total RNA as described [22]. Double stranded cDNAs were purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and quantified with a fluorimeter (Victor 2, Perkin Elmer, Wellesley, MA, USA). To estimate cDNA quality and fragment length, samples were separated on a 1.5% agarose gel. cDNA from olives characterized by different stages of fly attack were pooled together to obtain a representative 'damaged' sample. Two cDNA libraries (damaged and undamaged olives) per cultivar ('Ortice' and 'Ruveia') were sequenced. Approximately five µg of cDNA were sheared into small fragments via nebulization. The four shotgun cDNA libraries were sequenced using a 454 GS FLX+ Titanium Sequencer (Roche Diagnostics Corporation, Basel, Switzerland). The adaptor-trimmed 454 reads were assembled using the GSAssembler Software (Roche Diagnostics Corporation, Basel, Switzerland). To annotate unigenes (contigs and singletons), a blastX-based similarity search ( $e\text{-value} \leq 1e\text{-5}$ ) was used, querying the NCBI non-redundant (nr) database. We mapped the GI identifiers (<http://www.ncbi.nlm.nih.gov/>) of the best blastX hits to the UniprotKB protein database (<http://www.uniprot.org/>) in order to extract Gene Ontology (GO, <http://www.geneontology.org/>) and KEGG orthology (KO, <http://www.genome.jp/kegg/>) terms.

### Microarray preparation and hybridization

CustomArray™ (CustomArray Inc., Bothell, WA, USA) microarrays containing over 90k probes were built at the ENEA—Trisaia Research Center (Rotondella, MT, Italy). Probes were designed using the ProbeWeaver software (CustomArray Inc., Bothell, WA, USA), based on the 454 pyrosequencing results of four cDNA libraries obtained from damaged and undamaged drupes of the 'Ortice' and 'Ruveia' varieties.

The chip layout consists of 61,825 olive probes (60,706 non-redundant) out of the 87,720 sequences of the pooled library. The layout includes around 20,000 quality control spots. Probes were 35- to 40-mers with a melting temperature of 70–75°C and were synthesized on microarrays through the CustomArray Synthesizer™ (CustomArray Inc., Bothell, WA, USA).

Total RNA was isolated from 200 mg of finely ground olive tissue. A purification step was performed in a 1:1:2 (v/v/v) solution of phenol, chloroform and RNA Extraction Buffer (1 M Tris-HCl pH 8.5, 5 M NaCl, 0.5 M EDTA pH 8.0, 10% SDS) twice, followed by a chloroform purification. RNA isolation from the aqueous phase and quality control were performed as described [28]. Samples with a 260/280 nm absorbance ratio higher than 1.8 and a 260/230 nm absorbance ratio higher than 2.0 were used for subsequent experiments. For each biological condition, three independent samples (i.e. from different trees) were obtained as pools of three to five independently extracted technical replicates. Two µg of total RNA were retrotranscribed using the RNA ampULSe Amplification and Labeling Kit for CustomArray™ microarrays (Kreatech Biotechnology). cRNA labeling was carried out with the Cy5-ULS (Cyanine-Universal Linkage System). Unincorporated dye was removed with a KREApure purification column (Agilent). Labeling yield and quality were assessed using a NanoDrop 1000 (Thermo Scientific). cRNAs were fragmented for 20 min at 95°C in 200 mM Tris-Acetate (pH 8.1), 500 mM potassium acetate and 150 mM molybdenum acetate. The fragmented Cy5-cRNAs were added to the hybridization solution (6X SSPE, 0.05% Tween-20, 20mM EDTA, 25% formamide, 100 ng/µl salmon sperm DNA, 0.04% SDS) and poured in the hybridization chambers of the pre-hybridized microarrays, following the CustomArray™ Hybridization protocol. After an overnight incubation at 45°C in a rotating rotisserie oven, microarrays were treated with six washing steps: firstly, in 6X SSPE and 0.05% Tween-20 for 5' at 45°C, then with 3X and 0.5X SSPE and 0.05% Tween-20 for 1' at room temperature, with 2X PBS and 0.1% Tween-20, and finally twice with 2X PBS for 1' at room temperature.

For the validation of the microarray results, Real time PCR was carried out as described [14]. Primers and their main features are reported in [S1 Table](#).

### Microarray data analyses

Microarray slides were scanned with the GenePix Pro microarray scanner (Axon Instruments) and data processed with the CustomArray™ Microarray Imager software (CustomArray Inc., Bothell, WA, USA). A maximum threshold of 0.20 for the coefficient of variance (of the spots corresponding to identical probes) was applied to control intra-chip hybridization variability. The correlation among the three technical replicates for each experimental condition was assured by a minimum Pearson coefficient (R) of 0.99. Stripping and preparation of the microarrays for re-hybridization was performed twice, considering the three technical replicates, according to the manufacturer's instruction (CustomArray Inc., Bothell, WA, USA). Raw values were normalized on the median of the intensities by the ProbeWeaver software (CustomArray Inc., Bothell, WA, USA) using the quantile algorithm. Pairwise analysis of differential expression was assessed with a Welch's t-test, followed by a Bonferroni-Hochberg FDR adjustment with a cut off p-value of 0.05, using R [29]. To filter out weakly expressed sequences, we calculated the average and standard deviation of the expression value of all empty and negative controls and set as threshold the mean value plus two times the standard deviation [30]. Of the filtered, significantly differentially expressed probes, only those with greater than 2-fold increase or 2-fold decrease in expression compared to the control condition were used for further analysis. Principal Component Analysis was performed in R [29]. K-means clustering analysis of the differentially expressed genes (DEGs) was carried out with the Multi Experiment Viewer (MeV) software [31]. Dataset grouping was performed with the average linkage

clustering methods based on Euclidean distances. Different number of clusters ( $k$  fixed from 8 to 16) were compared and  $k = 10$  was selected to maximize partitioning and to avoid empty or very small groups. The corresponding expression graphs were visualized with MeV. We followed the Minimum Information About a Microarray Experiment (MIAME) guidelines for microarray analysis and verification [32].

### Functional annotation of differentially expressed genes

Functional annotation was carried out by sequence analysis using the Blast2GO software [33]. Briefly, for each sequence corresponding to a differentially expressed probe, a blastX similarity search ( $e$ -value  $< 1e-6$ ) against the non-redundant (nr) NCBI database (Non-redundant GenBank CDS translations including RefSeq, PDB, SwissProt, IR and PRF) was performed to retrieve a maximum of 20 top homologous hits per query. The GO-term mapping and annotation were retrieved using NCBI as well as non-redundant reference protein database (PSD, UniProt, RefSeq, GenPept, PDB Full Gene Ontology DB). Sequences with a blast hit that could not be mapped and annotated were then blasted (blastX) against the *Arabidopsis thaliana* protein sequences and the *Oryza sativa* protein sequences database. Additional annotations (e.g. the recovery of implicit "Biological Process" and "Cellular Component" GO-terms from "Molecular Function" annotations) were implemented using ANNEX 5.0 [34]. Completion of the functional annotation with protein domain information was performed with InterProScan 5.0. A plant GO-Slim reduction was carried out to summarize the functional content of the dataset. To complete the functional annotation Sma3s was also used [35]. Plant taxonomic division from UniProt was selected as source of annotations, and the transcriptomic sequences were enriched with keywords, InterPro domains and motifs, and GO terms. We generated multi-pie chart summary to present meaningful GO identifiers that are at hand yet not excessively general. When possible, sequences that could not be mapped were named (but not annotated) after a tblastX search against the nr NCBI database to identify sequences similar to the query based on their coding potential.

### Results

The already reported different susceptibility to *B. oleae* of the two varieties was evaluated by measuring the percentage of attack at the time of harvesting. The susceptible variety 'Ortice' had an attack index 2.5 times higher than the 'Ruveia' ( $p < 0.01$ ). At sample harvest, the infestation level was 13.3% in 'Ortice' and 5.3% in 'Ruveia'.

Pyrosequencing of the four cDNA libraries from damaged (pool of different stages of fly attack) and control undamaged fruits of 'Ortice' and 'Ruveia' generated 695,511 reads, with an average length of 361 bp (N50: 421 bp; N75: 359 bp) (Table 1).

We assembled the raw reads from the four libraries together. More than 80% of raw sequences were included in the assembly and 72,662 remained as singleton (Table 2).

The assembly yielded 15,058 contigs, with an average length of 884 bp and 78% of the sequences longer than 500 bp (Panel A in S1 Fig). Approximately 40% of the contigs were composed of 2–10 reads. The majority of the contigs (28%) were low-coverage (6–10 reads) and around 5% were high-coverage ( $> 100$  reads). The unigenes list and their annotation is reported in S2 Table. An overview of the putative functions of the unigenes, based on a blastX-similarity search annotation, is presented in S2 Fig. About 55% of the unigenes matched to a protein product. The remaining 39,657 had no function assigned (Panel A in S2 Fig). The best blast-hits belonged to *Vitis vinifera* (44%), *Populus trichocarpa* (15%) and *Ricinus communis* (5%) (Panel A in S2 Fig). Approximately 24.4% and 6.1% of the unigenes was assigned to at least one Gene Ontology (GO) and KEGG orthology (KO) terms, respectively. Unigenes were

**Table 1. Overview of the raw data output from the 454-FLX Titanium sequencing.**

Sample	Total bases	Reads	Mean length (bp)
'Ortice' control	63,611,956	173,118	367
'Ortice' damaged	73,464,937	197,782	371
'Ruveia' control	52,792,933	146,765	360
'Ruveia' damaged	61,220,898	177,846	344
Total	251,090,724	695,511	

<https://doi.org/10.1371/journal.pone.0183050.t001>

distributed in 14 GO-terms for the Biological Process ontology, 9 for Cellular Component and 10 for Molecular Function. Metabolic process sub-category, consisting of 11,167 genes, was dominant in biological process. Binding and cell part subcategories, consisting of 13,820 and 8,292 genes, were dominant in molecular function and cellular component, respectively. A considerable number of genes was included in cellular process, catalytic activity and intracellular sub-categories (Panel B in [S2 Fig](#)). To identify main biological pathways, we mapped the annotations to reference canonical pathways in the KEGG database by using KO identifiers (i.e. a classification of orthologous genes defined by KEGG). The 5,355 KO allocated genes were assigned to 265 KEGG categories. The most abundant processes/metabolic pathways were "biosynthesis of secondary metabolites" (185 unigenes); "microbial metabolism in diverse environments" (89); "spliceosome" (74); "biosynthesis of plant hormones" (72); "ribosome" (61); "RNA transport" (61) and "protein processing in endoplasmic reticulum" (60) (Panel B in [S2 Fig](#)).

### Differential gene expression

To profile the variation in gene expression of olives during *B. oleae* feeding and to characterize the differences between the two varieties with contrasting susceptibility to the olive fruit fly, unigenes were used to build a CustomArray™ (CombiMatrix Corporation). The chip layout consists of 61,825 olive probes out of the 87,720 assembled sequences (Panel B in [S1 Fig](#)). The DNA microarrays were then hybridized with labeled cRNA prepared from undamaged olives of the varieties 'Ortice' or 'Ruveia' and corresponding samples with feeding larvae.

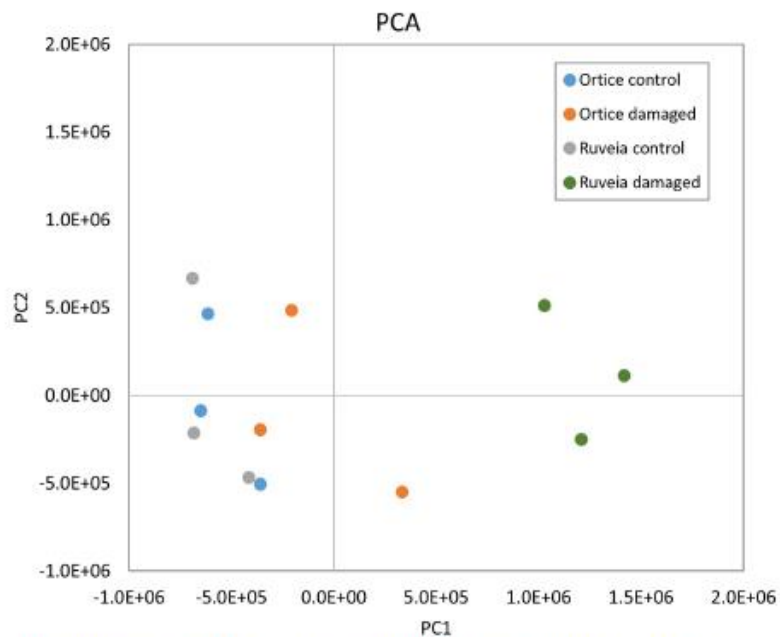
After normalization ([S3 Fig](#)), the expression data relative to all the olive probes were analysed by PCA to explore the variation among the different conditions based on gene expression states ([Fig 1](#)).

The variability among the different conditions and biological replicates could be summarized and visualized in two dimensions. The first component, which accounted for 46.0% of total variance, appears to represent the overall induction due to the fruit fly, as it discriminates the different biological response of the two varieties. The second component, which explained 14.1% of the variance, mainly discriminated the biological replicates for each condition. Because of a clear *B. oleae* effect, the limited difference among the control conditions of the

**Table 2. Main indices and statistics of the assembly.**

Samples in assembly	Reads in Assembly (%)	Contigs		Singletons	
		Number	Average length (bp)	Number	Average length (bp)
'Ortice' control	570,878 (82.93)	15,058	884	72,662	333
'Ortice' damaged					
'Ruveia' control					
'Ruveia' damaged					

<https://doi.org/10.1371/journal.pone.0183050.t002>



**Fig 1. Principal Component Analysis of the normalized microarray data of the olive probes.** The percentage of the total variation explained by the first two principal components (PC1 and PC2) was 46.0 and 14.1, respectively. The PCA plot shows separation of the four experimental groups, indicated by different colors, along the first component and clustering of the three biological replicates, represented by a dot, within these groups.

<https://doi.org/10.1371/journal.pone.0183050.g001>

two cultivars and the dispersion of the biological replicates, we retained all conditions and replicates for further analysis.

To identify differentially expressed probes, we applied the following selection criteria: a  $p$ -value  $< 0.05$  (Welch  $t$ -test, followed by a Bonferroni-Holchberg FDR correction for multiple testing) and a  $|\log_2\text{Ratio}| > 1$ . The number of differentially expressed genes, after removal of duplicates, is reported in [Table 3](#).

DEGs and their annotation are listed in [S1 File](#). The heatmap illustrates a weak linkage among the four conditions ([S4 Fig](#)). To validate the microarray results, the expression of six genes (four differentially expressed genes and two non-affected by *B. oleae*) was analyzed by Real Time-PCR. The results were consistent to the microarray data ([S5 Fig](#)).

**Table 3. Differentially expressed genes (DEGs) in pairwise comparisons.**

	Upregulated	Downregulated
'Ruveia' control vs 'Ortice' control	23	7
'Ortice' damaged vs 'Ortice' control	17	75
'Ruveia' damaged vs 'Ruveia' control	1071	1528
'Ruveia' damaged vs 'Ortice' damaged	714	537

<https://doi.org/10.1371/journal.pone.0183050.t003>

The smallest difference, in terms of differentially expressed genes, was present between the undamaged olives of the two cultivars. The response to the *B. oleae* feeding of the susceptible cultivar 'Ortice' also involved a limited set of genes and the majority (77%) were down-regulated. The response of the more tolerant variety 'Ruveia' involved the highest number of both up-regulated and down-regulated genes. When attacked, olives of the 'Ruveia' cultivar differentially expressed more than 1000 genes compared to attacked drupes of the 'Ortice'. To investigate the commonality between the olive varieties and their response, DEGs in the pairwise comparisons were matched. The response of 'Ruveia' is highly specific, as indicated by the limited number of common genes. Around half of the genes down-regulated in 'Ortice' were also down-regulated in 'Ruveia' (Fig 2).

The majority of the genes overexpressed in 'Ruveia' after *B. oleae* attack were also upregulated when comparing the damaged 'Ruveia' and 'Ortice' drupes. Overall, the comparison of the various response to the *B. oleae* indicated that two varieties have a markedly different reaction.

### Transcript clustering

Considering the limited overlap of the molecular response of the two varieties, the expression level of the statistically significant genes in at least one of the four comparisons was processed by K-means clustering analysis to infer possible co-expressed or co-regulated genes (Fig 3).

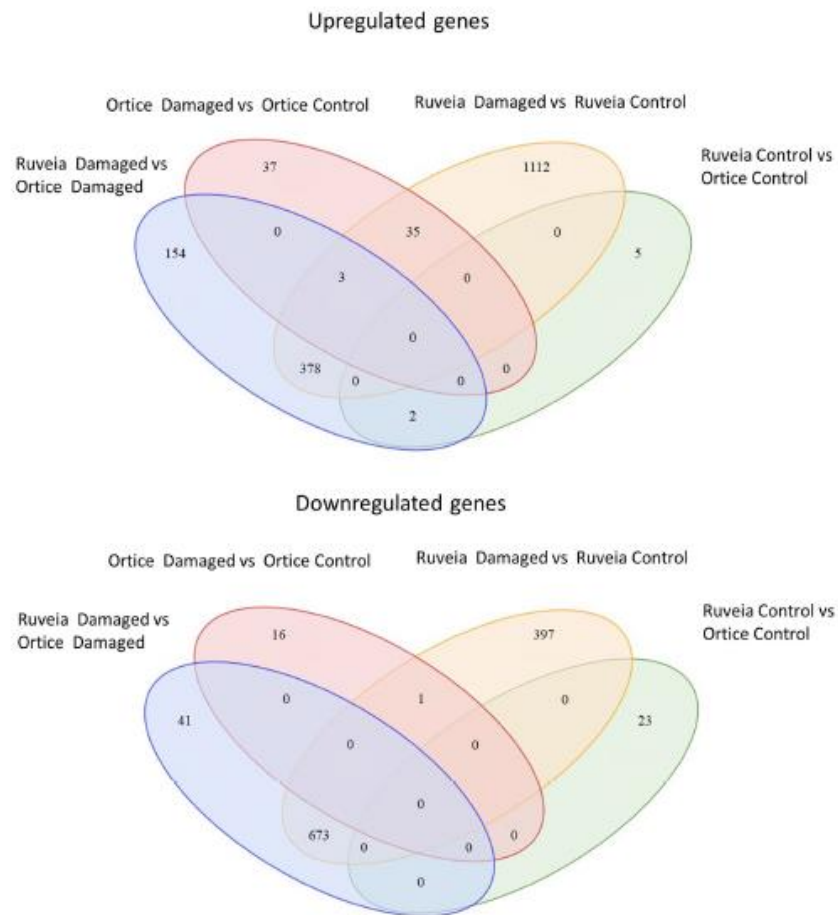
This analysis allowed to group the 2848 DEGs in 10 clusters (S3 Table). K-means clustering indicated that it was not possible to identify a cluster in which genes are overexpressed in the two attacked varieties with a similar pattern, further indicating a very limited overlap between the responses of the two olive varieties. The GO annotation (level 4) in the "Biological Process" domain of the genes belonging to the clusters (sequence cut-off 5%) is reported in Fig 4.

To ease the comparison, the graph reports, for each cluster, the different GO categories in relative terms. The annotation indicated that K7 and K4 are the most complex clusters because they include, respectively, 12 and 11 out of the 15 considered GO terms. Taking as reference the average expression pattern, clusters K7 and K4 mainly comprise genes that are, respectively, highly and mildly up-regulated exclusively in the damaged drupes of the more tolerant cultivar 'Ruveia'. Similarly, genes belonging to clusters K1 and K6, specifically modulated in 'Ruveia' drupes, were annotated with a high number of GO categories.

The most present GO terms in the different clusters were related to chemical reactions and pathways involving macromolecule, such as "macromolecule metabolic process" (present in all the ten clusters), "cellular macromolecule metabolic process" and "cellular nitrogen compound metabolic process" (present in nine clusters). K4 is the only cluster containing transcripts related to "carbohydrate metabolic processes" while the "post-embryonic development" term was exclusively present in K7. This analysis indicated the degree of specificity of the molecular events and biological processes characterizing the response of the tolerant cultivar 'Ruveia' to *B. oleae* attack.

### Expression analysis of the 'Ruveia' response to fruit fly

We used GO terms association (obtained by a sequence similarity search) to elucidate the biological objective to which the DEGs contribute in the response to the *B. oleae* feeding. GO analysis indicated that the transcriptional reconfiguration involved a range of biological processes (S6 Fig). To provide an optimal view of the dataset's most relevant terms, a summary of the annotation results is presented as a multi-level pie that shows the lowest GO terms per branch that fulfil our annotation criterion weight (Fig 5).

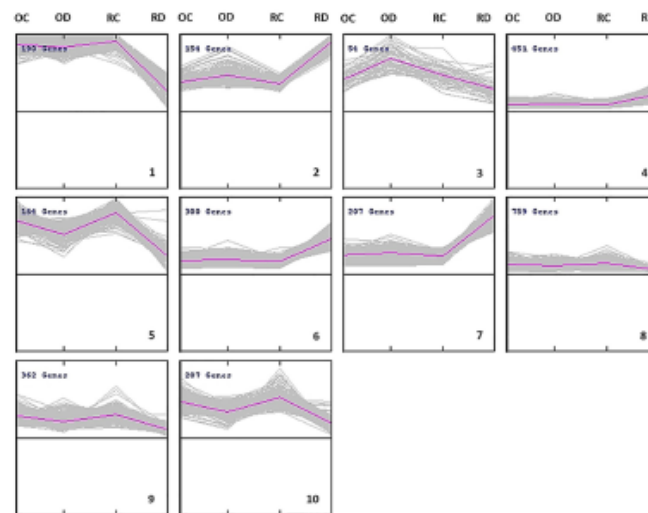


**Fig 2. Venn diagram of the differentially expressed genes.** The intersecting portions of the Venn diagrams report the number of common genes among the different comparisons between the two varieties ('Ruveia' and 'Ortice') in the two experimental conditions (control and *B. oleae* damaged).

<https://doi.org/10.1371/journal.pone.0183050.g002>

The functional profile of both down and upregulated sequences indicated that the most abundant GO term categories were related to chemical reactions and pathways resulting in the formation of substances, including protein modification processes. Collectively, response to stress and to abiotic or biotic stimulus ranked the first most affected biological process for the down-regulated genes and at the eighth place for the up-regulated sequences.

Among the up-regulated genes involved in the response to stress, there were transcripts coding for proteins involved in signaling, including receptors and transcription factors, such as two leucine-rich repeat receptors, nine WD-repeat containing proteins, five WRKY transcription factors, and two rpm 1-like proteins. The latter are essential regulator of plant defense

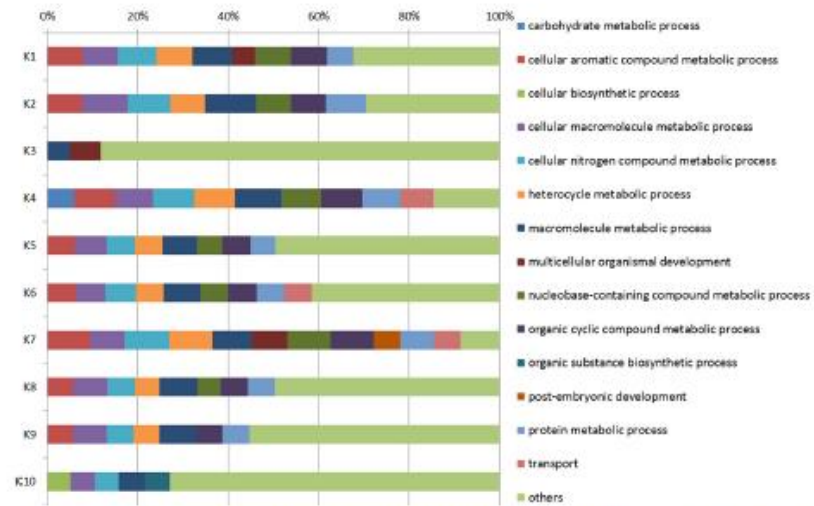


**Fig 3. K-means clustering of the differentially expressed gene (FDR < 0.05 and  $|\log_2 \text{Ratio}| > 1$ ).** In each box, grey lines represent genes, whose expression level vary in the four experimental conditions namely, 'Ortice' undamaged (OC), 'Ortice' damaged (OD), 'Ruveia' undamaged (RC) and 'Ruveia' damaged (RD). A pink line represents the average expression profile of all genes in each cluster. In the top-left corner, it is reported the total number of genes in the cluster. A progressive cluster (K) number is indicated in the bottom-right corner.

<https://doi.org/10.1371/journal.pone.0183050.g003>

and are typically associated to the resistance to pathogens [36]. Other up-regulated genes typically associated to pathogenesis included two PR proteins, one disease resistance response protein, one late blight resistance protein homologue and a NBS domain resistance protein. Among the genes that can directly deter insect growth, there were three beta-glucosidases. In fruits, beta-glucosidase activity plays an important role in the oleuropein metabolism, catalyzing its hydrolysis into a toxic glutaraldehyde-like structure that acts as defense mechanism against insects [37]. Moreover, four serine carboxypeptidase-like (SCPL) genes were up-regulated. In tomato, Arabidopsis and rice, some SCPL proteins are wound inducible and connected to jasmonic acid (JA) pathway [38, 39]. Other up-regulated genes possibly related to JA pathway include two phospholipases. Genes typically associated to abiotic stress were also up-regulated, such as two transcripts coding for dehydration-induced proteins, two heat-shock proteins and a bobber1-like protein. The 'Ruveia' response to *B. oleae* also included genes involved in phytohormone signaling such as those coding for proteins involved in auxin (i.e. auxin binding protein, auxin efflux carrier family protein, auxin response factor, indole-3 acetic acid amino-acid hydrolase), brassinosteroid (i.e. brassinosteroid insensitive 1-associated receptor kinase 1 like, bri1 suppressor 1 like) and ethylene signaling (i.e. four ethylene responsive transcription factors). Besides stress hormones, the plant perception and the signal transduction of *B. oleae* infestation seems to involve also other stress-related cellular messengers such ROS and calcium, as indicated by the up-regulation of five calcium-dependent CBL-interacting serine/threonine protein kinases, a calcium transporting ATPase plasma membrane, four glutaredoxin, and one oxoglutarate genes.

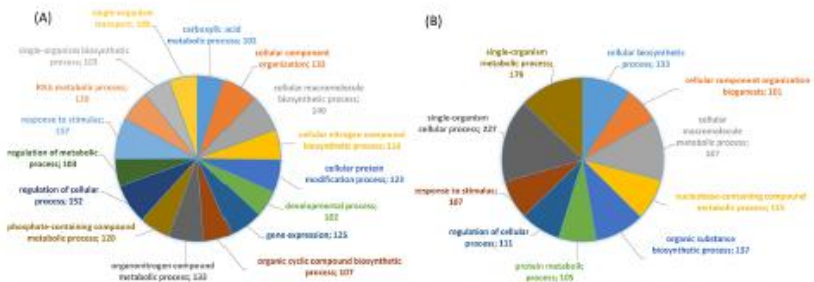




**Fig 4. Functional categories of the clusters of the DEGs (FDR < 0.05 and |log<sub>2</sub> Ratio| > 1).** Each horizontal bar refers to the clusters (numbered from K1 to K10; see Fig 3) deriving from K-means clustering. Bars are partitioned into colored segments whose length represents the relative number of sequences for each GO term (level 4; BP; sequence cut-off: 5%).

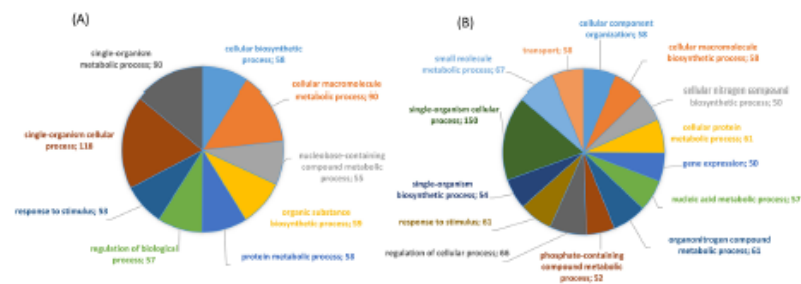
<https://doi.org/10.1371/journal.pone.0183050.g004>

KEGG pathway map analysis of annotated enzymatic activities indicated that purine metabolism ranked first as number of sequences. Among other physiological processes, variation in purine metabolism associates to fruit ripening and to uracil salvage activity during senescence. Moreover, plants produce toxic secondary metabolites from pyrimidines that act as defense compounds [40]. The "Biosynthesis of antibiotics" KEGG reference pathway ranked first considering the number of enzymatic activities. This molecular interaction network diagram comprises an ample number of reactions that in plants are associated to terpenoid backbone biosynthesis, the shikimate pathway and the biosynthesis of secondary metabolites. Starch and sucrose metabolism ranked second for both the number of sequences and enzymatic activities.



**Fig 5. Multilevel distribution of the differentially expressed annotated sequences by GO category.** The lowest terms with a minimum of 100 annotated sequences are shown. For each category it is indicated the number of annotated sequences. (A) downregulated genes; (B) upregulated genes.

<https://doi.org/10.1371/journal.pone.0183050.g005>



**Fig 6. Multilevel distribution of the differentially expressed annotated sequences by GO category.** The lowest terms with a minimum of 50 annotated sequences are shown. For each category, it is indicated the number of annotated sequences. (A) downregulated genes; (B) upregulated genes.

<https://doi.org/10.1371/journal.pone.0183050.g006>

### Differential expression analysis of the damaged drupes of the two varieties

To understand the inducible factors that can account for the different tolerance of the two varieties, we compared the gene expression levels between 'Ortica' and 'Ruveia' drupes when infested by *B. oleae*. The total number of DEGs was less than half compared with the response of the tolerant cultivar 'Ruveia' to the fruit fly. The proportion of the up-regulated genes (57%) was higher compared to the 'Ruveia' response, indicating that the different tolerance between the cultivars lies in the activation of a wide number of genes. In the Biological Process domain, differentially expressed sequences were assigned to a number of processes. At the GO level 2, the largest difference in ranking between the up-regulated and down-regulated sequences was for "multicellular organismal process". On the opposite, "metabolic process" in comparative terms, it is the most characterizing GO term in the down-regulated sequences (S7 Fig).

A summary of the GO annotation results for the down-regulated or the up-regulated sequences is presented as a multi-level pie (Fig 6).

The dominant GO term for both upregulated and downregulated sequences was "single-organism process". "Response to stimulus" ranked fourth for the up-regulated sequences and ninth for the down-regulated sequences.

Among the up-regulated genes, eight sequences were associated to the GO term "response to biotic stimulus" such as a WRKY transcription factor 17, two putative serine/threonine kinases, and a Serine carboxypeptidase like. Genes coding for proteins putatively involved in phytohormone signaling included an auxin efflux carrier family protein and abri1 suppressor 1 like. These genes were also upregulated in 'Ruveia' cultivar in response to the olive fruit fly. In total, almost 95% of the genes up-regulated in the comparison between the attacked drupes of the two cultivars were also upregulated in the damaged fruits of the tolerant variety ('Ruveia'). The data indicated that the different tolerance to the fruit fly in the two cultivars under investigation is mainly inducible.

### Discussion

The use of olive varieties that are highly tolerant to the fruit fly is an important element to reduce economic loss and use of chemical pesticides. Knowledge on the molecular aspects underlying the different resistance response to the fruit fly can assist the screening of more suitable genotypes and ultimately, contribute to the development of new integrated control strategies. For these reasons, we compared two olive cultivars with contrasting resistance levels

to *B. oleae*. To this aim, we first generated a large collection of unigenes from our NGS data. The assembly did not include about 20% of raw sequences, a limited number considering the abundance of short repeats and the high level of heterozygosity of the olive [27, 41]. The reduced number of unassembled reads should have been also affected by the smaller number of cultivars and plant tissues under investigation in comparison with other works [41]. The 454 DNA sequencing method confirmed to be an effective technology in revealing the expression of a large number of genes in a non-model organism mainly because its longer read length [42].

We pooled the pyrosequencing samples derived from a number of conditions to generate a microarray based on the drupe sequence information and obtained a number of unigenes comparable to other works in olive [41]. Similarly, the percentage of an annotation of our assembled dataset was comparable or higher to previous olive ESTs sequencing efforts [27, 41, 43]. The setup of a microarray allows analyses of gene expression and functional genomics studies in olive at a fraction of the cost of next generation sequencing. Moreover, microarray data are more computationally tractable and do not require an extended bioinformatics effort [44]. Our layout specifically focused on the drupe and was based on the widely used Combimatrix technology, hence extending currently available options [45, 46].

A key interest in this study was to compare the response of genotypes with different level of tolerance to the olive fruit fly. In absence of infestation, very little difference was noticed between the two cultivars. This is justified considering that we analysed plants present in the same environment, not only to reduce environmental variability but also to have the same *B. oleae* population level. A diverse expression profile between the tolerant and susceptible genotype was clearly observed only in the presence of larvae. Within a species, the scale and quality of response to herbivores as well as the accumulation of defence compounds may vary significantly [47–50]. In *Brassica oleracea*, there was a very little overlap between transcriptional responses of two varieties with contrasting level of resistance to the cabbage aphid (*Brevicoryne brassicae*), underlying that intraspecific variation in susceptibility to insect pests can be also explained by differences in induced transcriptional changes [51]. The here described strong difference between the tolerant and the susceptible variety also implies a significant genotype-specific response in olive. This would be consistent with the frequently reported genetic diversity and the large diversified phenotypic traits of the plant and of olive oil products [52]. The data suggest a strong relevance of the genetic component in the determination of the tolerant phenotype [53, 54] also because trees grew in the same environment using the same agronomic practices.

Plant defence pathways can be manipulated by also pathogens or pests [55, 56]. For instance, glucose oxidase, one of the principal components of *Helicoverpa zea* saliva, suppresses some induced resistance in *Nicotiana tabacum* by directly inhibiting the wound-signalling molecule jasmonic acid and/or by antagonizing its interaction with other signalling pathways [57]. In tomato, the spider mite *Tetranychus evansi* is able to suppress the induction of genes involved in the induced plant defences, such as proteinase inhibitors [58]. Evidence for the defence suppression by herbivores is not as large as for defence induction yet, it is conceivable that plant resistance mechanisms promote the selection of counter-adaptive mechanisms in biotic stressors, especially in compatible interactions involving monophagous pests [59]. Our data suggest that the successful colonization of drupes by *B. oleae* is likely to be due to a weak reaction in the more susceptible variety and that drupes of the more tolerant variety are a less favourable nourishment because of a more active and composite molecular response [60].

The gene expression in the attacked drupes of cv. 'Ruveia' identified a number of sequences directly and indirectly involved in the induced resistance mechanism. Plant response to

herbivorous pest encompasses a number of mechanisms [61] and therefore, several molecular processes are activated by the drupe defence to the fruit fly. Upregulated genes include those involved in the defence mechanisms against biotic stress, both wounding and pathogen attack, as well as abiotic stress, such as drought and high or low temperature. The overexpression of genes involved in different defense pathways suggests that the inducible response to *B. oleae* larvae has a degree of overlap with the response to pathogens, which was also noted for the interaction between olive and *S. oleaginea* [62]. Moreover, our data is consistent with other research works that indicated that beta-glucosidases are important element of the olive defense to *B. oleae* [14, 15, 37]. Beta-glucosidases promote the formation of toxic glutaraldehyde-like structure from oleuropein, a phenolic compound that has been linked to the different susceptibility of olive cultivars [16]. DEGs were also involved in the production or response to phytohormones and molecules involved in pest response (e.g., jasmonic acid and ROS).

Besides genes associated to stress resistance, the Gene Ontology analysis suggested that biological processes related to secondary metabolism, cation transport and transmembrane transport are also affected, giving reasons to believe that numerous changes in plant primary metabolism should occur in response to larva feeding [63]. Recently, it has been reported that *B. oleae* infestation causes significant changes in mineral elements (such as P, K, Fe and Mg) in fruits [23]. Also abiotic stress, such as a moderate drought, affects different metabolites during fruit development, including terpenes [64]. Collectively, the existing molecular and metabolomics data suggest that the defensive reaction of the tolerant cultivar to the persisting feeding of the *B. oleae* comprises a larger than anticipated metabolic reprogramming in infested tissues, yet to be fully described [7, 65].

## Conclusions

The *B. oleae* feeding influences pathways with a known role in defence, oxidative stress responses, and genes involved in plant structure and metabolism. Defence against the long-lasting fruit fly larva feeding is a complex trait and involves multiple molecular mechanisms. The complexity of the drupe response suggests that a number of features, metabolites and signalling pathways effectively limit the fruit fly infestation. An additional level of complexity derives from the marked genotype-specific difference present in olive that is suggested by our study [66]. The extended gene expression difference between the tolerant and susceptible olive varieties under investigation also indicated that it would be possible to identify genetic factors that are associated with a higher level of tolerance to *B. oleae*.

## Supporting information

**S1 Fig. Length distribution of the assembled sequences (A) and of those used for the microarray (B).**  
(TIF)

**S2 Fig. Functional annotation of the assembled unigenes.** A) Distribution of the unigenes according to the presence (matched) or absence (non-matched) of hits retrieved from the NCBI database. The pie of the pie chart illustrates the relative distribution of the best-BlastX hits according to the plant species. B) Classification of the unigenes in the GO domains Biological Process, Cellular Component and Molecular Function. For each domain, the table reports the GO-terms (category) ranked in decreasing order according to the number of sequences (singleton/TCs). The bar chart illustrates for each GO-term the relative amount in relation to the number of sequences annotated in the GO-domain and the total number of sequences

assembled.

(TIFF)

**S3 Fig. Box plots of signal intensity after quantile normalization for each microarray hybridization.** Legend: Ort: 'Ortice'; Ruv: 'Ruveia'; S: control condition; T: test condition; Roman numbers denote, per each condition, the biological replicate.

(TIFF)

**S4 Fig. Heatmap of the differentially expressed genes.** The heatmap shows the relative expression level of the DEGs in the four experimental comparisons. 1: 'Ruveia' Test vs 'Ruveia' Control condition; 2: 'Ruveia' Test vs 'Ortice' Test condition; 3: 'Ortice' Test vs 'Ortice' Control condition; 4: 'Ruveia' Control vs 'Ortice' Control condition. Gradation from green to red is relative to the  $\log_2$  Fold Change (FC) values. Similarities were calculated using Euclidean distances and agglomeration was performed according to the complete-linkage algorithm.

(TIFF)

**S5 Fig. Real time RT-PCR validation of the microarray results.** The expression level of four DEGs (1, 2, 3 and 4) and two transcripts that were not affected by *B. oleae* (5 and 6) was analysed by real-time RT-PCR in drupes of the 'Ortice' cultivar. Quantities are reported on a linear scale relative to the calibrator condition (undamaged olives). 1: G0MWCVVW03GE3QK; 2: G0MWCVVW01A1C2H; 3: G0MWCVVW03FSPSU; 4: G0MWCVVW04JNBZP; 5: contig04878; 6: G0MWCVVW02DVZFG. See [S1 Table](#) for details on the transcripts. For each transcript, an asterisk indicates a significant difference with the control condition (Student t-test; \*;  $p < 0.05$ ).

(TIFF)

**S6 Fig. Level 2 chart summary of GO term association for the overexpressed (orange bars) and under expressed (blue bars) sequences following *B. oleae* feeding of the 'Ruveia' drupes.**

(TIFF)

**S7 Fig. Level 2 chart summary of GO term association for the up-regulated (orange bars) and down-regulated (blue bars) sequences of the comparison between infested drupes of the 'Ruveia' and 'Ortice' varieties.**

(TIFF)

**S1 Table. Primer used for qRT-PCR validation of the microarray results.**

(XLSX)

**S2 Table. List of all unigenes and their predicted functional annotation.**

(XLSX)

**S3 Table. K-means clusters of the differentially expressed genes.**

(XLSX)

**S1 File. Differentially expressed genes and their annotation.** Table A. List and annotation of the underexpressed genes ( $FDR < 0.05$ ;  $\log_2 \text{Ratio} < 1$ ) between attacked drupes of the Ruveia and attacked drupes of the Ortice cultivar. Table B. List and annotation of the overexpressed genes ( $FDR < 0.05$ ;  $\log_2 \text{Ratio} > 1$ ) between attacked drupes of the Ruveia and the attacked drupes of the Ortice cultivar. Table C. List and annotation of the underexpressed genes ( $FDR < 0.05$ ;  $\log_2 \text{Ratio} < 1$ ) between attacked and control drupes of the Ruveia cultivar. Table D. List and annotation of the overexpressed genes ( $FDR < 0.05$ ;  $\log_2 \text{Ratio} > 1$ ) between attacked and control drupes of the Ruveia cultivar. Table E. List and annotation of the underexpressed genes ( $FDR < 0.05$ ;  $\log_2 \text{Ratio} < 1$ ) between attacked and control drupes of the Ortice cultivar.

F. List and annotation of the overexpressed genes (FDR:  $<0.05$ ;  $\log_2\text{Ratio} >1$ ) between attacked and control drupes of the Ortice cultivar. Table G. List and annotation of the under-expressed genes (FDR:  $<0.05$ ;  $\log_2\text{Ratio} <1$ ) between control drupes of the Ruveia and control drupes of the Ortice cultivar. Table H. List and annotation of the overexpressed genes (FDR:  $<0.05$ ;  $\log_2\text{Ratio} >1$ ) between control drupes of the Ruveia and control drupes of the Ortice cultivar.  
(XLSX)

### Acknowledgments

We thank the Azienda Agricola Sperimentale Regionale "Improsta" (Eboli, SA) for the kind collaboration and free access to olive germplasm and Anna Rana for assistance during qRT-PCR experiments.

### Author Contributions

**Conceptualization:** Fabrizio Carbone, Luciana Baldoni, Rosa Rao.

**Formal analysis:** Filomena Grasso, Mariangela Coppola, Fiammetta Alagna, Gaetano Perrotta, Antonio J. Pérez-Pulido, Giandomenico Corrado.

**Funding acquisition:** Fabrizio Carbone, Luciana Baldoni, Gaetano Perrotta, Rosa Rao.

**Investigation:** Filomena Grasso, Mariangela Coppola, Fiammetta Alagna, Antonio J. Pérez-Pulido, Antonio Garonna, Paolo Facella, Loretta Daddiego, Loredana Lopez, Alessia Vitello, Giandomenico Corrado.

**Supervision:** Fabrizio Carbone, Gaetano Perrotta, Antonio Garonna, Rosa Rao, Giandomenico Corrado.

**Writing – original draft:** Fabrizio Carbone, Gaetano Perrotta, Antonio J. Pérez-Pulido, Giandomenico Corrado.

**Writing – review & editing:** Luciana Baldoni, Rosa Rao, Giandomenico Corrado.

### References

1. Nardi F, Carapelli A, Boore J, Roderick G, Dallai R, Frati F. Domestication of olive fly through a multi-regional host shift to cultivated olives: Comparative dating using complete mitochondrial genomes. *Molecular phylogenetics and evolution*. 2010; 57(2):678–86. <https://doi.org/10.1016/j.ympev.2010.08.008> PMID: 20723608
2. Zygouridis N, Augustinos A, Zalom F, Mathiopoulos K. Analysis of olive fly invasion in California based on microsatellite markers. *Heredity*. 2009; 102(4):402–12. <https://doi.org/10.1038/hdy.2008.125> PMID: 19107137
3. Fletcher B. The biology of dactynotus fruit flies. *Annual review of entomology*. 1987; 32(1):115–44.
4. Daane KM, Johnson MW. Olive fruit fly: managing an ancient pest in modern times. *Annual review of entomology*. 2010; 55:151–69. <https://doi.org/10.1146/annurev.ento.54.110807.090553> PMID: 19961328
5. Malheiro R, Casal S, Baptista P, Pereira JA. A review of *Bactrocera oleae* (Rossi) impact in olive products: From the tree to the table. *Trends in Food Science & Technology*. 2015; 44(2):226–42.
6. Gómez-Caravaca AM, Cerretani L, Bendini A, Segura-Carretero A, Fernández-Gutiérrez A, Del Carlo M, et al. Effects of fly attack (*Bactrocera oleae*) on the phenolic profile and selected chemical parameters of olive oil. *Journal of Agricultural and Food Chemistry*. 2008; 56(12):4577–83. <https://doi.org/10.1021/jf800118t> PMID: 18522402
7. Gucci R, Caruso G, Canale A, Loni A, Raspi A, Urbani S, et al. Qualitative changes of olive oils obtained from fruits damaged by *Bactrocera oleae* (Rossi). *HortScience*. 2012; 47(2):301–6.

8. Angerosa F, Giacinto LD, Solinas M. Influence of *Dacus oleae* infestation on flavor of oils, extracted from attacked olive fruits, by HPLC and HRGC analyses of volatile compounds. *Grasas y Aceites*. 1992; 43(3):134–42.
9. Medjkouh L, Tamendjari A, Keciri S, Santos J, Nunes MA, Oliveira M. The effect of the olive fruit fly (*Bactrocera oleae*) on quality parameters, and antioxidant and antibacterial activities of olive oil. *Food & Function*. 2016; 7(6):2780–8.
10. Vontas J, Hejazi M, Hawkes NJ, Cosmidis N, Loukas M, Hemingway J. Resistance-associated point mutations of organophosphate insensitive acetylcholinesterase, in the olive fruit fly *Bactrocera oleae*. *Insect molecular biology*. 2002; 11(4):329–36. PMID: [12144698](#)
11. Tsolakis H, Ragusa E, Tarantino P. Control of *Bactrocera oleae* by low environmental impact methods: NPC methodology to evaluate the efficacy of lure-and-kill method and copper hydroxide treatments. *Bulletin of Insectology*. 2011; 64(1):1–8.
12. Daane KM, Wang X, Nieto DJ, Pickett CH, Hoelmer KA, Blanchet A, et al. Classic biological control of olive fruit fly in California, USA: release and recovery of introduced parasitoids. *BioControl*. 2015; 60(3):317–30.
13. Hoelmer KA, Kirk AA, Pickett CH, Daane KM, Johnson MW. Prospects for improving biological control of olive fruit fly, *Bactrocera oleae* (Diptera: Tephritidae), with introduced parasitoids (Hymenoptera). *Biocontrol science and technology*. 2011; 21(9):1005–25.
14. Corrado G, Alagna F, Rocco M, Renzone G, Varricchio P, Coppola V, et al. Molecular interactions between the olive and the fruit fly *Bactrocera oleae*. *BMC Plant Biology*. 2012; 12(1):1.
15. Alagna F, Kallenbach M, Pompa A, De Marchis F, Rao R, Baldwin IT, et al. Olive fruits infested with olive fly larvae respond with an ethylene burst and the emission of specific volatiles. *Journal of Integrative Plant Biology*. 2015.
16. Iannotta N, Scalerio S. Susceptibility of Cultivars to Biotic Stresses. In: Mazzalupo I, editor. *Olive Germplasm—The Olive Cultivation, Table Olive and Olive Oil Industry in Italy*: INTECH Open Access Publisher; 2012. p. 81–106.
17. Rizzo R, Caleca V, editors. Resistance to the attack of *Bactrocera oleae* (Gmelin) of some Sicilian olive cultivars. Proceedings of Olivebioteq 2006, Second International Seminar "Biotechnology and quality of olive tree products around the Mediterranean Basin" November 5th–10th, Mazara del Vallo, Marsala, Italy; 2006.
18. Di Vaio C. Il germoplasma dell'olivo in Campania. Napoli (IT): Assessorato all'Agricoltura Regione Campania; 2012. 92 p.
19. Pugliano G. La risorsa genetica dell'olivo in Campania. Napoli (IT): Regione Campania; 2000. 158 p.
20. Corrado G, Garonna A, Cabanás CG-L, Gregoriou M, Martelli GP, Mathiopoulos KD, et al. Host Response to Biotic Stresses. In: Rugini E, Baldoni L, Muleo R, Sebastiani L, editors. *The Olive Tree Genome*. Springer International Publishing; 2016. p. 75–98.
21. Malheiro R, Casal S, Cunha SC, Baptista P, Pereira JA. Identification of leaf volatiles from olive (*Olea europaea*) and their possible role in the ovipositional preferences of olive fly, *Bactrocera oleae* (Rossi) (Diptera: Tephritidae). *Phytochemistry*. 2016; 121:11–9. <https://doi.org/10.1016/j.phytochem.2015.10.005> PMID: [26603276](#)
22. Lo Scalzo RL, Scarpati ML, Verzegnassi B, Vita G. *Olea europaea* chemicals repellent to *Dacus oleae* females. *Journal of Chemical Ecology*. 1994; 20(8):1813–23. <https://doi.org/10.1007/BF02066224> PMID: [24242710](#)
23. Garantonakis N, Varkou K, Markakis E, Birouraki A, Sergentani C, Psarras G, et al. Interaction between *Bactrocera oleae* (Diptera: Tephritidae) infestation and fruit mineral element content in *Olea europaea* (Lamiales: Oleaceae) cultivars of global interest. *Applied Entomology and Zoology*. 2016; 51(2):257–65.
24. Scarpati ML, Lo Scalzo RL, Vita G, Gambacorta A. Chemotropic behavior of female olive fly (*Bactrocera oleae* Gmel.) on *Olea europaea* L. *Journal of Chemical Ecology*. 1996; 22(5):1027–36. <https://doi.org/10.1007/BF02029952> PMID: [24227622](#)
25. Rizzo R, Caleca V, Lombardo A. Relation of fruit color, elongation, hardness, and volume to the infestation of olive cultivars by the olive fruit fly, *Bactrocera oleae*. *Entomologia Experimentalis et Applicata*. 2012; 145(1):15–22.
26. Uceda M, Frias L. Harvest dates. Evolution of the fruit oil content, oil composition and oil quality. Proc II Seminario Oleícola Internacional, Córdoba (Spain), 1975:6–17.
27. Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, et al. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics*. 2009; 10(1):1.
28. Coppola V, Coppola M, Rocco M, Digilio MC, D'Ambrosio C, Renzone G, et al. Transcriptomic and proteomic analysis of a compatible tomato-aphid interaction reveals a predominant salicylic acid-dependent plant response. *BMC Genomics*. 2013; 14(1):1.

29. R Developmental Team Core. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2009.
30. Voigt C. Synthetic Biology: Methods for part/device characterization and chassis engineering. Academic Press; 2011.
31. Saeed A, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. 2003; 34(2):374. PMID: [12613259](#)
32. Zimmermann P, Schildknecht B, Craigon D, Garcia-Hernandez M, Gruissem W, May S, et al. MIAME/Plant—adding value to plant microarray experiments. *Plant Methods*. 2006; 2(1):1.
33. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*. 2008; 36(10):3420–35. <https://doi.org/10.1093/nar/gkn176> PMID: [18445632](#)
34. Myhre S, Tveit H, Mollestad T, Laegreid A. Additional gene ontology structure for improved biological reasoning. *Bioinformatics*. 2006; 22(16):2020–7. <https://doi.org/10.1093/bioinformatics/btl334> PMID: [16787968](#)
35. Muñoz-Mérida A, Viguera E, Claros MG, Trelles O, Pérez-Pulido AJ, Sma3s: a three-step modular annotator for large sequence datasets. *DNA research*. 2014; 21(4):341–53. <https://doi.org/10.1093/dnares/dsu001> PMID: [24501397](#)
36. Selote D, Kachoo A. RIN4-like proteins mediate resistance protein-derived soybean defense against *Pseudomonas syringae*. *Plant Signaling & Behavior*. 2010; 5(11):1453–6.
37. Koudounas K, Banilas G, Michaelidis C, Demoliou C, Rigas S, Hatzopoulos P. A defence-related *Olea europaea*  $\beta$ -glucosidase hydrolyses and activates oleuropein into a potent protein cross-linking agent. *Journal of Experimental Botany*. 2015; 66(7):2093–106. <https://doi.org/10.1093/jxb/erv002> PMID: [25697790](#)
38. Liu H, Wang X, Zhang H, Yang Y, Ge X, Song F. A rice serine carboxypeptidase-like gene OsBISCP1 is involved in regulation of defense responses against biotic and oxidative stress. *Gene*. 2008; 420(1):57–65. <https://doi.org/10.1016/j.gene.2008.05.006> PMID: [18571878](#)
39. Moura DS, Bergery DR, Ryan CA. Characterization and localization of a wound-inducible type I serine-carboxypeptidase from leaves of tomato plants (*Lycopersicon esculentum* Mill.). *Planta*. 2001; 212(2):222–30. <https://doi.org/10.1007/s004250000380> PMID: [11216843](#)
40. Kafer C, Zhou L, Santos D, Guirgis A, Weers B, Park S, et al. Regulation of pyrimidine metabolism in plants. *Front Biosci*. 2004; 9:1611–25. PMID: [14977572](#)
41. Muñoz-Mérida A, González-Plaza JJ, Blanco AM, del Carmen García-López M, Rodríguez JM, Pedrola L, et al. De novo assembly and functional annotation of the olive (*Olea europaea*) transcriptome. *DNA research*. 2013; 20(1):93–108. <https://doi.org/10.1093/dnares/dss036> PMID: [23297299](#)
42. Thudi M, Li Y, Jackson SA, May GD, Varshney RK. Current state-of-art of sequencing technologies for plant genomics research. *Briefings in Functional Genomics*. 2012; 11(1):3–11. <https://doi.org/10.1093/bfpg/bfr045> PMID: [22345601](#)
43. Ozdemir Ozgenturk N, Oruç F, Sezerman U, Kuçukural A, Vural Korkut S, Toksoz F, et al. Generation and analysis of expressed sequence tags from *Olea europaea* L. Comparative and functional genomics. 2010; 2010.
44. Armstead I, Huang L, Ravagnani A, Robson P, Ougham H. Bioinformatics in the orphan crops. *Briefings in bioinformatics*. 2009; bfp036.
45. García-López MC, Vidoy I, Jiménez-Ruiz J, Muñoz-Mérida A, Fernández-Ocaña A, de la Rosa R, et al. Genetic changes involved in the juvenile-to-adult transition in the shoot apex of *Olea europaea* L. occur years before the first flowering. *Tree Genetics & Genomes*. 2014; 10(3):585–603.
46. González-Plaza JJ, Ortiz-Martín I, Muñoz-Mérida A, García-López C, Sánchez-Sevilla JF, Luque F, et al. Transcriptomic analysis using olive varieties and breeding progenies identifies candidate genes involved in plant architecture. *Frontiers in Plant Science*. 2016; 7.
47. Schuman MC, Heinzl N, Gaquerel E, Svatos A, Baldwin IT. Polymorphism in jasmonate signaling partially accounts for the variety of volatile produced by *Nicotiana attenuata* plants in a native population. *New Phytologist*. 2009; 183(4):1134–48. <https://doi.org/10.1111/j.1469-8137.2009.02894.x> PMID: [19538549](#)
48. Wu J, Hettenhausen C, Schuman MC, Baldwin IT. A comparison of two *Nicotiana attenuata* accessions reveals large differences in signaling induced by oral secretions of the specialist herbivore *Manduca sexta*. *Plant Physiology*. 2008; 146(3):927–39. <https://doi.org/10.1104/pp.107.114785> PMID: [18218965](#)
49. Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, et al. Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiology*. 2001; 126(2):811–25. <https://doi.org/10.1104/pp.126.2.811> PMID: [11402209](#)



50. Windsor AJ, Reichelt M, Figuth A, Svatoš A, Kroymann J, Kliebenstein DJ, et al. Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). *Phytochemistry*. 2005; 66(11):1321–33. <https://doi.org/10.1016/j.phytochem.2005.04.016> PMID: [15913672](https://pubmed.ncbi.nlm.nih.gov/15913672/)
51. Broekgaarden C, Poelman EH, Steenhuis G, Voorrips RE, Dicke M, Vosman B. Responses of *Brassica oleracea* cultivars to infestation by the aphid *Brevicoryne brassicae*: an ecological and molecular approach. *Plant, cell & environment*. 2008; 31(11):1592–605.
52. Grasso F, Paduano A, Corrado G, Ambrosino ML, Rao R, Sacchi R. DNA diversity in olive (*Olea europaea* L.) and its relationships with fatty acid, biophenol and sensory profiles of extra virgin olive oils. *Food Research International*. 2016;86:121–130.
53. Kliebenstein DJ, Figuth A, Mitchell-Olds T. Genetic architecture of plastic methyl jasmonate responses in *Arabidopsis thaliana*. *Genetics*. 2002; 161(4):1685–96. PMID: [12196411](https://pubmed.ncbi.nlm.nih.gov/12196411/)
54. Sauge MH, Mus F, Lacroze JP, Pascal T, Kervella J, Poësse JL. Genotypic variation in induced resistance and induced susceptibility in the peach-*Myzus persicae* aphid system. *Oikos*. 2006; 113(2):305–13.
55. Walling LL. Avoiding effective defenses: strategies employed by phloem-feeding insects. *Plant Physiology*. 2008; 146(3):859–66. <https://doi.org/10.1104/pp.107.1.13142> PMID: [18316641](https://pubmed.ncbi.nlm.nih.gov/18316641/)
56. Pieterse CM, Dicke M. Plant interactions with microbes and insects: from molecular mechanisms to ecology. *Trends in Plant Science*. 2007; 12(12):564–9. <https://doi.org/10.1016/j.tplants.2007.09.004> PMID: [17997347](https://pubmed.ncbi.nlm.nih.gov/17997347/)
57. Musser RO, Hum-Musser SM, Eichenseer H, Peiffer M, Ervin G, Murphy JB, et al. Herbivory: caterpillar saliva beats plant defences. *Nature*. 2002; 416(6881):599–600. <https://doi.org/10.1038/416599a> PMID: [11948341](https://pubmed.ncbi.nlm.nih.gov/11948341/)
58. Sarmiento RA, Lemos F, Bleeker PM, Schuurink RC, Pallini A, Oliveira MGA, et al. A herbivore that manipulates plant defence. *Ecology letters*. 2011; 14(3):229–36. <https://doi.org/10.1111/j.1461-0248.2010.01575.x> PMID: [21299823](https://pubmed.ncbi.nlm.nih.gov/21299823/)
59. Glas JJ, Alba JM, Simoni S, Villarreal CA, Stoops M, Schimmel BC, et al. Defense suppression benefits herbivores that have a monopoly on their feeding site but can backfire within natural communities. *BMC biology*. 2014; 12(1):98.
60. Tayeh C, Randoux B, Tisserant B, Khong G, Jacques P, Reignault P. Are ineffective defence reactions potential target for induced resistance during the compatible wheat-powdery mildew interaction? *Plant Physiology and Biochemistry*. 2015; 96:9–19. <https://doi.org/10.1016/j.plaphy.2015.07.015> PMID: [26218548](https://pubmed.ncbi.nlm.nih.gov/26218548/)
61. Wu J, Baldwin IT. New insights into plant responses to the attack from insect herbivores. *Annual review of genetics*. 2010; 44:1–24. <https://doi.org/10.1146/annurev-genet-102209-163500> PMID: [20649414](https://pubmed.ncbi.nlm.nih.gov/20649414/)
62. Benitez Y, Botella MA, Trapero A, Alsalmiya M, Caballero JL, Dorado G, et al. Molecular analysis of the interaction between *Olea europaea* and the biotrophic fungus *Spilocaea oleagina*. *Molecular plant pathology*. 2005; 6(4):425–38. <https://doi.org/10.1111/j.1364-3703.2005.00290.x> PMID: [20565668](https://pubmed.ncbi.nlm.nih.gov/20565668/)
63. Zhou S, Lou Y-R, Tzin V, Jander G. Alteration of plant primary metabolism in response to insect herbivory. *Plant physiology*. 2015; 169(3):1488–98. <https://doi.org/10.1104/pp.15.01405> PMID: [26378101](https://pubmed.ncbi.nlm.nih.gov/26378101/)
64. Martinelli F, Remorini D, Saia S, Massai R, Tonutti P. Metabolic profiling of ripe olive fruit in response to moderate water stress. *Scientia Horticulturae*. 2013; 159:52–8.
65. La Camera S, Gouzerh G, Dhondt S, Hoffmann L, Fritig B, LeGrand M, et al. Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. *Immunological reviews*. 2004; 198(1):267–84.
66. Rossi L, Borghi M, Francini A, Lin X, Xie D-Y, Sebastiani L. Salt stress induces differential regulation of the phenylpropanoid pathway in *Olea europaea* cultivars Frantoio (salt-tolerant) and Leccino (salt-sensitive). *Journal of Plant Physiology*. 2016; 204:8–15. <https://doi.org/10.1016/j.jplph.2016.07.014> PMID: [27497740](https://pubmed.ncbi.nlm.nih.gov/27497740/)