

**University of Naples Federico II**



**Department of Agricultural Sciences  
Division of Plant Genetics and Biotechnology**

**“DOTTORATO DI RICERCA IN AZIENDA” POR  
*Campania FSE 2007-2013, Asse IV***

**Ph.D. Scienze Agrarie e Agro-alimentari  
AGROBIOLOGIA e AGROCHIMICA  
Miglioramento Genetico e Orticoltura  
Cycle XXVII  
Academic year 2014-2015**

**GENETIC AND GENOMIC APPROACHES TO  
INVESTIGATE ZYMV RESISTANCE IN *Cucurbita pepo***

**Tutor: Dr. Maria Raffaella Ercolano  
Co-tutor: Dr. Gelsomina Formisano  
Coordinator: Prof. Domenico Carputo**

**Ph.D. Student: Claudio Capuozzo**



## INDEX

<b>SOMMARIO</b>	1
<b>ABSTRACT</b>	5
<b>1. INTRODUCTION</b>	8
1.1 Towards sustainable agriculture	8
1.2 Cucurbita pepo	8
1.3 Breeding for disease resistance in Cucurbita pepo	9
1.3.1 Plant virus resistance mechanism	10
1.4 Molecular marker in breeding for resistance to biotic stress	12
1.5 Sequencing and Next Generation Sequencing	13
1.5.1 RNA-Seq	15
1.6 Aims of the work	16
<b>2. MATERIALS AND METHODS</b>	17
2.1 Plant materials	17
2.2 Resistance to ZYMV in vivo assay in segregating populations	17
2.3 Molecular analysis	18
2.3.1 DNA isolation	18
2.3.2 Genotyping assay	18
2.3.3 CAPS analysis	18
2.3.4 HRM analysis	19
2.4 Statistical analysis goodness of fit	19
2.5 Analysis of genome scaffolds containing ZYMV markers	20
2.6 Transcriptomic analysis	21
2.6.1 RNA isolation, quantity and quality evaluation	21
2.6.2 RNA-Sequencing	21
2.6.3 Functional annotation of reference transcriptome	22
2.6.4 RNA-Sequencing data analysis	22
<b>3. RESULTS</b>	24
3.1 Genetic analysis of resistance to ZYMV	24
3.1.1 Phenotypic evaluation of resistance in segregating populations	24
3.1.2 Molecular screening of parental lines	25
3.1.3 Genotyping of segregating populations	26
3.1.4 Statistical analysis goodness of fit	29

3.1.5 Analysis of genome scaffolds containing ZYMV markers	34
3.2 Transcriptome analysis of zucchini reference genome and of susceptible and resistant lines	36
3.2.1 Reference transcriptome assembling and functional annotation	36
3.2.2 Transcriptome sequencing and mapping of zucchini resistant and susceptible genotypes	38
3.2.3 Identification of Differential Expressed Genes	41
3.2.4 R-genes prediction	43
3.2.5 SNP calling analysis	44
3.2.6 Identification of new candidate markers for ZYMV resistance	46
<b>4. DISCUSSIONS</b>	49
<b>5. CONCLUSIONS</b>	54
<b>6. REFERENCES</b>	56
<b>7. HOST COMPANY</b>	72
<b>8. STAGE</b>	72
<b>9. PUBLICATIONS</b>	72



## SOMMARIO

*Cucurbita pepo* è una specie molto importante sia dal punto di vista economico che sociale ed è coltivata in tutto il mondo (ISTAT, 2007). Questa coltura è soggetta a numerosi attacchi virali. Uno dei virus più dannosi per le colture di zucchini è rappresentato dallo Zucchini Yellow Mosaic Virus (ZYMV). L'identificazione di fonti di resistenze e il loro trasferimento in cultivar commerciali rivestono un ruolo fondamentale nella lotta a questo patogeno. Negli ultimi anni sono state descritte diverse fonti di resistenza a ZYMV (Paris e Brown, 2005) alcune delle quali già impiegate in programmi di miglioramento genetico di zucchini (Provvidenti, 1997; Paris e Cohen, 2000).

È importante individuare la corretta locazione sul genoma dei geni di resistenza caratterizzati, affinché si possano individuare marcatori molecolari ad essi associati da impiegare nelle successive fasi di selezione genetica.

Il principale obiettivo di questo lavoro è l'identificazione di marcatori molecolari associati ai geni noti per conferire resistenza a ZYMV, che potrebbero risultare utili per il miglioramento genetico di varietà di zucchini tra cui la varietà locale 'San Pasquale' tipica della regione Campania. Il secondo obiettivo di questo lavoro è rappresentato dall'individuazione di nuovi geni candidati, e relativi marcatori SNP ad essi associati, coinvolti nel meccanismo di resistenza a ZYMV in *Cucurbita pepo* mediante un approccio trascrittomico. L'ultimo obiettivo è quello di effettuare una predizione genica dei geni di resistenza R nel genoma di zucchini.

In primo luogo, sono state ottenute due popolazioni segreganti F<sub>2</sub> dall'incrocio di varietà di zucchini suscettibili ('True French' e 'San Pasquale') con una varietà resistente al virus ('381e') e tre BC<sub>1</sub> dal successivo reincrocio con la 'San Pasquale'. Le valutazioni fenotipiche dei sintomi effettuate sulle 5 popolazioni sono in accordo con i rapporti di segregazione proposti da Paris e Cohen (2000) per 3 geni complementari dominanti sono candidati per la resistenza a ZYMV.

I dati provenienti dall'uso di una piattaforma SNP (Illumina GoldenGate; Esteras *et al.*, 2012) hanno consentito di valutare la distanza genetica delle linee 'True French' (TFR) '381e' e 'San Pasquale' (SPQ) e di individuare 6 polimorfismi SNP tra le accessioni suscettibili e quelle resistenti al virus. Questi SNP sono stati valutati sui parentali e le relative popolazioni segreganti, già caratterizzate fenotipicamente per la resistenza al virus. Le analisi molecolari HRM, hanno evidenziato tre marcatori SNP putativamente associati ai geni di resistenza (*Zym-1*, *Zym-2* e *Zym-3*) proposti da Paris e Cohen (2000). I risultati dell'analisi di segregazione sono in accordo con il rapporto atteso di 3 resistenti su 1 suscettibile ( $\chi^2$  P=0,8). Tale rapporto è stato dedotto ipotizzando che il gene *Zym-1* sia il gene maggiore rispetto ai geni minori *Zym-2* e *Zym-3*, ipotesi confermata

anche dalla percentuale di co-segregazione calcolata tra il genotipo e il fenotipo di ogni singola pianta e dalla percentuale di co-segregazione tra i singoli marcatori testati ed il fenotipo. Il test del chi-quadrato suggerisce che esiste una forte relazione tra lo SNP1 ed il gene maggiore *Zym-1*. Anche la valutazione molecolare effettuata sulla popolazione F2 'TFRx381e', attraverso un'analisi CAPS, ha confermato l'associazione tra lo SNP1 ed il gene maggiore *Zym-1* ( $\chi^2 P=0,7$ ). Inoltre, il test di correlazione Mantel condotto sui dati genotipici e fenotipici delle due popolazioni segreganti F2 'TFRx381e' e 'SPQx381e' ha mostrato che esiste una forte correlazione tra le due matrici in esame ( $r(AB) = 0,112$ ).

Al fine di identificare nuovi geni candidati e nuovi marcatori SNP associati alla resistenza a ZYMV è stato effettuato un esperimento di sequenziamento del trascrittoma (RNA-seq) di piante inoculate del genotipo suscettibile 'TFR' e del genotipo resistente '381e'. L'analisi dei geni differenzialmente espressi (DEG) ha evidenziato 206 geni comuni nelle due linee isogeniche analizzate, 1956 espressi esclusivamente nel genotipo suscettibile e 1006 trascritti nel genotipo resistente a ZYMV.

Al fine di ottenere una classificazione in categorie funzionali 'Gene Ontology' (GO) è stata effettuata l'annotazione funzionale tramite il software Blast2GO (B2GO; Conesa *et al.*, 2005) del trascrittoma di riferimento. L'analisi ha rivelato che il 36,54 % dei trascritti è ascrivito in almeno una delle tre categorie funzionali GO (molecular function, biological process e cellular component). Sulla base di questa analisi è stato possibile annotare i trascritti differenzialmente espressi nei due genotipi. L'annotazione funzionale dei DEG nei genotipi analizzati ha permesso di identificare, una relativa abbondanza di trascritti che codificano per domini LRR e Serina/Treonina chinasi in entrambi i genotipi, mentre una più alta concentrazione di trascritti codificanti Elicasi ATP-dipendenti è stata riscontrata nel genotipo resistente.

I dati provenienti dal sequenziamento del trascrittoma di 'TFR' e '381e' sono stati impiegati anche per effettuare un'analisi di SNP calling nei confronti del genoma di riferimento. Tale analisi ha permesso di evidenziare 3357 e 14327 SNP unici caratterizzanti rispettivamente 'TFR' e '381e'. Gli SNP1 e SNP2, putativamente correlati al gene *Zym-1* e *Zym-2*, sono stati identificati tra i 14327 SNPs del genotipo '381e', mentre lo SNP3 nei 3357 SNP unici del genotipo 'TFR'.

Dal confronto delle liste dei geni contenenti gli SNP e le liste dei DEG è stato possibile identificare i geni regolati in modo positivo o negativo, espressi nei due genotipi in relazione all'attacco del virus e polimorfici rispetto al genoma di riferimento. Tra i geni differenzialmente espressi e polimorfici è stata evidenziata un'alta percentuale di geni codificanti per 'Protein Kinase', 'Chloroplast Protein Kinase' e 'Leucine Rich Repeats' in entrambe i genotipi oltre ad una rilevante quantità di altre proteine come Taumatina, Ubiquitina, Fosfatasi, ed ormoni quali Auxine ed Etilene, tipicamente coinvolti nelle reazioni di risposta della pianta a stress biotici.

Al fine di identificare gli SNP unici dei due genotipi correlati a geni di resistenza, è stato interrogato il proteoma di *Cucurbita pepo* con un predittore specifico per geni di resistenza chiamato MATRIX R (Hermoso Pulido *et al.*, 2014). I risultati hanno evidenziato che dei 3357 e 14327 SNP unici dei genotipi 'TFR' and '381e', 251 SNP e 1895 SNP, rispettivamente, ricadono in geni o domini tipici coinvolti in meccanismi di difesa della pianta. In particolare, 100 SNP del genotipo '381e' ricadono in soli 3 geni R localizzati sullo scaffold2494 in cui è presente anche lo SNP1 associato al gene maggiore *Zym-1*, così come 19 dei 251 SNP identificati nel genotipo 'TFR' ricadono su in 3 geni R dello scaffold1174 in cui risiede lo SNP3 associato al gene minore *Zym-3*.

Concludendo, lo studio dell'ereditarietà della resistenza a ZYMV attraverso le analisi fenotipiche e genotipiche, condotte su cinque popolazioni segreganti, hanno messo in evidenza il coinvolgimento di un gene principale e di due geni minori nei meccanismi di resistenza a tale virus. I saggi *in vivo* e le analisi molecolari hanno permesso di mostrare una forte relazione tra i tre marcatori SNP identificati e questi geni di resistenza, consentendo di avere uno strumento molecolare utile per una selezione più mirata. Le analisi trascrittomiche condotte su due linee isogeniche 'True French' e '381e', hanno permesso di identificare geni R polimorfici tra il genotipo resistente e quello suscettibile espressi in modo diverso durante la risposta di resistenza a ZYMV. Inoltre, la predizione dei geni R ha fornito ulteriori geni candidati potenzialmente implicati nella resistenza a ZYMV. La differente espressione genica, evidenziata in questo lavoro tra le due linee isogeniche, sarà utile per costruire un modello esplicativo sul processo di risposta dello zucchini a ZYMV. Infine, i risultati ottenuti saranno molto utili per velocizzare la selezione di cultivar locali di zucchini resistenti a ZYMV.





## ABSTRACT

Cucurbits species are economically important crops grown worldwide. In many regions, the most serious threat to their production is Zucchini yellow mosaic virus (ZYMV). A monogenetic source of resistance to ZYMV was identified in *Cucurbita moschata* and introgressed in a *Cucurbita pepo*. However, in cultivated zucchini the major dominant gene indicated as *Zym-1* interact with other genes, making difficult the breeding transfer. Until now, not effective markers associated to ZYMV resistance have been found in *C. pepo*.

Main goal of this work was to investigate the mode of inherence of ZYMV resistance in zucchini and to identify molecular markers associated with resistance genes to ZYMV, in order to make easier, faster and economically sustainable the introgression of this resistance in zucchini local varieties 'San Pasquale'. Moreover, a genome-wide transcriptional analysis of a resistant and a susceptible genotype to ZYMV was conducted in order to identify new candidate genes for resistance to ZYMV and to catalogue SNP polymorphic markers.

Two F<sub>2</sub> and three BC<sub>1</sub> segregating populations, obtained from the cross of '381e' (resistant variety) with 'True French' and 'San Pasquale' (susceptible varieties) were assayed for ZYMV resistance. Phenotypic analysis highlighted a complex inheritance of resistance with the involvement at least of three genes (*Zym-1*, *Zym-2* and *Zym-3*). Molecular segregation analysis conducted using High Resolution Melt (HRM) technology on 3 SNP loci strongly suggests a relationship of these SNPs with resistance genes *Zym-1*, *Zym-2* and *Zym-3*. Moreover, nearly 3000 differentially expressed genes between 'True French' and '381e' isogenic lines for resistance to ZYMV were identified in transcriptome of susceptible and resistant varieties inoculated with the virus. A high number of SNPs between the two varieties were also evidenced. A close co-localization of SNP1 with a CNL and an ATP-dependent RNA helicase, suggests that one of those genes could be a good candidate for exerting resistance function to ZYMV.

The study of ZYMV resistance response in zucchini at genetic and genomic level, performed in this work, can be useful for the genetic improvement of local variety as the susceptible 'San Pasquale' that have a good marketability in Campania region.





## 1. INTRODUCTION

### 1.1 Towards sustainable agriculture

Intensive agriculture practices have dramatically changed the 'Earth' ecosystem. In order to achieve maximum production with minimum efforts the modern agricultural system is based on high crop specialization, motorization and mechanization. The massive use of pesticides involves enormous damages to ecosystems leading to a gradual contamination of the soil and the water, the same that is used for irrigation, inducing resistance reactions by pests and destroying beneficial organisms. Therefore, a more sustainable agriculture system, to apply both in developed and in developing countries, is required to meet "the needs of the present without compromising the ability of future generations to meet their own needs". Sustainable agriculture should maintain the adaptive capacity of farming systems (Hansen, 1996; Gold, 1994; Hansen and Jones, 1996; Park and Seaton, 1996). Agriculture is sustainable when it is ecologically and, economically viable, socially just, culturally appropriate, and based on a holistic scientific approach" (Madden *et al.*, 1997). Although no one set of farming practices were established for sustainable agriculture, the cultivation of disease resistant varieties can meet the sustainable agriculture purpose (Horrihan *et al.*, 2002).

### 1.2 *Cucurbita pepo*

Morphologically, *Cucurbita pepo* is the most polymorphic species with respect to fruit characteristics (Duchesne, 1786; Naudin, 1856). The edible forms of this species can be grouped into eight morphotypes: Pumpkin, with round or nearly round fruits; Vegetable Marrow, with short, tapered and cylindrical-shaped fruits; Cocozelle, with long bulbous fruits; Zucchini, with uniformly cylindrical shaped fruits; Acorn, with furrowed, turbinate fruits; Scallop, with flat, scalloped fruits; Crookneck, with narrow, usually curved and warty, necked fruits and Straightneck, with short-necked or constricted fruits, usually warty (Paris, 1986). The great variability of types represented in the Botanical works of the 16<sup>th</sup> and 17<sup>th</sup> centuries, and the observations of the first European explorers in America, seem to indicate that Pumpkins, Scallops and Acorns were developed under the guidance of Native Americans in pre-Columbian times (Paris, 2001; Smith, 1997; Whitaker, 1947). The subsequent arrival in Europe of *C. pepo* fruits and seeds generated an extraordinary variability of new phenotypes through hybridization and recombination. New cultivars were developed, particularly the elongated forms of *ssp. pepo*: Vegetable Marrow, Zucchini and Cocozelle.

The great importance of the zucchini cultivation is due to the use of fruits for the greatly appreciated organoleptic characteristics that are a rich source of nutrients such as vitamins (vitamin A, C, folate) and minerals (potassium). About 24 million tons of zucchini (also reported as squash), together with pumpkins and gourds, were produced in the world in 2012 for a cultivated area of 1.8 million of hectares. China is the largest producer followed by India and Russian Federation. Italy, with around 520,000 tons, is the ninth producer country (FAOSTAT 2012; <http://faostat.fao.org>). In Italy, zucchini are among the first twenty commodities for production, covering more than 50% of European production, and is particularly developed in south regions (ISTAT 2007). The major production centres are located in Southern Italy (Sicily 2,920 ha; Lazio 1,338 ha; Puglia 1,510 ha; Calabria 1,342 ha; Campania 793 ha), but also in the North (Piemonte has 1,401, the Veneto has 1,027 and Emilia Romagna 116 ha) (ISTAT 2007).

### **1.3 Breeding for disease resistance in *Cucurbita pepo***

In the last century, breeders have made great efforts in zucchini breeding in order to develop new and improved cultivars and hybrids, nowadays among the main objectives of the genetic breeding there is also the recovery of local landraces. Breeding methods are based on crosses followed by selection pedigrees and backcross (Allard, 1960) and can be distinguish in pedigree and backcross methods. The first one, pedigree method, involves self-fertilization and selection for characters of interest in each generation and is commonly used when both of parental plants have positive characteristics. The backcross provides the selection of a parental showing a character of interest and then it is crossed with plants missing of that character and is commonly used to obtain plants resistant to several diseases. Disease resistance is considered the most important goal 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 insects attack. Among fungi, powdery mildew, caused by *Podosphaera xanthii* or by the association of *Erysiphe chicoracearum* and *Sphaerotheca fuliginea*, is a serious disease affecting the leaves, stems and fruits of cucumber and zucchini squash grown in greenhouses and in open field. On the other hand, insects, in particular *Aphis gossypii*, represents one of the major pests of cucurbits and ornamental plants inducing leaves deformation and stunted growth. In addition, insects represent a serious problem affecting *Cucurbitaceae* because they are responsible of the transmission of viruses, especially the Zucchini Yellow Mosaic Virus (ZYMV).

The *potyvirus* ZYMV is one of the most destructive and widespread viral pathogens on cucurbits worldwide (Zellning *et al.*, 2014). It was first described in 1973 in northern Italy (Lisa *et al.*, 1981).

ZYMV infection dramatically alters normal leaf and fruit shape, size and colour. Leaf symptoms include veinal chlorosis, vein banding, blistering, and strapleaf. Fruits display pronounced chlorotic bumps, malformations, and stunting. Mixed field infections of cucurbit species with ZYMV, PRSV, WMV and Cucumber mosaic virus (CMV) were reported to be common in cucurbit species and cause considerable damage worldwide and severe epidemics in cucurbit fields (Grafton-Cardwell *et al.*, 1996; Lecoq *et al.*, 1981). A study of ZYMV synergism was performed on the interaction of ZYMV with CMV, as these viruses are commonly found together in nature and can be transmitted by the same aphid (Lecoq *et al.*, 1981) resulting in some cases in plants death (Wang *et al.*, 2002). A complete list of resistance genes involved in breeding programs has been identified in the Cucurbita species and published in 2005 by H. S. Paris. There are about 66 genes reported for *C. pepo*, 21 for *C. moschata*, 19 for *C. maxima* genes and 29 for the interspecific cross of *C. maxima*-*C. ecuadorensis* (25 out of these genes are enzymatic isoforms). One or two genes were also identified in 4 wild species (*C. okechobeensis*, *C. lundelliana*, *C. ecuadorensis* and *C. fetidissima* HBK) and in several other interspecific crosses. Furthermore, a low number of resistance genes was introgressed in commercial zucchini cultivars from genotypes of related species, such as *C. moschata* Duchesne, and wild relatives, such as *C. okechobeensis* (small) Bailey. These species represent an existing source of genetic variation to incorporate required traits, but their use is limited by cross-ability barriers, including reduced fertility and quasi-linkage.

### 1.3.1 Plant virus resistance mechanism

Plant viral diseases cause serious economic losses reducing yield and quality of crops and, although viruses are relatively simple genetic entities, much remain unknown about the mechanisms by which the many symptoms of disease are generated, and by which plants resist these effects.

The study of plant resistance raises many fundamental questions regarding the molecular, biochemical, cellular, and physiological mechanisms involved in the plant-virus interaction and the co-evolution. The first step in the study of genetics of viral resistance is to determine if the resistant response is inherited, and the number of genes involved. More than 80% viral resistance reported is monogenically controlled (the remaining shows oligogenic or polygenic control) (Kang *et al.*, 2005). Only slightly more than half of all reported monogenic resistance traits show dominant inheritance. In most but not all cases, dominance has been reported as complete and dominant resistance is often associated with the hyper-sensitive response (HR) (Fraser, 1986).

The majority of characterized viral resistance genes from plants fall into the NBS-LRR class of resistance genes, such as *HRT*, *RTM1*, *RTM2*, *RCY1*, from *Arabidopsis*, and from *solanaceous*

hosts, *N*, *Rx1*, *Rx2*, *Sw5*, and *Tm-22*, all sharing structural similarity (Kang *et al.*, 2005). However, the strong structural similarity does not ensure the same spectrum of action. *A. thaliana* *RCY1* (resistance to C strain Y1) and *HRT* (HR to turnip crinkle virus) encode proteins that share 91% similarity (Takahashi *et al.*, 2002), but confer resistance to unrelated viruses such as cucumber mosaic virus (CMV, a cucumovirus) and turnip crinkle virus (TCV, a carmovirus), respectively (Soosaar *et al.*, 2005). In cultivated tomato, ToMV (Tomato mosaic virus) infections are controlled by the introgressed *Tm-1*, *Tm-2* and *Tm-22* genes. The *Tm-22* resistance gene was shown to be durable (Hall *et al.*, 1980; Pelham, 1966) and it has been cloned and well characterized (Lanfermeijer *et al.*, 2003). Susceptible tomato plants transformed with the *Tm-22* gene displayed resistance against ToMV infection and the resistance was conserved in all transgenic lines. In Citrus, it has been reported the resistant gene dominant gene *Ctv-R* that conferring broad-spectrum resistance against Citrus tristeza virus (CTV), one of the major pathogen of citrus (Rai *et al.*, 2006). Seo and collaborators (2006) identified the TIR-NBS-LRR gene *RT4-4* involved in a viral resistance response in common bean (*Phaseolus vulgaris* cv. Othello). The functional analysis revealed that the *RT4-4* gene in transgenic *N. benthamiana* lines is up-regulated in a non-virus-specific manner, although *RT4-4* did not confer resistance to the reporter virus, it activated a resistance-like response (systemic necrosis) to Cucumber mosaic virus (CMV) (Seo *et al.*, 2006). Studies on potyviruses resistance genes led to the identification of resistance genes corresponding to mutations in translation initiation factors, eukaryotic initiation factors 4E (eIF4E) and to a lesser extent, the eukaryotic initiation factor 4G (eIF4G) (Piron *et al.*, 2010). The eIF4E gene mutation providing resistance to several *Potyviridae* family viruses has been identified in pepper (*pvr1*), pea (*sbm1*), lettuce (*mo1 (1)*, *mo1 (2)*), tomato (*pot1*), melon (*nsv*) and barley (*rym4/5*) (Robaglia *et al.*, 2006; Ruffel *et al.*, 2005). The modification in eIF4G is responsible for resistance of rice to Yellow mottle virus (Albar *et al.*, 2006) and in Arabidopsis to Cucumber mosaic virus and Turnip crinkle virus (Yoshii *et al.*, 2004).

Moreover, there are several examples of dominant and recessive genes that seem to control a wide range of viral genotypes. The most representative examples involve members of the *Potyviridae*. The *I* gene of common bean (Ali, 1950) was first reported to control a dominant resistance to ten different related potyviruses, Azuki mosaic virus, Bean common mosaic virus (BCMV), Bean necrotic mosaic virus, Blackeye cowpea mosaic virus, Cowpea aphid-borne mosaic virus, Passionfruit woodiness virus-K, Soybean mosaic virus (SMV), Thailand Passiflora virus, Watermelon mosaic virus (WMV), and Zucchini Yellow Mosaic Virus. As well in *Cucumis melo* a dominant gene (*Ag*) confer resistance to CMV transmitted by *Aphis gossypii* (Pitrat and Lecoq, 1980).



Genetic resistance is considered the most cost-effective and reliable approach to protect crops from virus infection. Several resistance genes have been tagged with molecular markers including RFLP, AFLP, RAPD, and various other PCR-based markers and these molecular approaches has been used used for indirect selection via genotype, for localizing and isolating R genes in plant genomes.

#### **1.4 Molecular marker in breeding for resistance to biotic stress**

The construction of molecular maps and marker-assisted selection can aid the introgression of resistance to pathogens. Marker-assisted selection (MAS) provides opportunities for enhancing the response from selection because molecular markers can be applied at the seedling stage, with high precision and reductions in cost. The use markers associated to genes conferring monogenic resistances and quantitative trait loci (QTL) for quantitative disease resistances have been reported in several crop (St. Clair, 2010).

Traditional DNA-based methods include PCR-based methods and hybridization-based methods, i.e. restriction fragment length polymorphisms (RFLPs). PCR-based methods are extremely sensitive, often faster than other technologies, and are widely used in agriculture and food control (Grassi *et al.*, 2006; Labra *et al.*, 2004; Mane *et al.*, 2006; Pasqualone *et al.*, 2007). Random amplification polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), inter simple sequence repeats (ISSRs), microsatellites or simple sequence repeats (SSRs), cleaved amplified polymorphic sequence (CAPS) and single nucleotide polymorphisms (SNPs). CAPS proves useful for genotyping, positional or map based cloning and molecular identification studies (Konieczny and Ausubel, 1993; Komori and Nitta, 2005; Spaniolas *et al.*, 2006; Weiland and Yu, 2003). Recently, SNP markers have become the most used markers in genetic characterization studies as well as in translational genomic (Varshney *et al.*, 2014).

Single nucleotide polymorphisms are the most abundant forms of genetic variation among individuals of a species. SNPs are bi-allelic and allow an easy comparison between results collected from different laboratories using different chemicals and technologies. SNP markers were widely investigated in several crop. In corn, the frequency of polymorphism found in inbred lines is 1 SNP every 31 bp in the non-coding regions, and 1 SNP every 124 bp in the coding regions (Ching *et al.*, 2002). In soybeans, genetic diversity studies in 22 different genotypes have detected the presence of 1.64 SNPs every 1000 bp in the coding regions and 4.85 SNPs every 1000 bp in the non-coding regions (Zhu *et al.*, 2003). A research on barley from Kanazin *et al.* (2002) reports a frequency of 1 every 189 bp mutation. In Zucchini, Blanca and colleagues performed the first study based on SNP polymorphisms in 2011 identifying 9,043 high confidence SNPs between subsp. *pepo* cv. Zucchini

and subsp. *ovifera* cv Scallop, 4 SNPs per 100 bases in the coding regions. The popularity of SNPs as valuable and efficient molecular markers has increased the need to have more accurate SNPs detection methods. High resolution melting (HRM) analysis has been developed to detect SNPs in small PCR amplicons because it is an easy, require low-time (Hoffmann *et al.*, 2007; Hung *et al.*, 2008; Liew *et al.*, 2004) and discriminates homozygous and heterozygous (Bennett *et al.*, 2003; Herrmann *et al.*, 2007; Stephens *et al.*, 2008). The evolution of the HRM techniques from traditional melting curve analysis was implemented by the invention of new generation intercalating dyes (SYBR green I dye) and new instrumentation (Slinger *et al.*, 2007). Melting curve analysis using fluorescent dyes has proven to be a highly sensitive method for mutation discovery and SNP genotyping (Liew *et al.*, 2004; Cho *et al.*, 2008; Smith *et al.*, 2008).

The development of high-throughput methodologies allows the detection of many thousands of SNPs through genotyping platforms, facilitating the construction of high-density genetic maps (Gupta *et al.*, 2008). Different SNPs genotyping technologies are now available: TaqMan® probes (Livak *et al.*, 1995), FP-TDI (Chen *et al.*, 1999), Invader (Olivier 2005), IPLEX (Wright *et al.*, 2008), Kaspar (Nijman *et al.*, 2008), SNaPshot (Applied Biosystems; Sanchez *et al.*, 2003), GoldenGate (Illumina, San Diego, USA; Fan *et al.*, 2003) and Infinium (Illumina, San Diego, USA; Gunderson 2009). These genotyping techniques have been used extensively in humans (International HapMap Consortium, 2003), in several animal species (McKay *et al.*, 2008; Kijas *et al.*, 2009; Malhi *et al.*, 2011) and are also available for plant species (Pavy *et al.*, 2008; Hyten *et al.*, 2008; Akhunov *et al.*, 2009; Close *et al.*, 2009; Muchero *et al.*, 2009; Yan *et al.*, 2009; Deulvot *et al.*, 2010; Lepoittevin *et al.*, 2010; Massman *et al.*, 2011). The development of these new methods for the analysis of hundreds or thousands of SNPs at the same time allowed the possibility to build a dense SNP-genetic map in *Cucurbita pepo* (Esteras *et al.*, 2012).

## 1.5 Sequencing and Next Generation Sequencing

In 1977, Maxam and Gilbert first reported an approach for DNA sequencing in which terminally labeled DNA fragments were subjected to base-specific chemical cleavage and the reaction products were separated by gel electrophoresis (Maxam and Gilbert, 1977).

In the last 30 years, the sequencing Sanger method, based on chain-terminating dideoxynucleotide analogs that caused base-specific termination of primed DNA synthesis (Sanger *et al.*, 1977), had been the main approach to improve knowledge about genome sequences and, among the others, the conclusion of the first revision of the human genome in 2004 (International Human Genome Consortium, 2004).

From 2005, the use of massively parallel pyrosequencing platforms led the scientific community in the new era of high-throughput genomic analysis (Next Generation Sequencing-NGS). Starting about ten years ago, several instruments for sequencing were produced, such as Genome Sequencer from Roche 454 Life Sciences ([www.454.com](http://www.454.com)), the Solexa Genome Analyzer from Illumina ([www.illumina.com](http://www.illumina.com)), the SOLiD System from Applied Biosystems ([www.appliedbiosystems.com](http://www.appliedbiosystems.com)), the Heliscope from Helicos ([www.helicos.com](http://www.helicos.com)) and the commercialized Polonator ([www.polonator.org](http://www.polonator.org)). All these machines are able to give higher yield with simpler process and to produce sequences in a cost-efficient way. Although each instrument is different from the others, all massively parallel sequences machines share, more or less, the same way of action (Mardis, 2011). First, the amplification of the library fragments by PCR, involving multiple primer pairs. Following the PCR, the resulting fragments have platform-specific adapters ligated to their ends to form a library that is ready for sequencing. After sequencing, the followed step is the computationally conversion of image data into sequence reads, the so-called “base calling”, to generate read sequences and to highlight the error probability–related quality scores for each called base. The quality values calculated during the sequencing is related to the historically relevant Phred score, introduced in 1998 for Sanger sequence data (Ewing *et al.*, 1998a; Ewing *et al.*, 1998b). The Phred score quality value,  $q$ , uses a mathematical scale to convert the estimated probability of an incorrect call,  $e$ , to a log scale:  $q = -10 * \log_{10}(e)$ . Miscall probabilities of 0.1 (10%), 0.01 (1%), and 0.001 (0.1%) yield Phred scores of 10, 20, and 30, respectively. This score resulting to be crucial for eliminating low quality reads and bases, improving alignment accuracy and determining consensus-sequence and variant calls (Li *et al.*, 2008).

Finally, the obtained reads are used to make an assembly, or a hierarchical data structure, that maps the sequence data in order to reach a putative reconstruction of the target. Reads are grouped in contigs (provide a multiple sequence alignment of reads plus the consensus sequence) and contigs into scaffolds (called also supercontigs or metacontigs, define the contig order and orientation and the sizes of the gaps between contigs).

To date, several genomes are sequenced *de novo* or resequenced (Wheeler *et al.*, 2008; Margulies *et al.*, 2005; Pearson *et al.*, 2007; Smith *et al.*, 2008; Quinn *et al.*, 2008; Satkoski *et al.*, 2008; Borneman *et al.*, 2008; Wang *et al.*, 2008) and, moreover, NGS platforms have been used to detect new structural variation, such as deletions, insertions and rearrangements (Korbel *et al.*, 2007; Campbell *et al.*, 2008; Kim *et al.*, 2008; Chen *et al.*, 2008).

### 1.5.1 RNA-Seq

One of the most popular NGS approaches is the RNA sequencing (RNA-seq) employed to map and quantify transcripts in biological samples. This technology include the analysis of a population of RNA where each molecule, with or without amplification, is sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). The alignment of reads is the consequential step, if a reference genome is available, or *de novo* assembling of transcriptome, a transcription map that consists of both the transcriptional structure and level of expression for each gene is required (Wang *et al.*, 2009).

By counting reads that map on sequenced exons (assembled in transcripts corresponding to genes) is possible to obtain the expression levels of each transcripts. The quantity of mapped reads per transcript must be normalized in order to compare the expression levels of transcripts within the same sample and between different samples. The most used normalization methods are RPKM/FPKM (reads or paired-end fragments per kilobase of exon model per million mapped reads) (Mortazavi *et al.*, 2008), TMM (trimmed mean of M-values) (Robinson *et al.*, 2010) and CPM (count per million) (Robinson *et al.*, 2010).

The biggest revolution mediated by the sequencing of RNA is the ability to give an accurate view about gene expression levels and, on the other side, the possibility to detect and study the transcriptome for non-model organisms that have no genomic sequences determined (Vera *et al.*, 2008; Marioni *et al.*, 2008). Moreover, the RNA-seq strategy can also reveal sequence variations, such as SNPs, isoform and novel splice junctions, and information of antisense regulation and intragenic expression (Cloonan *et al.*, 2008; Morin *et al.*, 2008; Carninci *et al.*, 2005; Nagalakshmi *et al.*, 2008; Graveley *et al.*, 2011; Trapnell *et al.*, 2010).

Although the RNA sequencing is a relatively new approach, it has been already applied to several crop plants (Bagnaresi *et al.*, 2012; Gao *et al.*, 2013; Mesarich *et al.*, 2014).

## **1.6 Aims of the work**

Main goal of this work was to investigate the mode of inheritance of ZYMV resistance in zucchini and to identify molecular markers associated with resistance genes to ZYMV, in order to make easier, faster and economically sustainable the introgression of this resistance in local varieties of zucchini, as 'San Pasquale'. Second aim of our work was to conduct a genome-wide transcriptional analysis of a resistant and a susceptible genotype to ZYMV, in order to identify new candidate genes for resistance to ZYMV, to perform a R genes prediction and to catalogue new SNP markers linked to the genes conferring ZYMV resistance. Results obtained constitute an important compendium of methods and information that can improve selection procedures employed in *Cucurbita pepo* breeding program.

## **2. MATERIALS AND METHODS**

### **2.1 Plant materials**

Four zucchini varieties (Nano Verde di Milano, True French, 381e and 968Rb), four Coccozelle varieties (Romanesco, Ortolana Di Faenza, San Pasquale and Bianca di Trieste) and one Pumpkin accession (Tondo Chiaro di Nizza) were used for genotyping assay. The susceptible open-pollinated zucchini ‘True French’ (TFR) (Thompson & Morgan, Ipswich, England, 1978) was crossed with its near-isogenic resistant counterpart, Accession ‘381e’ indicated from hereafter as ‘381e’ to obtain a F<sub>2</sub> populations. The latter variety had been derived from six generations of backcrossing to ‘True French’, selecting for resistance, followed by four successive generations of self-pollination and selection for resistance (Paris and Cohen, 2000). ‘San Pasquale’ referred hereafter as ‘SPQ’ obtained from La Semiorto Sementi, Sarno, Italy, is an open-pollinated cocozelle cultivar having extremely long, striped fruits and susceptible to ZYMV, was also crossed with ‘381e’ to produce F<sub>2</sub> populations. Three BC<sub>1</sub>-F<sub>2</sub> populations were obtained from the cross of different F<sub>2</sub> ‘SPQx381e’ individuals (plants 64, 47 and 28) with the recurrent parent ‘SPQ’.

### **2.2 Resistance to ZYMV *in vivo* assay in segregating populations**

Seeds of parental and filial generations were sown in multi-cellular trays consisting of 4 cm diameter pots that had been filled with peat, one per pot. After the seedlings were transferred into pots 15 cm diameter and were grown under glasshouse condition at 22-24°C using supplement lighting to maintain 12 photoperiod at the Research Centre for Plant Pathology (CRA-PAV) in Rome. Each individual plant was assigned an identification number.

An isolate of ZYMV from a naturally infected plant of field-grown summer squash was used for experiments. This isolate cause the typical symptoms of ZYMV disease including yellow mosaic, vein banding, blistering and malformation of leaves. Symptomatic leaves of artificially infected zucchini plants were crushed, and the raw juice was extracted at a ratio 1:10 w/v in 0.1 M phosphate buffer pH 7.2. Both cotyledons of test plants, sprinkled with the abrasive powder “celite”, were inoculated with approximately 20 µl of the diluted extract, and subsequently washed with distilled water. Symptoms were observed and recorded from 6 to 30 days after inoculation (dpi), when the plants had developed 6–7 true leaves. Plants each of the three parents as well as F<sub>1</sub> plants were used as resistant, susceptible and heterozygous controls. According to the symptoms observed after 12 days after inoculation and based on what reported by Paris and Cohen (2000), the plants were classified in three groups. Group A, asymptomatic plants or with few yellow spots on new leaf.

Group B, plants with yellow spots spread on new entire leaf and with symptoms of "vein banding" after 12 dpi and then recovery after 25 dpi. Group C, plants with symptoms of "vein banding" and mottle on leaves or clearly susceptible plants with leaf malformation.

To ensure that the asymptomatic plants were really resistant and not simply escaped by inoculation of the virus, a serological ELISA test was performed. The presence of the virus into infected plants would prove by an intense serological reaction (yellow) compared to healthy controls.

## **2.3 Molecular analysis**

### **2.3.1 DNA isolation**

In the young leaves, using the protocol described by Fulton *et al.* (1995), was isolated total genomic DNA of parental and segregating population samples. Each sample was represented by a bulk of three leaves. The DNA amount was quantified spectrophotometer by NanoDrop ND-1000 Spectrophotometer (NanoDropTechnologies) and the quality was checked electrophoretically with a 1.2% agrose gel staining with a SYBR® Safe DNA Gel Stain (Takara). Samples were then diluted to a 15-25 ng/µl working concentration and stored at -20°C.

### **2.3.2 Genotyping assay**

An Illumina GoldenGate genotyping assay, based on a 384 SNPs panel, was performed on nine varieties of *Cucurbita pepo* ssp *pepo*. These varieties include representatives of three morphotypes of the subspecies *pepo* such as Zucchini varieties (Nano Verde di Milano and three isogenic lines True French, 381e and 968Rb); Coccozelle varieties (Romanesco, Ortolana Di Faenza, San Pasquale and Bianca di Trieste) and Pumpkin accession (Tondo Chiaro di Nizza). The genotyping was performed through a Principal Component Analysis (PCA) obtained using GenAIEx 6.41 and MEGA5 software.

### **2.3.3 CAPS analysis**

The F<sub>2</sub> segregating population obtained from 'True French' and '381e' was tested with a CAPS marker designed thanks to the GoldenGate platform data. The CAPS analysis consisted of a PCR reaction and subsequent digestion with BglIII restriction enzyme that recognizes A<sup>^</sup>GATCT sites. PCR amplification was carried out in a final volume of 25 µL, Buffer 1×, 0,1 mM dNTPs, 0,2 µM, Taq polymerase 0,5 U, and 0,2 µM each primer, DNA 25-35 ng. Amplification was performed

under the following program: 30 s at 95° C; 40 cycles of 10 s at 95° C, 30 s at 57° C, and 30 s at 72° C; and extra extension for 1 min and 30 s at 72° C. 15 µL of amplified products was treated for 3 hours at 37° C with 1 U BglIII restriction enzyme in a final volume of 25 µL. Polymorphism was detected by separating the whole volume of treated DNA on 2% agarose gel (1× Tris-Acetate EDTA) containing SYBR® Safe DNA Gel Stain (Takara), and visualizing it under a UV transilluminator. We used 1Kb+ (Invitrogen) fragments as size markers.

#### **2.3.4 HRM analysis**

Real-Time PCR mixture was prepared using 15-25 ng of DNA, 0.6 U of Takara ex Taq™ (Takara), 20 mM of Mg<sup>2+</sup>, 2.5 µl of 10X ex Taq Buffer 10 mM, 2.5 mM dNTPS and primers were mixed to the final concentration of 10 uM. The intercalating dye used was the EvaGreen (Biothium) at final concentration of 0.5X. Real-Time PCR reactions and HRM analysis were performed on Rotor-Gene 6000™ (CorbettResearch, Cybeles, Thailand). The Real-Time PCR profile comprised of one initial cycle of 30 s at 95°C and followed by 40 cycles of 10 s at 95°C, 30 s at 58-60°C, 30 s at 72°C. After Real-Time PCR amplification, HRM was performed using melting profile from 70° to 90°C rising at 0.1°C per sec. The software provided by the company normalized the melting curves. The specificity of primers used in the experiment was confirmed by a single amplicon melting analysis. Data from HRM analysis was determined using Rotor-Gene 6000 series 1.7 software. All of the analyzed data was achieved according to the manufacturer's protocol (HRM guide [www.appliedbiosystems.com](http://www.appliedbiosystems.com)).

#### **2.4 Statistical analysis goodness of fit**

ZYMV resistance phenotypic segregation of F<sub>2</sub> 'TFRx381e' and F<sub>2</sub> 'SPQx381e' population was analyzed by chi-square test to establish the degree of coincidence between the expected and the obtained segregation ratio 45:19. The results from HRM analysis for all molecular markers were subjected to 3:1  $\chi^2$  analysis, too. Moreover, the segregation into resistant and susceptible genotypes was also calculated as the frequency of resistant plants  $f(R)$  – number of resistant plants/number of tested plants per combination. In the two BC<sub>1</sub>-F<sub>2</sub> populations the expected ratio of resistant/susceptible was 1:1 both in phenotypic and genotypic evaluations.

Furthermore, a Mantel test (Mantel, 1967) was executed between phenotypic and genotypic data to find a correlation between the two trends. In order to obtain the phenotypic and genotypic matrices was performed the Ascendant Hierarchical Coefficient (HAC) by the XLStat tool. MEGA5 tool was



used to perform, on the obtained matrices, the dendrograms based on the Euclidean distance between the distributions. The Pearson's correlation coefficient was evaluated on the two matrices through a correlation/association test (Mantel test).

## **2.5 Analysis of genome scaffolds containing ZYMV markers**

In order to identify genes physically closer to our markers, a functional prediction of protein comprising domains involved in the defense mechanisms was performed. Firstly, the three unigenes extracted from the GoldenGate panel have been searched on the genome and obtaining the scaffolds containing the unigenes. GenScan tool was used to translate the scaffolds sequences in the peptides in order to verify, via HMM profiling using InterProScan software version 4.8 (Zdobnov and Apweiler, 2001), the presence of conserved domains and motifs characteristic of proteins (NBS; LRR; TIR; KINASE; SERINE/THREONINE). In this analysis, recovered sequences were compared with the following databases: Hidden Markov Model Panther (HMMPanther), Hidden Markov Model Tigr (HMMTigr), patternScan, FPrintScan, HMMPfam, ProfileScan, High-Quality Automated and Manual Annotation of Microbial Proteomes (HAMAP), SignalPHMM PROSITE, SuperFamily PRINTS (Fingerprint database), HMMPfam (Protein family; <http://pfam.wustl.edu/hmmsearch.shtml>), Blast Protein Domain database (BlastProDom), and Hidden Markov Model Simple Modular Architecture Research Tool (HMMSMART) protein motif analyses (<http://smart.embl-heidelberg.de/>; Schultz *et al.*, 1998). The TMHMM database was also accessed to verify the presence of transmembrane regions of the candidate proteins belonging to the class of transmembrane receptor (RLK and RLP proteins). The coils program (Lupas *et al.*, 1991; [http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html)) was used to detect CC domains in CNL proteins using a threshold = 0.9 followed by visual inspection.

InterPro is a resource that provides functional analysis of protein sequences by classifying them into families and predicting the presence of domains and important sites. To classify proteins in this way, InterPro uses predictive models, known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro consortium. The software package InterProScan allows sequences to be scanned against InterPro's signatures (more information are available in <http://www.ebi.ac.uk/InterProScan/>).

## 2.6 Transcriptomic analysis

### 2.6.1 RNA isolation, quantity and quality evaluation

The sampling was conducted on leaves of 3 groups of 6 plants each at 12th and 22th days after inoculation (dpi) on the two susceptible and resistant genotypes. Collected samples were immediately submerged in liquid nitrogen and then stored at -80° C. The total RNA extraction was performed on 100 mg of frozen zucchini leaves using TRIZOL (Invitrogen) protocol (Chomczynski and Mackey, 1995). A treatment with DNase I to remove DNA contaminations was done on total RNA. RNA samples were valuated quantitatively and qualitatively by NanoDrop ND-1000 Spectrophotometer (Nano-Drop Technologies, Wilmington, DE, USA) and by Bioanalyzer (Agilent Technologies). RNA integrity was checked by horizontal electrophoresis on a 1.2% (w/v) agarose gel prepared without addition of ethidium bromide. Two µg of each sample were prepared with 20 µl of 10 X RNA Loading Buffer composed by 400 µl Formamide, 10 µl 37% formaldehyde, µl loading buffer 10X (50% glycerol, 0.25% w/v bromophenol blue, 0.25% w/v xylene cyanol; Sigma), and 1. µl of 10 mg/µl ethidium bromide and than denatured at 6 °C for '. Gel visualization was performed using UV light (UV Gel Doc BIORAD). In order to obtain high quantity of RNA for each samples, total RNA from the six replicate was pooled for each 12 samples.

The samples, which have met the required standards of quality, were subjected to sequencing at the LabMedMolGe (Laboratory of Molecular Medicine and Genomics Department of Medicine and Surgery, University of Salerno).

### 2.6.2 RNA-Sequencing

Total purified RNA was converted to cDNA libraries and sequenced on Illumina HiSeq1500 platform at the LabMedMolGe (Laboratory of Molecular Medicine and Genomics Department of Medicine and Surgery, University of Salerno). The process generates millions (36 M) of short (100 bp) reads sequenced from both ends of each cDNA fragment (paired-end sequencing). The raw data for each sample consist of a long list of short sequences with associated quality scores (*fastq* format) (Ewing *et al.*, 1998a; Ewing *et al.*, 1998b). To evaluate the quality of the sequences the FastQC software was used (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>).

### 2.6.3 Functional annotation of reference transcriptome

Transcriptome (version 3.0), was mapped to the genome (version 2.0), in order to obtain the coordinates of the transcripts/genes over the genome. The PASA tool (version r20140417), which calls BLAT in order to align the transcripts over the genome and get the coordinates of the best hits (<http://pasa.sourceforge.net>), was employed. To perform the functional annotation of the reference transcriptome the BLAST2GO tool was used at the default parameters (Conesa *et al.*, 2005).

### 2.6.4 RNA-Sequencing data analysis

RNA-seq analysis was performed in three different steps: Reads processing; Statistical analysis of the data; Functional annotation of the identified loci. The high quality reads (obtained after trimming) were aligned against the *Cucurbita pepo* transcriptome (version v3) using TopHat v.2.0.13 (<http://ccb.jhu.edu/software/tophat>). Bowtie2 (included in TopHat) was selected as read aligner due to its high sensitivity. High quality reads were aligned against the transcriptome using minimum insert-length as zero and maximum insert-length as 500 (or more). The resulting alignment files (.bam files) were filtered for quality (quality score > 30) and duplicates. The obtained files were used as input for the calculation gene expression values. The output file includes one column for the transcript ID and one column for expected-read-counts for each sample. Data counts were then normalized using the RPKM method (Mortazavi *et al.*, 2008). All the statistical analyses were performed with R (version 3.1.0).

The overall quality of the experiment was evaluated, on the basis of the similarity between replicates, by a PCA analysis. Using this approach, we detected if any of the samples does not behave like the other replicates (outliers). The similarity between replicates was evaluated as well by the calculation of Euclidean distance between the samples and hierarchical clustering. The clustering of the samples was repeated after the removal. Once the consistency of biological replicates were evaluated, and divergent samples (outliers) were discarded, differential expression analysis was performed. The final dataset was composed by all the samples passing the Quality Control step.

The identification of the differentially expressed genes was performed with the package EBSeq (version 1.5.3). Sample 381-12dpi versus the sample TF-12dpi, 381-22dpi versus 381-12dpi, 381-22dpi versus TF-22dpi, TF-22dpi versus TF-12dpi have been compared. In addition, a more general analysis has been performed in order to identify DE transcripts in '381e' genotype respect to 'TFR' genotype. In this experiment, the time was not treated as factor. The transcripts were scored at q-value less than 0.05. A further filter was applied by selecting those with a fold change higher than

1.5-fold (UP-regulated) and lower than -1.5-fold (DOWN-regulated). The attachment of biological information to identified differential expressed transcripts was performed using the functional annotation file of the *Cucurbita pepo* transcriptome (version 3.0).

In order to detect new SNP markers on the transcripts of the susceptible 'TFR' and resistant '381e' accessions, a SNP calling analysis against the *Cucurbita pepo* reference genome was performed. SNP calling, based on the two filtered .bam files, produced two VCF (Variant Call Format) files ([http://www.broadinstitute.org/igv/viewing\\_vcf\\_files](http://www.broadinstitute.org/igv/viewing_vcf_files)) with the SNPs of each genotype in comparison to the genome. To obtain the SNPs between both genotypes we have found the common and unique SNPs between both VCF files, removing at the same time SNPs at low quality/coverage and inconsistent SNPs.

The protein sequences of our genes were interrogated by MATRIX RELOADED (Matrix-R) tools that, via HMM profiling, automatically retrieves, annotate and classified plant resistance genes. The protein sequences of the genes identified, were examined in order to underscore typical R gene domains. The output was a list of transcripts with a hit to any resistance domain Matrix-R was launched on transcripts with an ORF > 60 amino acids.

### 3. RESULTS

#### 3.1 Genetic analysis of resistance to ZYMV

##### 3.1.1 Phenotypic evaluation of resistance in segregating populations

Two F<sub>2</sub> populations deriving from the cross ‘True French’ x ‘381e’ (F<sub>2</sub> TRF) and ‘San Pasquale’ x ‘381e’ (F<sub>2</sub> SPQ) and three backcross populations obtained crossing F<sub>2</sub> ‘SPQ’ progenies with susceptible variety ‘SPQ’ were screened for resistance to ZYMV. Inoculated plants scored from 6 to 30 days post inoculation (dpi) exhibited wide scattered reaction to virus. Accession ‘381e’ and a number of resistant segregating progenies showed chlorotic spots, mild mottling then recovering or only a few chlorotic dots on new leaves. ‘True French’, ‘San Pasquale’ and all susceptible segregating plants showed severe mosaic, blistering and distortion on leaves and reduced growth.

As regards of ‘TFRx381e’ F<sub>2</sub> progeny, consisted of 175 individuals, 122 plants (70%) were classified as resistant, 102 plants showed the same symptoms of resistant parent, with mild mosaic and yellow spots at initial development stage (1-3rd leaf) and subsequent development of asymptomatic leaves (Group B). While, 20 plants showed no macroscopic symptoms on all leaves observed and classified as group A. Frequency of resistant plants on total of plants observed was calculated as  $f(R) = 0.70$ . The symptoms of the susceptible parent ‘TRF’ with malformations, mosaic and reduction of development were visualized in 53 individuals, about 30% of the full population screened. These individuals, exhibiting growth reduction, severe mosaic, blistering and leaf distortion were pooled in group C as previously described in materials and methods section. The observed segregation fitted with both the 3:1 and 45:19 ratio (Table 3.1), which suggested that one major gene control the resistant reaction in ‘381e’ but its action could be modulated from at least from other two other genes.

Also in the F<sub>2</sub> population ‘San Pasquale’ x ‘381e’ (n = 96), the segregation ratio was in accordance with a genetic control mediated from a major gene, putatively modulated from other genes. Phenotypic analysis of this population showed the occurrence of 62 resistant plants and of 34 susceptible,  $f(R) = 0.65$ . In particular, out of 62 plants judged resistant, 17 belong to the group A and 45 to the group B.

The F<sub>2</sub> individuals used to produce the backcross populations were carefully selected according to their allelic composition at putative loci *Zym-1* (A), *Zym-2* (B) and *Zym-3* (C) revealed previously by the molecular analysis carried out on the ‘SPQx381e’ F<sub>2</sub> population. The genotype ‘64-F<sub>2</sub>’ was characterized by a heterozygous locus Aa and two dominant homozygous loci BB and CC,

genotype '47-F<sub>2</sub>' had an allelic configuration Aa, bb, cc, '28-F<sub>2</sub>' is a triple heterozygote (Aa, Bb, Cc).

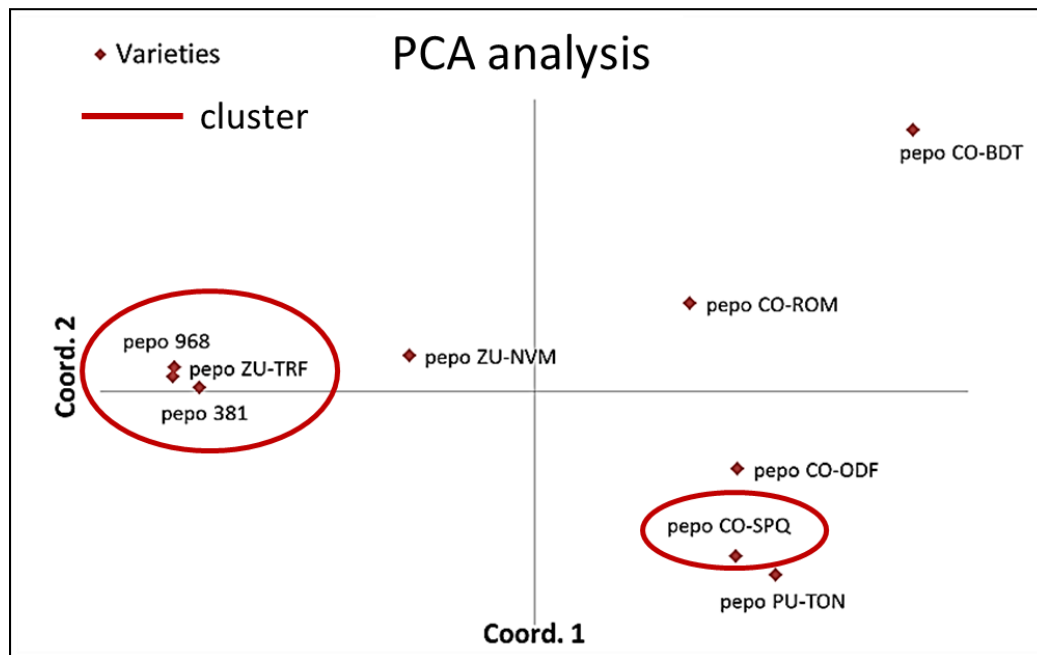
In the backcross of '64-F<sub>2</sub>', based on the theoretical allelic frequency, was expected to obtain 50% of resistant individuals belonging to genotypic class *Zym-1/zym-1 Zym-2/zym-2 Zym-3/zym-3* and 50% of susceptible belonging to genotypic class *zym-1/zym-1 Zym-2/zym-2 Zym-3/zym-3*. The '47-F<sub>2</sub>' should exhibit 50% of resistant plants (*Zym-1/zym-1 zym-2/zym-2 zym-3/zym-3*) and 50% of susceptible plants (*zym-1/zym-1 zym-2/zym-2 zym-3/zym-3*), whilst the resistant individuals originated from '28-F<sub>2</sub>' should be subdivided in 1/4 of individuals *Zym-1/zym-1 Zym-2/zym-2 Zym-3/zym-3*, 1/4 of individuals *Zym-1/zym-1 Zym-2/zym-2 zym-3/zym-3*, 1/4 of individuals *Zym-1/zym-1 zym-2/zym-2 Zym-3/zym-3* and 1/4 of individuals *Zym-1/zym-1 zym-2/zym-2 zym-3/zym-3*. The susceptible individuals should be subdivided in 1/4 *zym-1/zym-1 Zym-2/zym-2 Zym-3/zym-3*, 1/4 *zym-1/zym-1 Zym-2/zym-2 zym-3/zym-3*, 1/4 *zym-1/zym-1 zym-2/zym-2 Zym-3/zym-3* and 1/4 *zym-1/zym-1 zym-2/zym-2 zym-3/zym-3*.

The results obtained confirm the previsions in the '47-F<sub>2</sub>' and '28-F<sub>2</sub>' backcross. In particular, BC<sub>1</sub>-F<sub>2</sub> '47-F<sub>2</sub>xSPQ' population evidenced 23 susceptible plants and 29 with symptoms of tolerance, on a total of 52 plants tested,  $f(R) = 0.56$ . In BC<sub>1</sub>-F<sub>2</sub> '28-F<sub>2</sub>xSPQ' out of 68 tested plants, 36 showed signs of susceptibility, 32 symptoms of tolerance,  $f(R)$  was 0.47. Finally, out of 68 backcross plants obtained from '64-F<sub>2</sub>', 49 (72% instead of 50%) showed symptoms of susceptibility, 19 showed tolerance (28% instead of 50%), in the juvenile stages (20 dpi). Frequency of resistant plants was 0.28. Resistant plants were kept for further crosses useful to transfer the ZYMV resistance in the 'San Pasquale' genotype.

### 3.1.2 Molecular screening of parental lines

A 384 Illumina GoldenGate platform designed on a Zucchini Unigene-Dataset (Blanca *et al.*, 2011) was used in order to analyze the genetic relationships of three parental lines and of six lines used as control. The GoldenGate genotyping assay allowed us to identify 39% of polymorphic SNPs among varieties tested. Based on SNPs profiles obtained a PCA profile was built up. The phylogenetic distance and the correlation between the nine analyzed cultivars was evaluated as showed in figure 3.1. Only six SNPs detected variation between the two isogenic lines, '381e' (resistant to ZYMV) and 'True French' (susceptible to ZYMV). Out of six markers used to develop a Real-Time PCR-HRM assays, three produced a clear melting curve.

**Figure 3.1:** Phylogenetic relationships between nine *Cucurbita pepo* cultivars. In the red circle on the left were pooled the three isogenic lines ‘True French’, ‘968Rb’ (susceptible) and ‘381e’ (resistant) and on the right the cultivar ‘San Pasquale’ belonging a group of cocozelle.

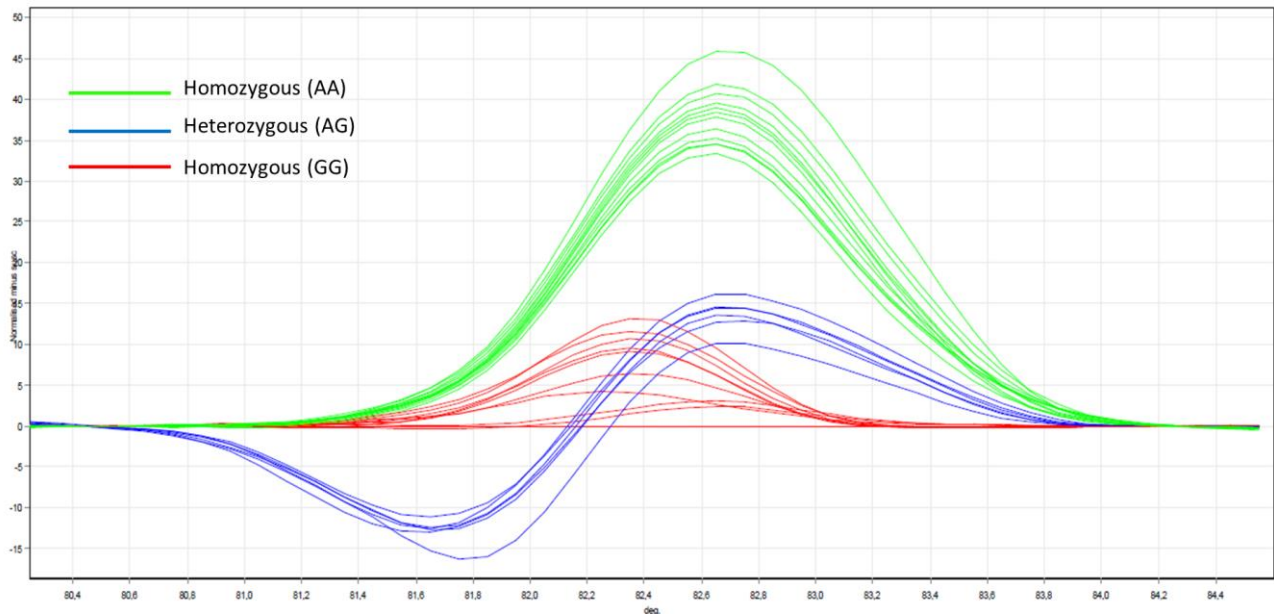


### 3.1.3 Genotyping of segregating populations

HRM assay with SNP1, SNP2 and SNP3 carried out on 5 segregating populations, characterized for resistance to ZYMV allowed a quick and efficient discrimination between individuals homozygous resistant, susceptible homozygous and heterozygous resistant at loci tested.

Thanks to the molecular analysis, we evaluated the putative association of SNP markers to three genes (*Zym-1*, *Zym-2* and *Zym-3*) involved in resistance to ZYMV in two F<sub>2</sub> populations and three backcross F<sub>2</sub> populations.

The subset of three SNPs were investigated on 174 F<sub>2</sub> ‘TFRx381e’ individuals, 91 F<sub>2</sub> ‘SPQx381e’ individuals, 60 BC<sub>1</sub>-F<sub>2</sub> ‘64-F<sub>2</sub>xSPQ’, 49 ‘47-F<sub>2</sub>xSPQ’ and 65 of ‘28-F<sub>2</sub>xSPQ’ individuals. The assay performed on the five segregating populations allowed the clear discrimination of allelic state at three loci in all individuals tested. High-Resolution Melting analysis revealed significant differences in melting curve shape that correlate SNP to the presence of a given allele. Therefore based on this assay it was possible to discriminate dominant homozygotes (AA), heterozygotes (Aa) and homozygous recessive (aa) individuals for each locus. Representative normalized melt curves of SNP1 marker are shown in figure 3.2.



**Figure 3.2:** High Resolution Melting Real-Time PCR. HRM profiles of SNP1 marker in a segregating population of zucchini plants. Normalized melting plot between resistant homozygous *R/R*, susceptible homozygous *r/r* and resistant heterozygous *R/r* are shown. Representative profiles of the melting curves of amplicons for  $F_2$  population.

HRM analysis performed on the ‘ $F_2$  TFR’ population revealed a segregating ratio of 130 A- plant to 44 aa plants,  $f(A-) = 0.75$  and on the  $F_2$  obtained from the ‘ $F_2$  SPQ’ showed a segregating ratio of 66 A- to 25 aa samples,  $f(A-) = 0.73$ . In these  $F_2$  populations, the resistant individuals were characterized by the configurations A-B-C-, AaBBcc, A-bbC-, A-bbcc, and the susceptible were aaB-C-, aaB-cc, aabbC-, aabbcc. These 27 allelic configurations are summarized in figure 3.3.

In the  $BC_1$ - $F_2$  ‘64- $F_2$ xSPQ’ the  $f(A-)$  calculated was 0.47. On 60 analyzed plants, 28 resulted with the allelic configuration AaBbCc and 32 aabbcc. While in the ‘47- $F_2$ xSPQ’ backcross the resistant plants was 21 (Aabbcc) and the susceptible 28 (aabbcc) with an  $f(A-) = 0.43$ . Lastly, in the  $BC_1$ - $F_2$  ‘28- $F_2$ xSPQ’ there was 36 resistant subdivided in  $\frac{1}{4}$  of individuals AaBbCc,  $\frac{1}{4}$  of individuals AaBbcc,  $\frac{1}{4}$  of individuals AabbCc and  $\frac{1}{4}$  of individuals Aabbcc as well as the 29 susceptible individuals were subdivided in  $\frac{1}{4}$  aaBbCc,  $\frac{1}{4}$  aaBbcc,  $\frac{1}{4}$  aabbCc and  $\frac{1}{4}$  aabbcc for a total of 65 ( $f(A-) = 0.55$ ). The results obtained by HRM analysis for each marker were subject to a chi-square test for a single dominant gene segregation ratio 3:1 and for three dominant genes segregation ratio 45:19. This test has highlighted a strong correlation between SNP observed frequency and expected frequency for both ratio.

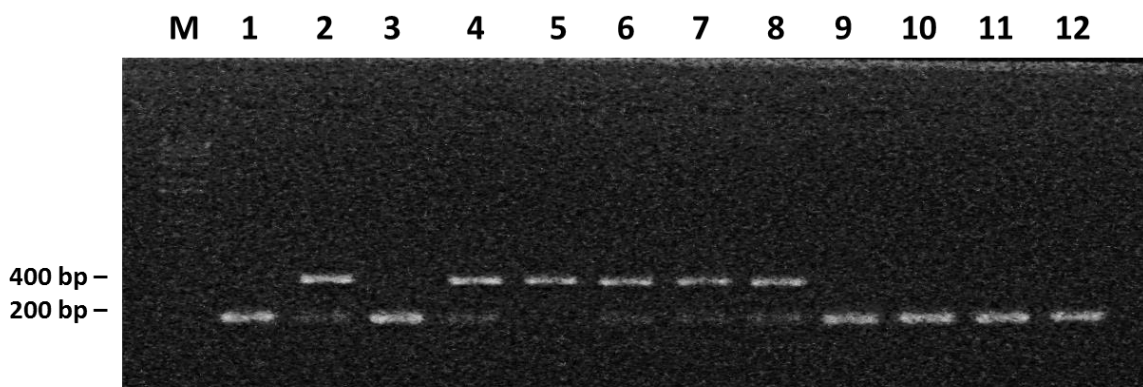


	ABC	ABc	AbC	aBC	abc	Abc	abC	aBc
ABC	AABBCC	AABBCCc	AABbCC	AaBBCC	AaBbCc	AABbCc	AaBbCC	AaBBCCc
ABc	AABBCCc	AABBcc	AABbCc	AaBBCCc	AaBbcc	AABbcc	AaBbCc	AaBBcc
AbC	AABbCC	AABbCc	AAbbCC	AaBbCC	AabbCc	AAbbCc	AabbCC	AaBbCc
aBC	AaBBCC	AaBBCCc	AaBbCC	aaBBCC	aaBbCc	AaBbCc	aaBbCC	aaBBCCc
abc	AaBbCc	AaBbcc	AabbCc	aaBbCc	aabbcc	Aabbcc*	aabbCc	aaBbcc
Abc	AABbCc	AABbcc	AAbbCc	AaBbCc	Aabbcc*	Aabbcc*	AabbCc	AaBbcc
abC	AaBbCC	AaBbCc	AabbCC	aaBbCC	aabbCc	AabbCc	aabbCC	aaBbCc
aBc	AaBBCCc	AaBBcc	AaBbCc	aaBBCCc	aaBbcc	AaBbcc	aaBbCc	aaBBcc

**Figure 3.3:** F<sub>2</sub> expected allelic configuration of three complementary genes involved in resistance mechanism. In green and red are labeled the resistant and susceptible genotypes, respectively. In this work, the genotypes Aabbcc\* and Aabbcc\* showed resistance variations among individuals.

The SNP1 marker showed a strong correlation with phenotypic data and it was used to develop also a CAPS marker able to validate results obtained by HRM assay on 167 F<sub>2</sub> ‘TFRx381e’ individuals.

The PCR reaction produced amplicons size of approximately 400bp. The BglII restriction enzyme digestion allowed us to discriminate susceptible from resistant samples. In particular, the susceptible samples have produced two fragments of 200bp (displayed on an agarose gel as a single band), those judged resistant a 400bp fragment, whereas the heterozygous samples produce both fragments (Figure 3.4).



**Figure 3.4:** A Cleaved Amplified Polymorphism Sequence (CAPS) marker. The polymorphism of the marker SNP1 is shown in segregating population F<sub>2</sub> generated crossing ‘381e’ (resistant parent) and ‘True French’ (susceptible parent). In lane 5: resistant homozygous genotypes (R/R 400 bp); in lane 1, 3, 9, 10, 11 and 12: susceptible homozygous genotype (r/r – 200 bp); in lane 2, 4, 6, 7 and 8: resistant heterozygous genotypes (R/r - 400 bp and 200 bp), M: ladder (1Kb<sup>+</sup>, Invitrogen).

### 3.1.4 Statistical analysis goodness of fit

‘True French’ x ‘381e’ F<sub>2</sub> population phenotypic evaluation showed 122 resistant samples and 53 susceptible samples, in accordance with the expected 45:19 ratio ( $\chi^2 = 0.03$  and  $P = 0.86$ ). The f (R) number of resistant plants/total number of tested plants is 0.70 as shown in table 3.1. In addition, the ratio of 62 resistant on 34 susceptible individual observed in the cross between ‘San Pasquale’ and ‘381e’ statistically fitted with an expected ratio 45:19 ( $\chi^2 = 1.51$  and  $P = 0.22$ ). In this two F<sub>2</sub> population, the observed ratio also fitted with the 3:1 segregating expected ratio, but with lower value at  $P > 0.05$  (Table 3.1). The 45:19 ratio was tested assuming that three complementary dominant genes are involved in resistance to ZYMV (Paris and Cohen, 2000).

The phenotypic evaluation of three backcross populations showed an observed ratio of 19 resistant and 49 susceptible in the ‘64-F<sub>2</sub>’ derived backcross population, 29:23 in the ‘47-F<sub>2</sub>xSPQ’ and 32:36 in the ‘28-F<sub>2</sub>’. In the latter two, the observed ratio fitted with the 1:1 expected ratio.

Phenotypic evaluations suggest that the response to ZYMV in a first stage of resistance process require the participation of three genes. In particular, one major gene *Zym-1* and two minor genes *Zym-2* and *Zym-3*, which may have a complementary effect, producing an F<sub>2</sub> ratio of 45:19 and a BC<sub>1</sub>-F<sub>2</sub> ratio of 1:1 seems to be involved. Analyzed plants were considered resistant assuming that the resistant genotypes should have at least one allele of the locus *Zym-1* dominant (*Zym-1/-*), complementary genes *Zym-2* and *Zym-3* (*Zym-2/-* and/or *Zym-3/-*) could be necessary to keep resistance in later stage of development.

**Table 3.1:** Phenotypic evaluation perfectly reflect the expected segregation ratio of 45:19 for the F<sub>2</sub> populations and 1:1 for the BC<sub>1</sub>-F<sub>2</sub> populations except for the ‘64-F<sub>2</sub>’ backcross line.

ID_Popuation	No. pt Resistant	No. pt Susceptible	No. pt Total	Expected ratio (R:S)	$\chi^2$ (P)	f (R)
F <sub>2</sub> (TFRx381)	122	53	175	45:19	0.86	0.70
F <sub>2</sub> (SPQx381)	62	34	96	45:19	0.22	0.65
F <sub>2</sub> (TFRx381)	122	53	175	3:1	0.11	0.70
F <sub>2</sub> (SPQx381)	62	34	96	3:1	0.02	0.65
BC <sub>1</sub> -F <sub>2</sub> (64-F <sub>2</sub> xSPQ)	19	49	68	1:1	0.001	0.28
BC <sub>1</sub> -F <sub>2</sub> (47-F <sub>2</sub> xSPQ)	29	23	52	1:1	0.71	0.56
BC <sub>1</sub> -F <sub>2</sub> (28-F <sub>2</sub> xSPQ)	32	36	68	1:1	0.88	0.47

Molecular results obtained with the HRM analysis conducted were also subject to chi-square test. For each marker was calculated the segregation ratio for a single dominant gene. Assuming that also the samples carrying the genotype *Zym-1/Zym-1 zym-2/zym-2 zym-3/zym-3* and *Zym-1/zym-1 zym-2/zym-2 zym-3/zym-3* were resistant, the segregating expected ratios were 3:1 in the two F<sub>2</sub> populations and 1:1 in the BC<sub>1</sub>-F<sub>2</sub> populations. The chi-square test performed on dataset obtained by the molecular analysis confirmed this expected ratios in all five cases suggesting a close linkage between analyzed SNPs and ZYMV resistance genes. Moreover, in order to envisage the tendency of genetic markers to co-segregate with the phenotypic data was calculated the percent of co-segregation between the genotype and the phenotype. Co-segregation was ascertained in all five population with a minimum of 47% in the BC<sub>1</sub>-F<sub>2</sub> ‘47-F<sub>2</sub>xSPQ’, considering as susceptible the genotype *Zym-1/Zym-1 zym-2/zym-2 zym-3/zym-3* and *Zym-1/zym-1 zym-2/zym-2 zym-3/zym-3*, to a maximum of 89% in the ‘TFRx381e’ F<sub>2</sub>. Table 3.2 reports the results of a chi-square goodness of fit test and the percent of co-segregation calculated between the genotype and the phenotype. A ratio of 3:1 segregation was observed for each F<sub>2</sub> segregating populations reflecting a close linkage between analyzed SNPs and ZYMV resistance genes.

**Table 3.2:** Based on the genotypic data, chi-square test was calculated for 45:19 and 3:1 expected ratio in the F<sub>2</sub> populations and for 1:1 expected ratio in BC<sub>1</sub>-F<sub>2</sub> populations. The ratio labeled with \* report the calculated  $\chi^2$  considering as resistant the genotype *Zym-1/Zym-1 zym-2/zym-2 zym-3/zym-3* and *Zym-1/zym-1 zym-2/zym-2 zym-3/zym-3*. The f (A-) and the percent of co-segregation of genotype vs phenotype was calculated, too.

ID_Population	No. pt Resistant	No. pt Susceptible	No. pt Total	Expected	$\chi^2$ (P)	f(A-)	Co-segregation genotype vs phenotype (%)
				ratio (R:S)			
F <sub>2</sub> (TFRx381)	130	44	174	3:1*	0.93	0.75	89
F <sub>2</sub> (SPQx381)	66	25	91	3:1*	0.59	0.73	80
F <sub>2</sub> (TFRx381)	130	44	174	45:19*	0.20	0.75	89
F <sub>2</sub> (SPQx381)	66	25	91	45:19*	0.64	0.73	80
F <sub>2</sub> (TFRx381)	121	53	174	3:1	0.10	0.70	84
F <sub>2</sub> (SPQx381)	62	29	91	3:1	0.13	0.68	78
F <sub>2</sub> (TFRx381)	121	53	174	45:19	0.82	0.70	84
F <sub>2</sub> (SPQx381)	62	29	91	45:19	0.65	0.68	78
F <sub>2</sub> (TFRx381) CAPS	123	44	167	3:1	0,70	0.74	96
F <sub>2</sub> (TFRx381) CAPS	123	44	167	45:19	0,32	0.74	96
BC <sub>1</sub> -F <sub>2</sub> (64-F <sub>2</sub> xSPQ)	28	32	60	1:1*	0.88	0.47	82
BC <sub>1</sub> -F <sub>2</sub> (47-F <sub>2</sub> xSPQ)	21	28	49	1:1*	0.61	0.43	80
BC <sub>1</sub> -F <sub>2</sub> (28-F <sub>2</sub> xSPQ)	36	29	65	1:1*	0.76	0.55	86
BC <sub>1</sub> -F <sub>2</sub> (64-F <sub>2</sub> xSPQ)	28	32	60	1:1	0.88	0.47	82
BC <sub>1</sub> -F <sub>2</sub> (47-F <sub>2</sub> xSPQ)	0	49	49	1:1	0.00	0	47
BC <sub>1</sub> -F <sub>2</sub> (28-F <sub>2</sub> xSPQ)	17	48	65	1:1	0.0004	0.26	72

The statistical analysis of chi-square and percent of co-segregation between the frequency of the single SNP markers in each populations and the phenotype were also performed.

In the two F<sub>2</sub> populations, the expected ratios for each SNPs were 3:1 considering the AA and Aa as resistant and aa susceptible. The obtained  $\chi^2$  (P) of F<sub>2</sub> 'TFRx381e' SNP1, SNP2 and SNP3 were 0.93, 0.19 and 0.15 respectively. In the F<sub>2</sub> 'SPQx381e' instead were  $\chi^2$  (P) = 0.95 for SNP1,  $\chi^2$  (P) = 0.59 for SNP2 and  $\chi^2$  (P) = 1.00 for SNP3. In the BC<sub>1</sub>-F<sub>2</sub> was calculated a  $\chi^2$  (P) > 0.5 in all SNPs analyzed.

In all the populations, the co-segregation between SNP1 marker and phenotypic data was higher of 80% except for the SNP1 valuation in 64-F<sub>2</sub>, indicating a strong relationship between the SNP1 and the ZYMV resistance genes. In particular, the analysis highlighted the SNP1 could be linked to the major candidate gene and the SNP2 and SNP3, showing a lower percent of co-segregation, to minor genes involved (Table 3.3).

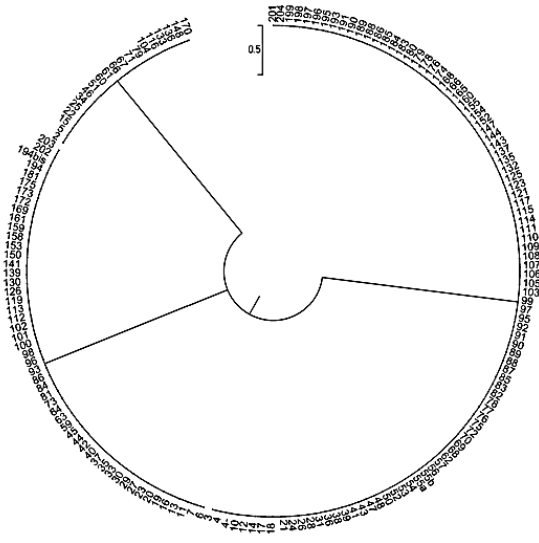
**Table 3.3:** Based on the associated markers data, chi-square test was calculated for 3:1 expected ratio in the F<sub>2</sub> populations and for 1:1, 1:0 and 0:1 expected ratio in BC<sub>1</sub>-F<sub>2</sub> populations. Again, the percent of co-segregation of SNPs vs phenotype was calculated.

ID_Popuation	SNP_ID	Linked markers	No. dominant alleles (A-)	No. recessive alleles (aa)	No. pt total	Expected ratio (R:S)	$\chi^2$ (P)	Percent co-segregation (%)
F <sub>2</sub> (TFRx381)	SNP1	Zym1	130	44	174	3:1	0.93	89
	SNP2	Zym2	138	36	174	3:1	0.19	64
	SNP3	Zym3	138	35	173	3:1	0.15	62
F <sub>2</sub> (SPQx381)	SNP1	Zym1	68	23	91	3:1	0.95	80
	SNP2	Zym2	66	25	91	3:1	0.59	72
	SNP3	Zym3	66	22	88	3:1	1	54
BC <sub>1</sub> -F <sub>2</sub> (64-F <sub>2</sub> xSPQ)	SNP1	Zym1	28	32	60	1:1	0.61	78
	SNP2	Zym2	68	0	68	1:0	1	28
	SNP3	Zym3	68	0	68	1:0	1	28
BC <sub>1</sub> -F <sub>2</sub> (47-F <sub>2</sub> xSPQ)	SNP1	Zym1	21	28	49	1:1	0.32	80
	SNP2	Zym2	0	51	51	0:1	1	57
	SNP3	Zym3	0	51	51	0:1	1	57
BC <sub>1</sub> -F <sub>2</sub> (28-F <sub>2</sub> xSPQ)	SNP1	Zym1	36	29	65	1:1	0.39	86
	SNP2	Zym2	31	33	64	1:1	0.80	45
	SNP3	Zym3	29	36	65	1:1	0.39	51

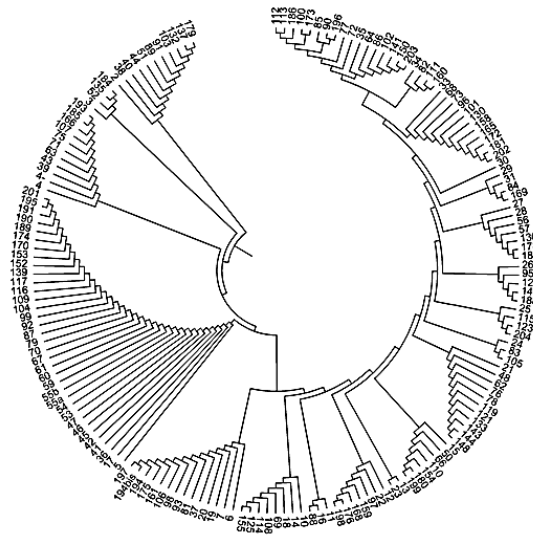
Results obtained from the CAPS analysis also fit with a 3:1 ratio (Table 3.2), confirming a strong relationship between SNP1 and *Zym-I* resistance genes ( $\chi^2 = 0.16$ ,  $P = 0.7$ ).

To calculate the overall relationship between phenotypic and genotypic results, we used the Mantel test (Mantel, 1967) on data obtained from 'TFRx381e' and 'SPQx381e' populations, as shown in figure 3.5. In both populations, the p-value  $< 0.0001$  was lower of significant level  $\alpha = 0.05$ . The p-value was calculated using the distribution of  $r(AB) = 0.186$  in the 'F<sub>2</sub> TFR' and  $r(AB) = 0.196$  in the 'F<sub>2</sub> SPQ', estimated on 10.000 permutations, allowing to confirm the correlation existing between dissimilarity matrices originated by phenotypic and genotypic dataset of the two F<sub>2</sub> segregating populations. Dissimilarity matrices, based on the Euclidean distance, were than used to produce a graphical representation of relationships existing among samples show in figure 3.5.

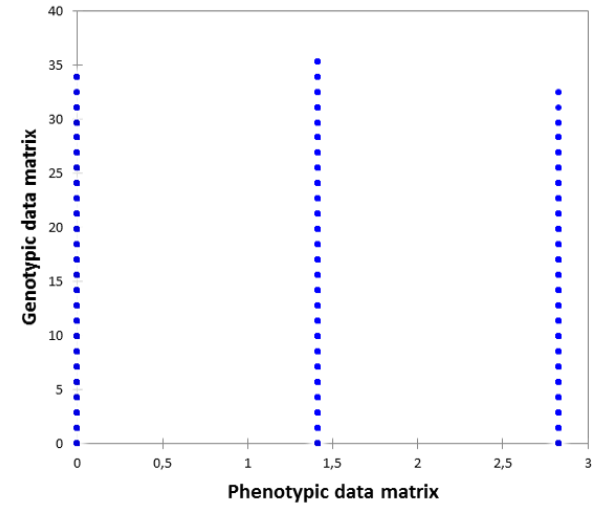
**Figure 3.5:** Samples clustering, based on phenotypic dataset, in three classes of evaluation of 'TFRx381e' (A) and 'SPQx381e' (D). Samples clustering, on genetic dataset, in the 23 out of 27 expected allelic configuration segregation of three genes in 'TFRx381e' (B) and in 'SPQx381e' (E). Mantel test output obtained from the correlation between phenotypic and genotypic matrices.



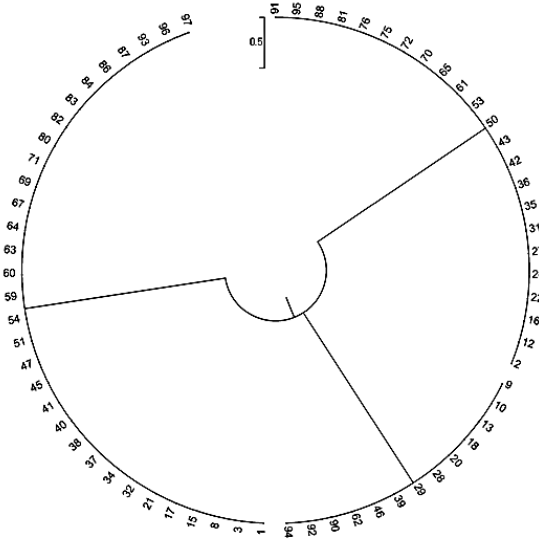
**A**



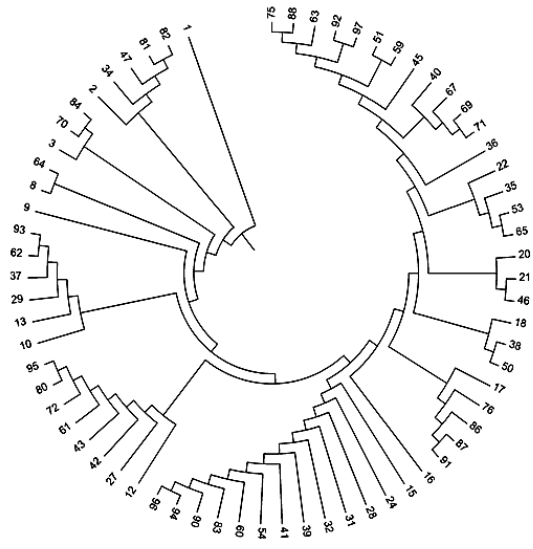
**B**



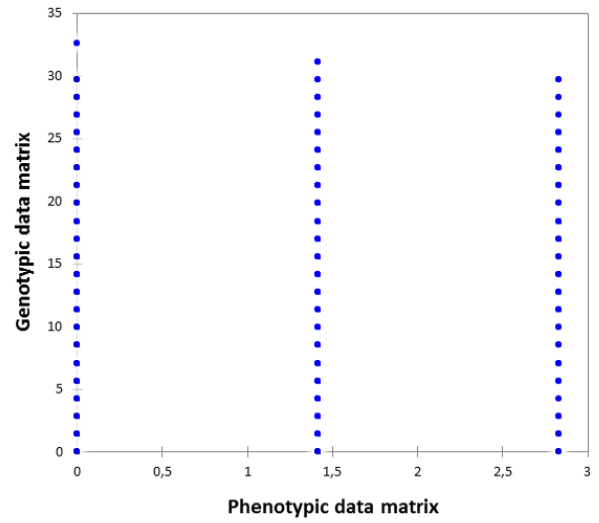
**C**



**D**



**E**



**F**

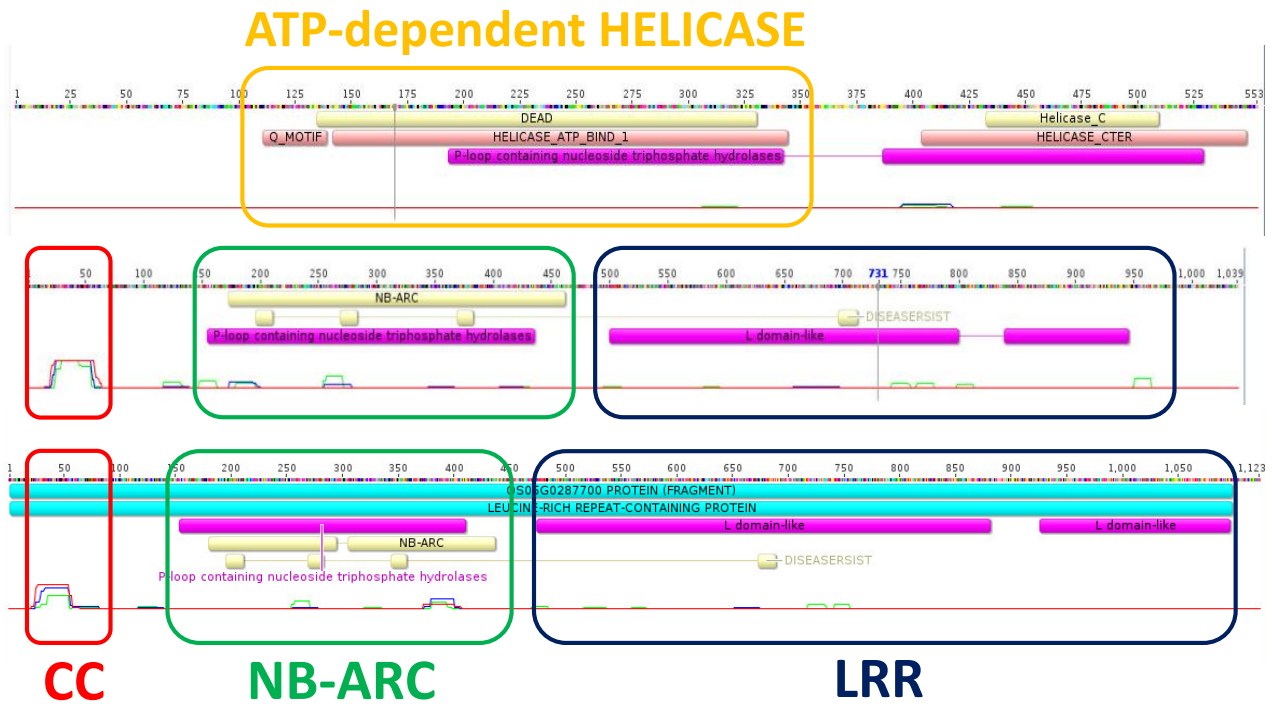
### 3.1.5 Analysis of genome scaffolds containing ZYMV markers

The main goal of our work was to develop SNP markers linked to the genes conferring ZYMV resistance. In order to identify genes physically closer to our markers, a functional prediction of protein containing domains involved in the resistance mechanisms was performed. Firstly, the three unigenes containing SNP1, SNP2 and SNP3 identified through the GoldenGate assay have been searched on the genome v2. Table 3.4 reports the scaffolds to which unigenes maps, the percentage of identity and of coverage.

**Table 3.4:** Based on the GenScan tool were detected the coordinate of unigenes on the scaffolds.

<b>Unigene_ID</b>	<b>Scaffold_ID</b>	<b>Identity</b>	<b>Coverage</b>	<b>Start</b>	<b>stop</b>
unigene8819	scaffold1377	97.1%	37.10%	254671	254367
unigene11643	scaffold1174	99.6%	80.85%	546127	546865
unigene11643	scaffold1174	98.4%	19.80%	544665	544847
unigene9577	scaffold2494	99.5%	100.00%	79461	80079

The scaffolds with the highest percentage of identity and coverage is the scaffold2494 for the unigene containing the SNP1, scaffold1377 for the unigene of SNP2 and the scaffold1174 for SNP3. The length of these three scaffolds was between 257.375 bp and 714.477 bp. GenScan tool was used to translate the scaffolds sequences in the peptides in order to interrogate by InterProScan tool. The InterProScan prediction carried out on scaffolds allow us to find a closely association of the SNP1 with two CNL domains, a RLK domain and an ATP-dependent RNA helicase as shown in figure 3.6. ATP-dependent RNA helicase, LRR and other domains potentially involved in pathogen recognition were also found on the scaffolds to which belong the SNP2 and SNP3.



**Figure 3.6:** By InterProScan analysis, SNP1 marker results to be associated to ATP-dependent RNA helicase, two CNL and a RLK (data not showed) in genome v2.

The close relationship found between the R-domains and SNP1 suggest us that such marker could be associated to one of three resistance genes.

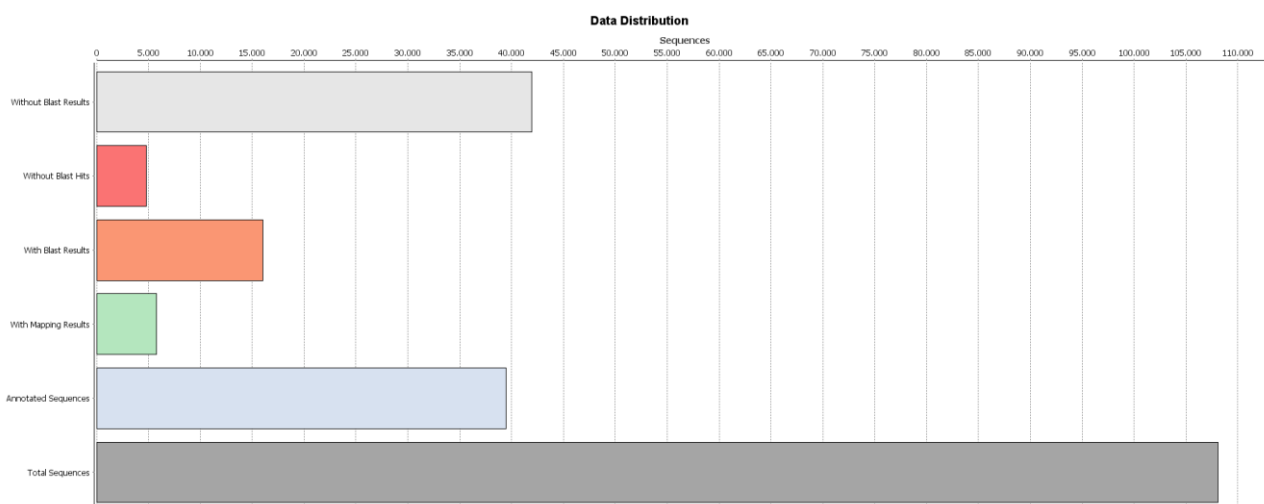


## 3.2 Transcriptome analysis of zucchini reference genome and of susceptible and resistant lines

### 3.2.1 Reference transcriptome assembling and functional annotation

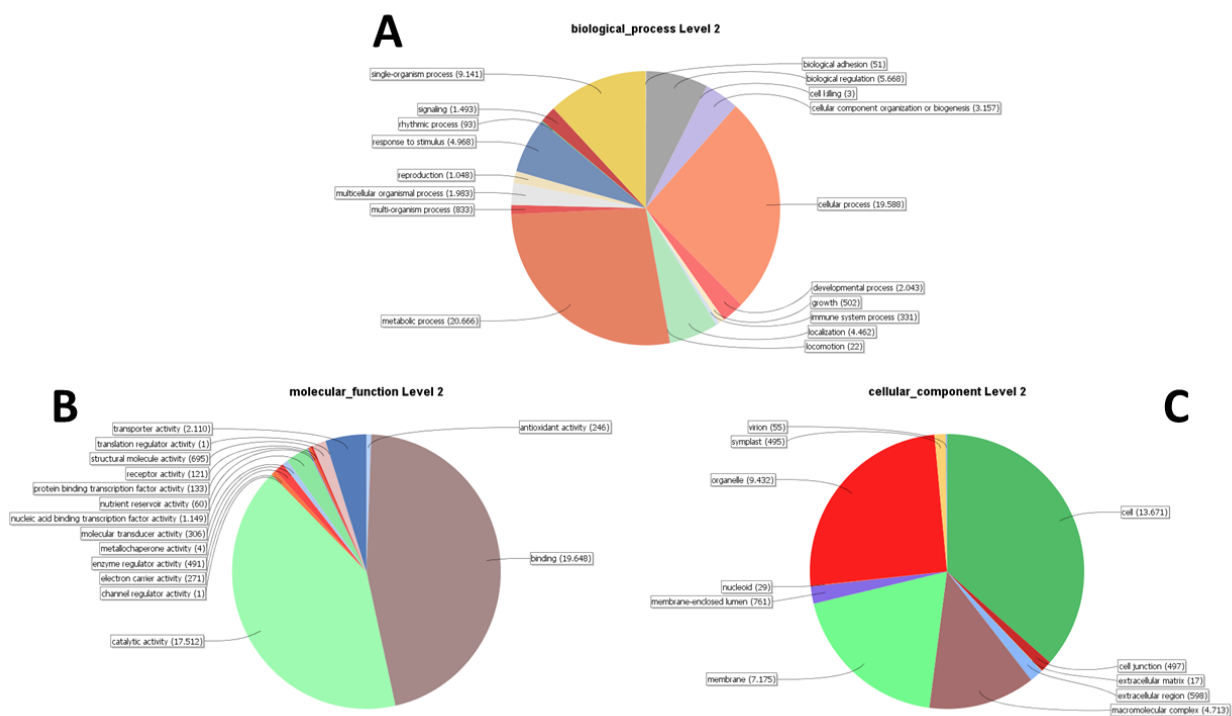
In order to perform a deep functional annotation of *C. pepo* reference transcriptome, the version 3 was assembled in 108,062 transcripts, revealing 90% of overall alignment of transcripts over the genome. Out of 108,062 reference transcripts 79,733 showed a full transcribed sequence and 28,330 fail to give a complete transcript. They were classified as not full transcripts and further explored in order to identify broken sequences at scaffold ends. Approximately 1000 cases were extracted and added to first dataset because full transcript was reconstructed.

A total of 108,062 transcripts were analyzed using the Blast2GO tool (B2GO; Conesa *et al.*, 2005), at the default parameters, in order to understand their putative functions. The analysis revealed that 4.44% of the transcripts do not showed similarity with known sequences collected in interrogated databases. On the other hand, only the 36.54% of the remaining transcripts (95%) were classified according the terms included in the three main ontology categories (molecular function, biological process and cellular component) proposed by the Gene Ontology consortium which provides an ontology of defined terms representing gene product properties. A relevant portion transcripts (39,486) showed a partial human readable annotation. The data distribution of B2GO analysis is showed in figure 3.11.



**Figure 3.11:** Distribution of B2GO analyzed transcripts belongs to reference transcriptome v3.

This analysis was performed in order to identify the biological function of each zucchini transcript. In figure 3.12 is reported the graphical representation of the distribution of GO terms among the three ontologic category (P) biological process, (F) molecular function and (C) cellular component, based on 108.062 transcripts. In the biological process, the two biggest functional categories were represented by the “metabolic process” and “cellular process”, with 20,666 and 19,588 transcripts respectively. Out of annotated transcripts in the molecular function category, 19,648 transcripts were involved in “binding” and 17,512 transcripts in “catalytic activity”. In the cellular component, the three most representative functional categories were “cell” (13,671), “organelle” (9,432) and “membrane” (7,175).



**Figure 3.12:** Transcripts level 2 classification in (A) Biological process, (B) molecular function and (C) Cellular component ontology categories.

### 3.2.2 Transcriptome sequencing and mapping of zucchini resistant and susceptible genotypes

RNA sequencing experiment was performed on three biological replicates of two resistant and susceptible genotypes inoculated with ZYMV collected at 12 and 22 dpi. A paired-end tag sequencing strategy was chosen, in which short and paired tags were extracted from the ends of long DNA fragments for ultra-high-throughput sequencing. The protocol included strand-specific library preparation followed by 100 base pair (bp) paired-end reads sequencing. Before to perform further analysis, a quality check was conducted on the raw sequencing data, removing low quality portions and preserving the longest high quality part of a NGS read. RNA-seq generated around 36 million reads of 100 bp length.

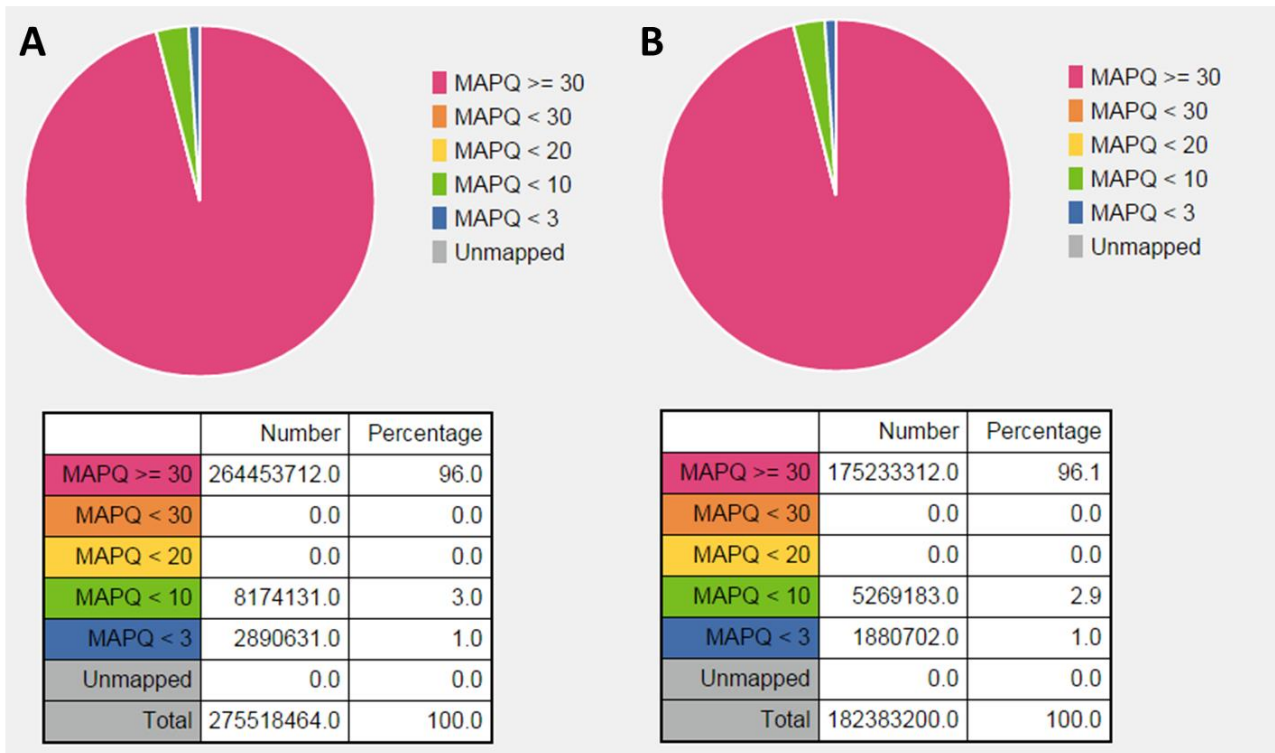
In order to ensure a high quality and reliability of the analysis was established a minimum sequence length of 25 bp and a sequence quality score of 35. The total number of reads per sample before and after the trimming, the number of sequenced bases and the coverage mapping is reported in table 3.6.

**Table 3.6:** Number of samples reads before and after the Trimming, Sequenced bases number and Coverage mapping.

	«381e»	«TFR»
<b>Library type</b>	Paired ends	Paired ends
<b>Sequenced reads length</b>	2 x 100 bp	2 x 100 bp
<b>Number of reads before trimming</b>	163,483,878.00	126,621,462.00
<b>Number of reads after trimming</b>	151,087,631.00	112,382,770.00
<b>Sequenced bases</b>	30.22 Gbp	22.48 Gbp

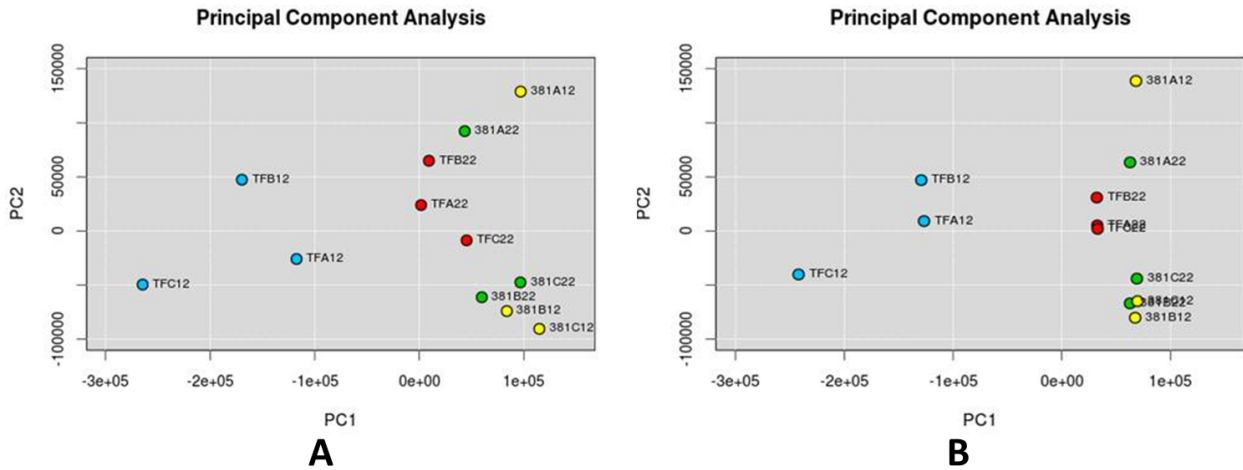
The high quality reads, obtained after trimming, were aligned against the *Cucurbita pepo* transcriptome v3, assigning a number of reads to each annotated locus and calculating the relative abundances. The resulting alignment files were used as input for the calculation gene expression values. The output file includes one column for the transcript ID and one column for expected-read-counts for each sample. Columns correspond to the three biological replicates for the four conditions under analysis ('True French' and '381e' collected both at 12 and 22 dpi). Data counts were normalized using the RPKM method (Mortazavi *et al.* 2008).

The quality control of the mapped reads of '381e' and 'TFR' genotypes using TopHat, allow us to remove the mapped reads with quality score < 30 and remove duplicates generated during the sequencing reaction. The two Quality BAM files produced were graphically represented in figure 3.8.



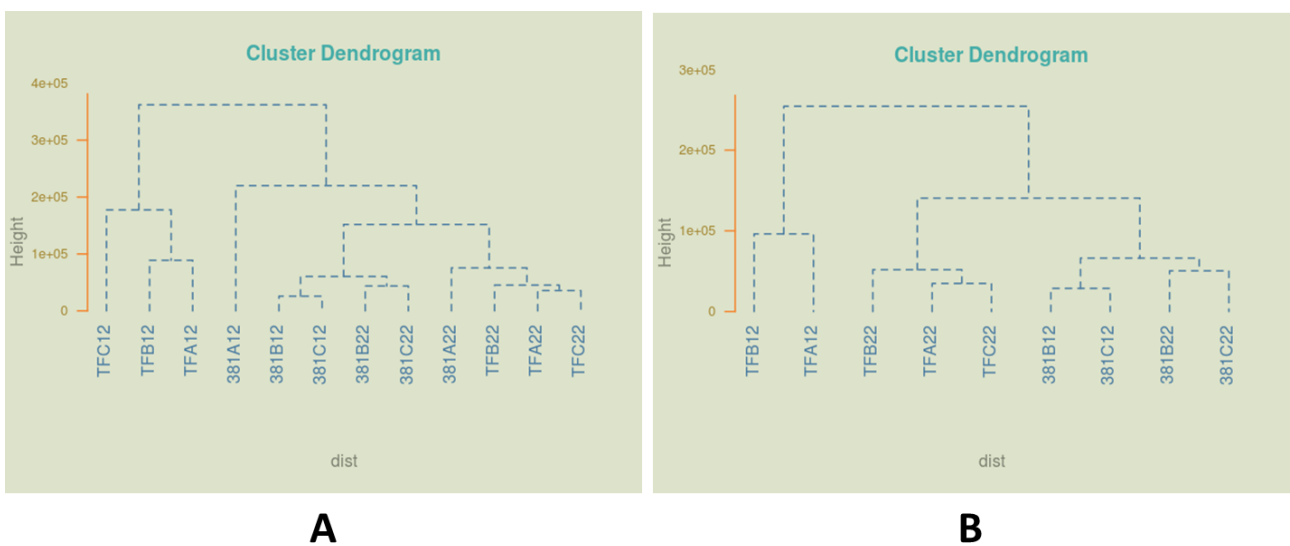
**Figure 3.8:** Graphical representation of number of alignments in various mapping quality (MAPQ) intervals and number of mapped sequences of '381e' genotype (A) and 'TFR' genotype (B).

The assembled transcripts, 73348 and 73609 for 'TFR' and '381e' respectively, have a mean length of 1052 bp, minimum length of 201 bp and a maximum length of 9917 bp. The overall quality of the biological replicates was evaluated, based on the similarity between replicates, by a PCA analysis. At first, the similarity was calculated using obtained reads reported in table 3.6. Since the zucchini available transcriptome version 3 is very fragmented and the annotation does not distinguish between splicing variants and transcripts of different loci. For this reason, we extracted the unique transcripts with expression values different from the others and for which the reads mapped univocally. Therefore, the 32,520 unique transcripts were selected. The quality control and the samples clustering is shown in figure 3.9.



**Figure 3.9:** Clustering of sequenced samples using all the reads (A) and only the unique transcripts obtained after the repeated quality control (B).

The similarity between replicates and the Euclidean distance between the samples was evaluated by hierarchical clustering. Using this approach, we ascertain if any samples behave differently by the other replicates (outliers). The cluster analyses highlighted that some replicates did not behave as expected showing too much divergence from the other replicates of the same group. Therefore, samples *381A12*, *381A22* and *TFC12* were removed from the downstream analysis. As the TF-22 replicates were the most uniform, all of them were used for the following analyses. The resulting plots, performed with the 32,520 unique transcripts, before (A) and after (B) the removal of the outliers are showed below in the figure 3.10.



**Figure 3.10:** Unique transcripts (32,520) clustering obtained after the repeated quality control of all sequenced samples (A) and after removed outliers (B).

Because of the low quality of the available reference transcriptome, the following step was to annotate the transcriptome to the genome, in order to obtain the coordinates of the transcripts/genes over the genome, before starting the RNA-seq analysis.

### 3.2.3 Identification of Differential Expressed Genes

Differential expression analysis has been performed on the final datasets without outliers. To discover the differentially expressed genes (DEGs), the following comparisons were included in the analysis:

- 381-12dpi vs TF-12dpi
- 381-22dpi vs 381-12dpi
- 381-22dpi vs TF-22dpi
- TF-22dpi vs TF-12dpi
- 381 (all times) vs TF (all times)

A change was deemed significant based on the following criteria: the transcript is scored and the q-value of the statistical test is less than 0.05. In order to get the list of the mostly interesting genes, a further filter was applied by selecting those with a fold change higher than 1.5-fold (UP-regulated) and lower than -1.5-fold (DOWN-regulated). The 381-12dpi vs TF-12dpi comparison aims to the identification of the transcripts differentially expressed (DE) in the '381e' resistant line respect to 'True French' susceptible line after 12 day from virus inoculation. A total of 680 up-regulated and 900 down-regulated genes were identified in the resistant line.

The 381-22dpi vs 381-12dpi comparison aims the identification of the transcripts differentially expressed (DE) in the 381-22dpi line respect to 381-12dpi line. Out of a total 1302 total DE transcripts, 656 were up-regulated and 646 down-regulated.

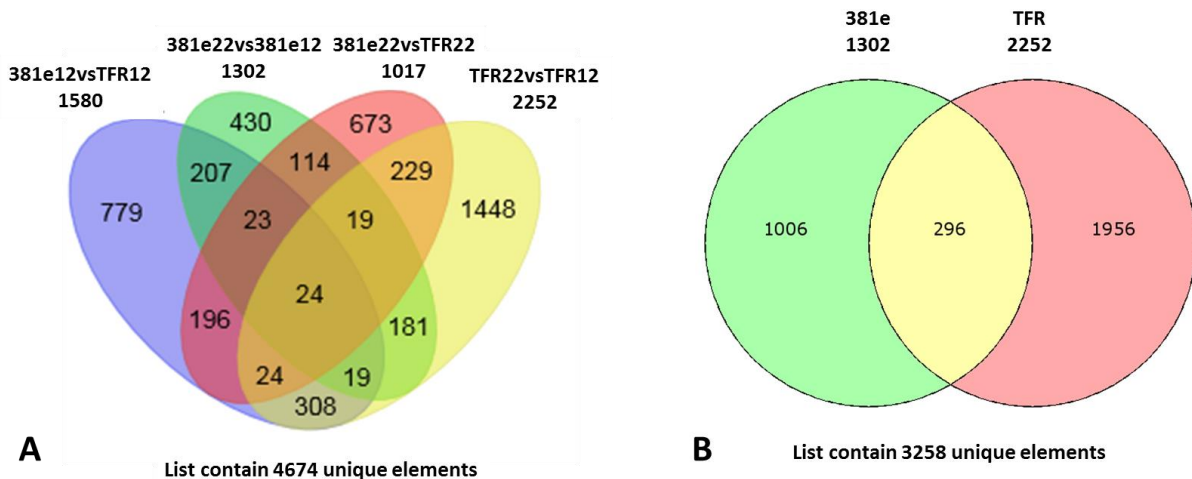
The 381-22dpi vs TF-22dpi comparison aims the identification of the transcripts differentially expressed (DE) in the 381-22 dpi line respect to TF-22dpi line. A total of 484 transcripts were found up-regulated in 381-22dpi line respect to TF-22dpi line and 533 down-regulated.

The TF-22dpi vs TF-12dpi comparison aims the identification of the transcripts differentially expressed (DE) in the TF-22dpi line respect to TF-12dpi line. A total of 1164 transcripts were up-regulated in TF-22dpi line respect to TF-12dpi line and 1088 down-regulated.

In addition to the comparisons already described, a more general analysis has been performed in order to identify DE transcripts in ‘381e’ genotype respect to ‘TFR’ genotype. In this experiment, the time was not treated as factor. This comparison aims the identification of the transcripts differentially expressed (DE) in the ‘381e’ genotype samples respect to ‘TFR’ genotype samples. In this comparison, a total of 350 transcripts were up-regulated in ‘381e’ genotype respect to ‘TFR’ genotype and 641 down-regulated.

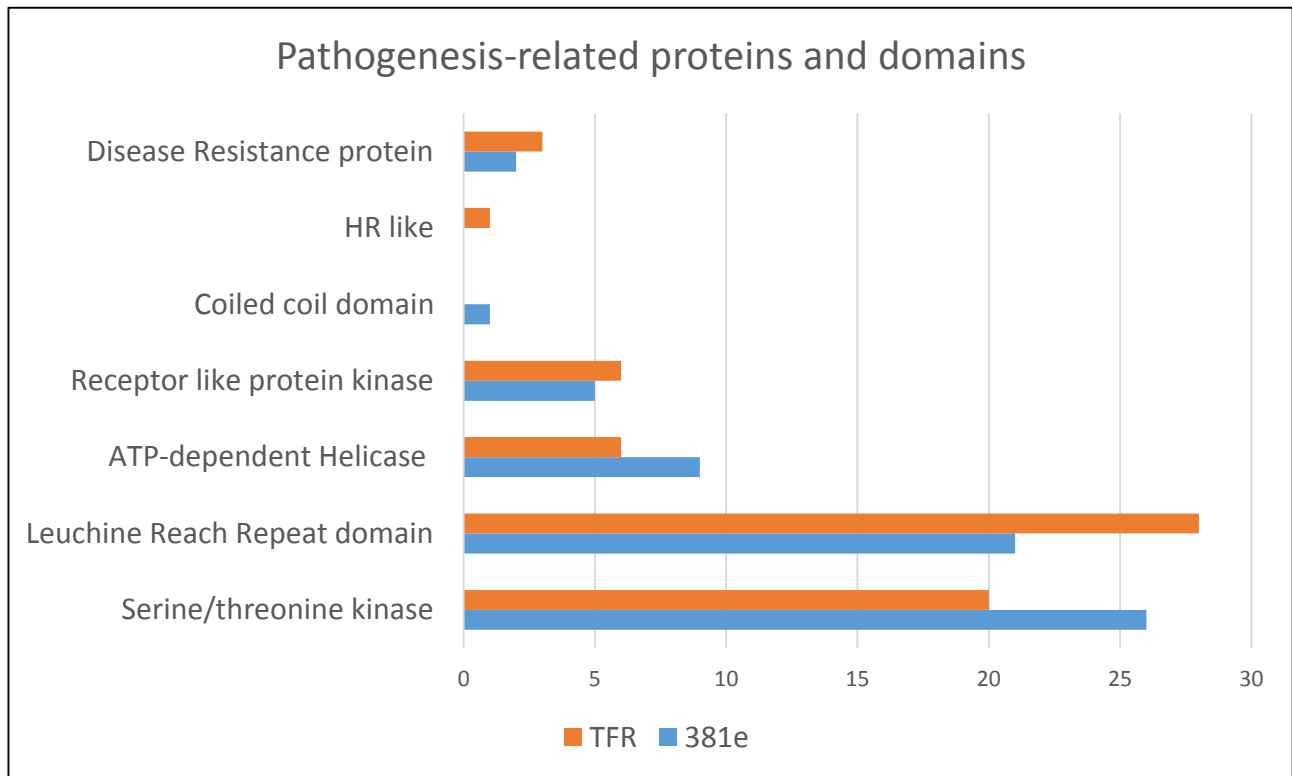
In order to identify the transcripts that are significantly expressed across the selected experimental conditions, the expressed genes for both genotype in both sampling times were recorded. Comparison of common and specific genes express in each studied condition was performed (Figure 3.13 A). A total number of 1,302 expressed genes were found for ‘381e’ resistant genotype, while ‘TFR’ susceptible genotype showed 2,252 expressed genes, regardless the sampling time. ‘381e’ presented 1,006 expressed genes absent in ‘TFR’ as well as ‘TFR’ presented 1,956 expressed genes absent in ‘381e’, while there were 296 common genes (Figure 3.13 B).

It is interesting to note, that in both diagram there is a high number of specific expressed genes and, in contrast, there is a reduced number of common genes for each sample, which may be related to the sampling time of inoculum.



**Figure 3.13:** Venn diagram showing the number of overlapping and non-overlapping expressed genes in the comparisons between the two sampling time of inoculum 12 and 22 dpi (A) and two genotypes (B).

Studying more thoroughly the unique genes expressed in the two genotypes (1006 and 1956), several pathogenesis-related proteins were identified. Those showed a relative abundance of transcripts coding for LRR domain and Serine/threonine kinase in both genotypes, meanwhile 9 ATP-dependent helicases was detected in the resistant genotype with respect the 6 found in susceptible genotype. More details were showed in figure 3.14.



**Figure 3.14:** Graphical representation of pathogenesis-related protein and domains identified in 1006 and 1956 genes expressed uniquely in '381e' and 'TFR' genotypes respectively.

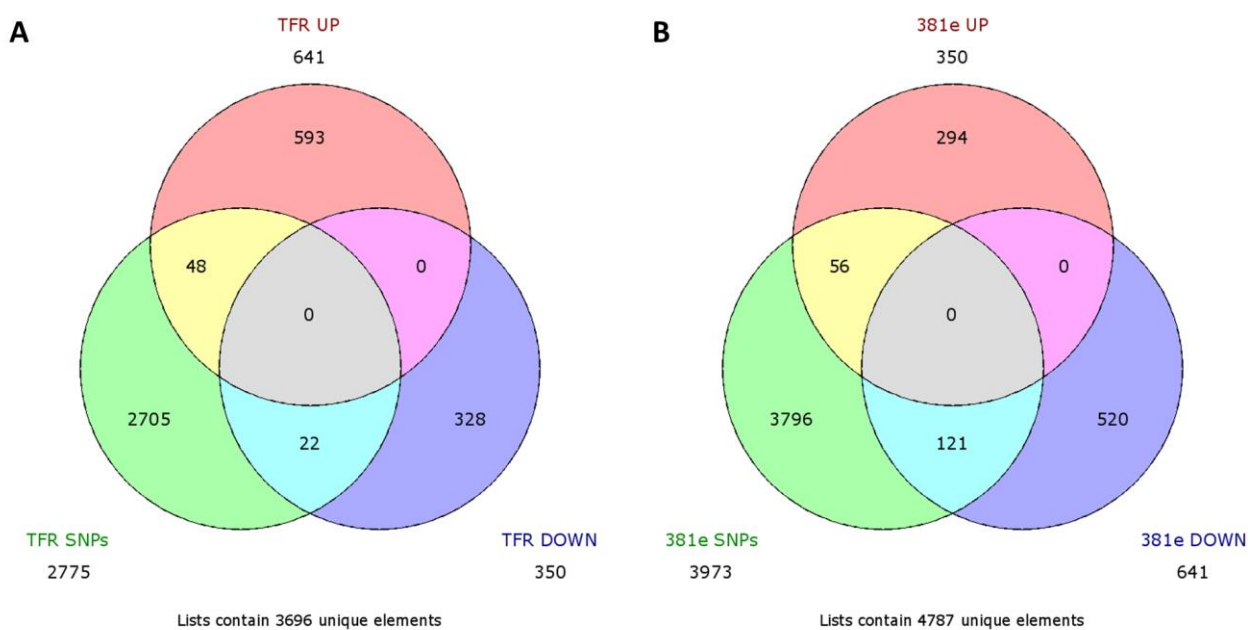
### 3.2.4 R-genes prediction

The *Cucurbita pepo* proteome was also interrogated by Matrix-R tools that, via HMM profiling, in order to retrieve, annotate and classify plant resistance genes. The protein sequences of the genes identified, were examined in order to underscore typical R gene domains. The output was a list of transcripts with a hit to a resistance domain . Out of 45016 only 2113 transcripts (belongs to the *C. pepo* reference transcriptome) have a score of "HIT" for at least one R-domain.



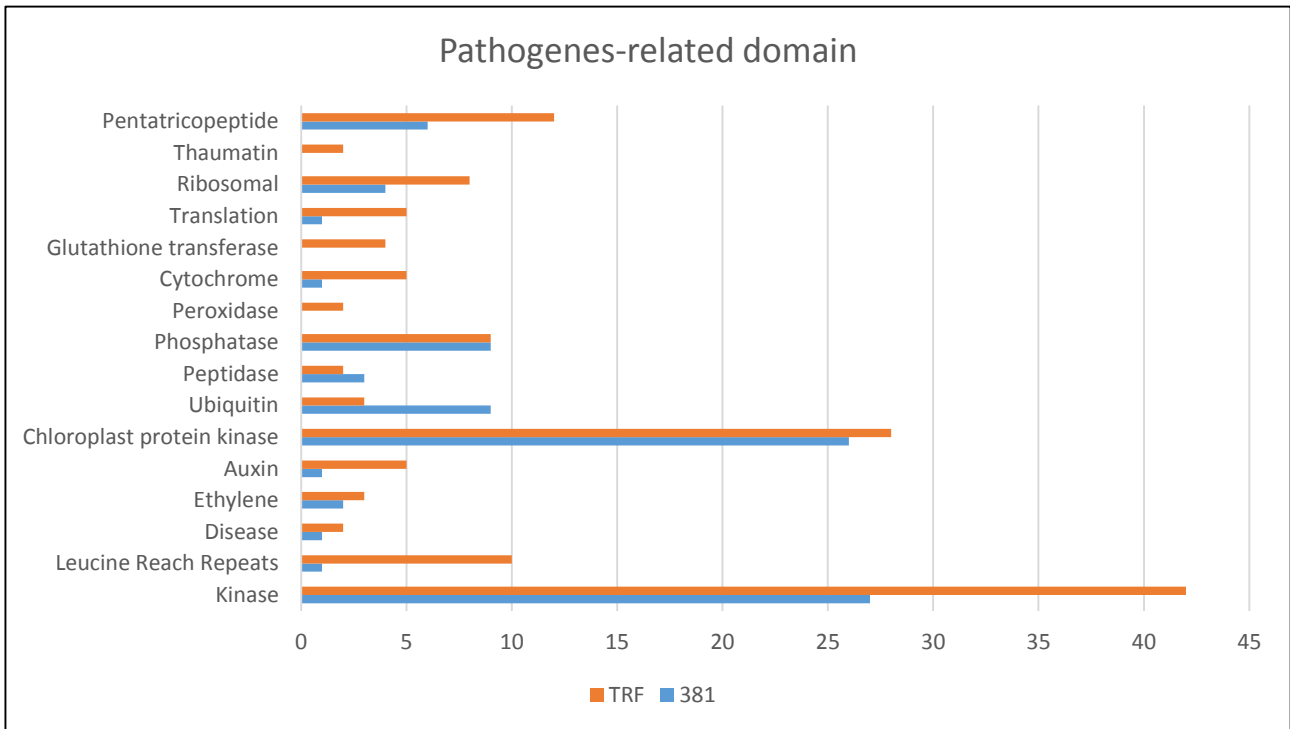
### 3.2.5 SNP calling analysis

In order to discover new SNP markers associated to genes involved in the defense mechanisms, a SNP calling analysis against the *Cucurbita pepo* reference genome was performed for transcripts of the susceptible ‘TFR’ and resistant ‘381e’ accessions. This analysis, allows to highlight the total number of 30275 and 6150 SNPs for ‘381e’ and ‘TFR’ respectively (figure 3.15). The lists of genes carrying SNPs obtained by SNP calling analysis were compared with the lists of genes up and down regulated of two genotypes highlighting the number of DE genes having SNPs as showed in Venn diagram reported in figure 3.11. In particular out of 350 up-regulated genes in ‘381e’, 56 contained SNPs, as well as of 641 down-regulated genes of ‘381e’ genotype, 121 carried least a SNP in comparison with the reference *C. pepo* genome. Likewise, out of 641 up-regulated and 350 down-regulated genes in ‘TFR’ genotype 48 and 22 respectively, contain at least a SNP.



**Figure 3.15:** Venn diagram showing the number of ‘TFR’ and ‘381e’ DEGs having SNPs compared to the reference *C. pepo* genome.

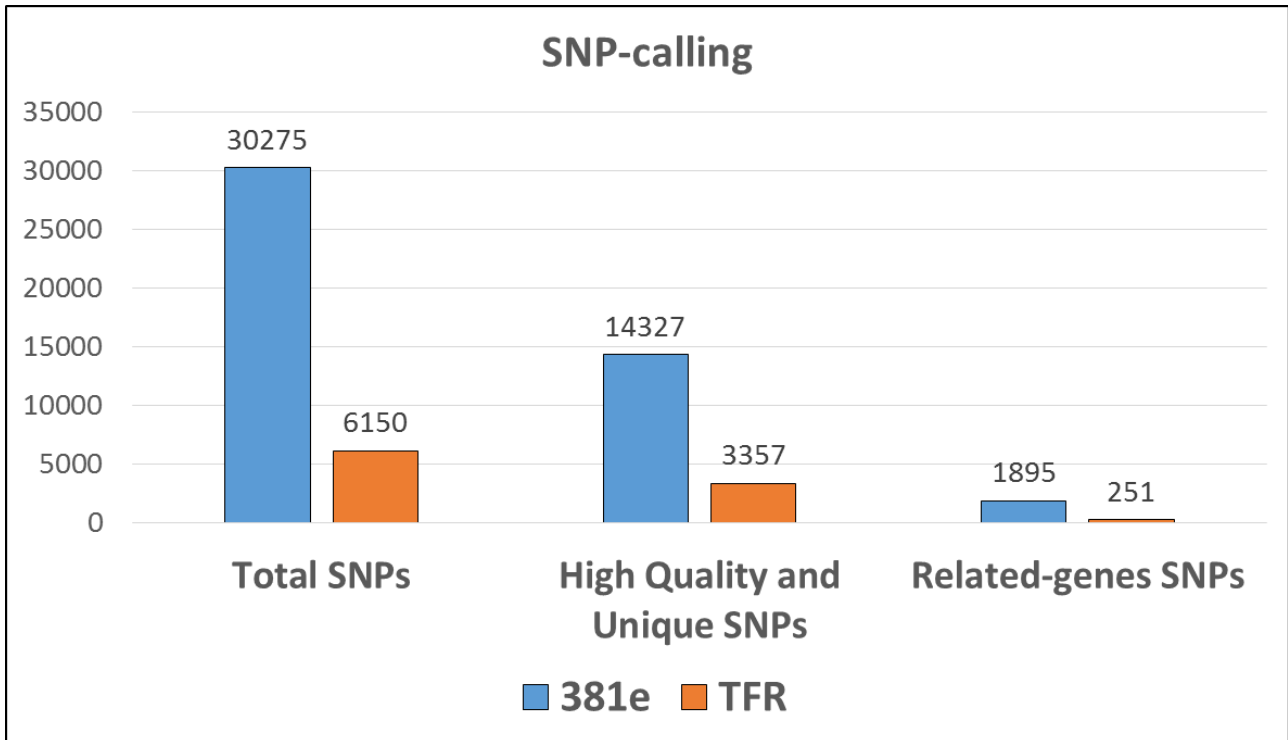
Out of total DE genes in ‘TFR’ and ‘381e’ carrying least a SNP marker, a high percentage coding for protein kinase (42 and 27), Chloroplast protein kinase (28 and 26), Leucine Rich Repeats (10 and 1) and other important protein such as Thaumatin, Ubiquitin, Phosphatase, but also hormones such as Auxin and Ethilene, typically involved in response reaction of plants to biotic stresses was found (Figure 3.16).



**Figure 3.16:** Graphical representation of pathogenesis-related protein identified in up and down regulated genes of ‘381e’ and ‘TFR’ genotypes carrying least a SNP marker.

Inconsistent and low quality/coverage SNPs were removed, identifying 3357 unique SNPs for ‘TFR’ and 14327 unique SNPs for ‘381e’ enlarged the number of SNPs markers available for *C. pepo* (Figure 3.17) and enhancing our possibility to find new markers associated with resistance to ZYMV. Interestingly, the presence of SNPs putatively associated to genes *Zym-1*, *Zym-2* and *Zym-3*, previously analyzed by the HRM assay on five different segregating populations was confirmed, demonstrating that they can be good candidate markers for resistance. Particularly, the SNP1 and SNP2 belongs to the 14327 SNPs of ‘381e’ genotype, while the SNP3 ranged in the 3357 SNPs of ‘TFR’ genotype.

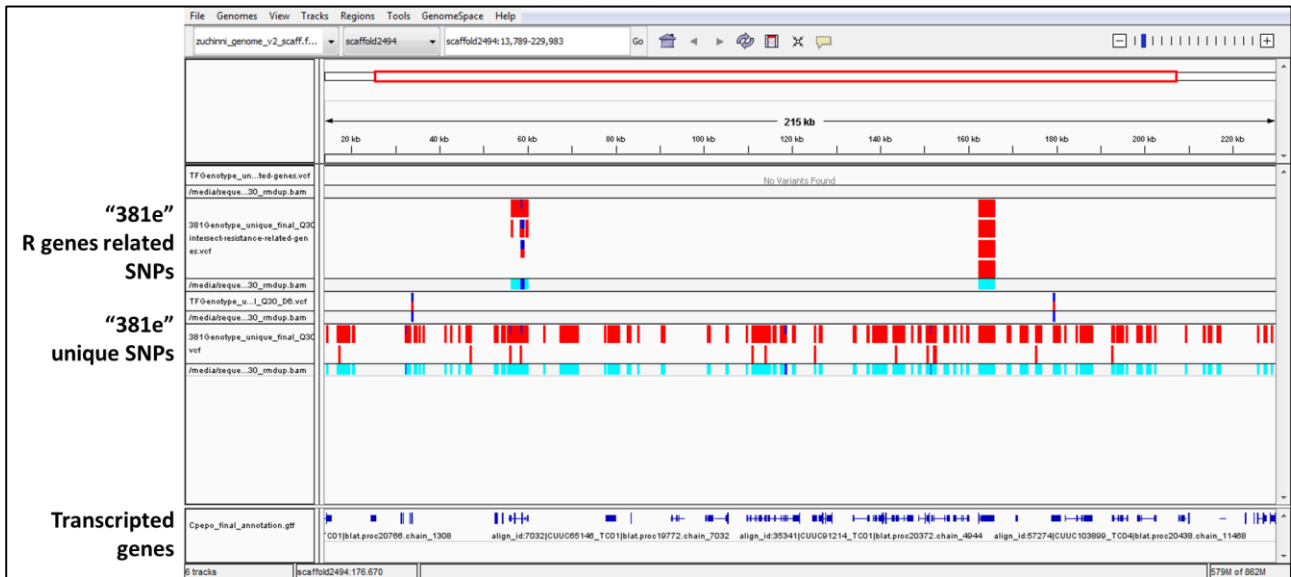
Next step was to select a pool of SNPs, targeting genes, between the two susceptible and resistant isogenic lines (Figure 3.17). The SNP calling was performed on the transcriptome of the two genotypes taking into account only the expressed genes in a given time, excluding from the assessment all the SNPs belongs the not expressed genes and the intronic SNPs. This procedure excluded intron alleles that can affect disease such as transcription regulators, leaving analysis to genome elements of large effect.



**Figure 3.17:** Graphical representation of total SNPs, unique/high quality SNPs and R gene-related SNPs identified in ‘381e’ and ‘TFR’ transcriptomes compared to *C. pepo* genome.

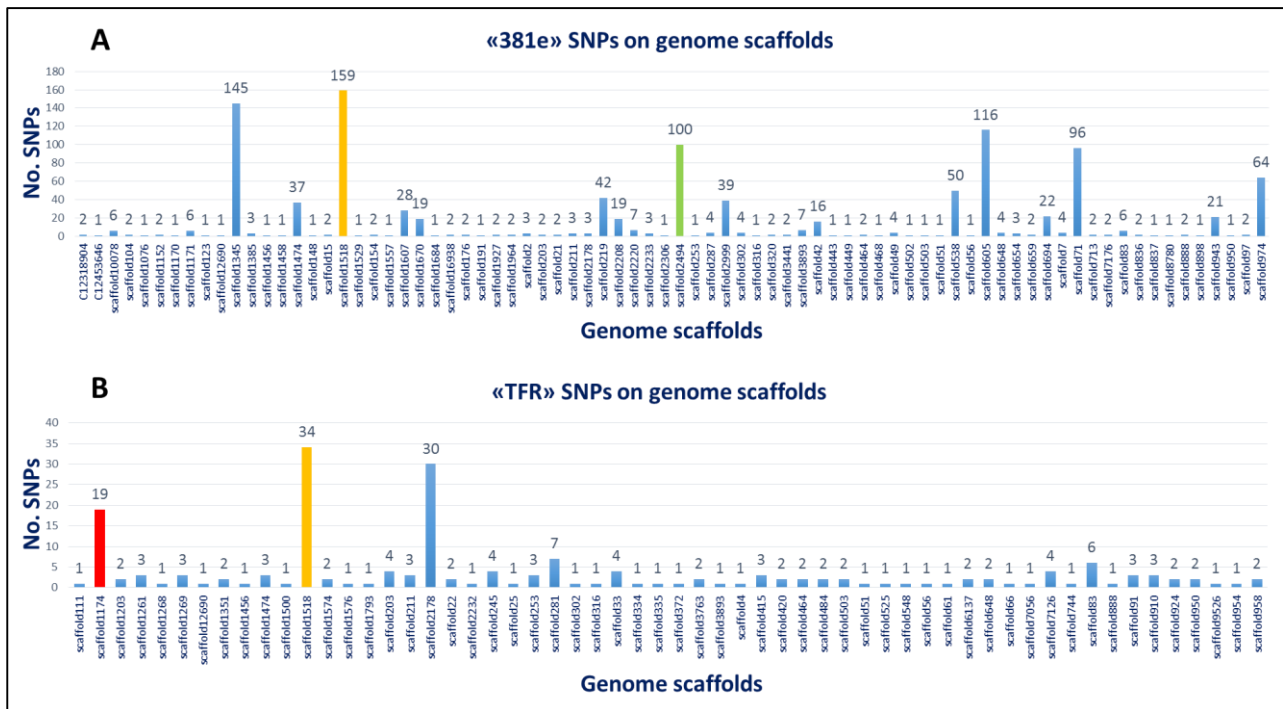
### 3.2.6 Identification of new candidate markers for ZYMV resistance

In order to identify new candidate markers for ZYMV resistance we merged SNP calling and Matrix-R data. Unique SNPs for each genotype were filtered in order to extract only SNPs located on resistance-related transcripts highlighted by Matrix-R analysis. Results showed that out of 3357 and 14327 unique SNPs identified in ‘TFR’ and ‘381e’ genotypes respectively, 251 SNPs and 1895 SNPs, fall in R genes or in domains involved in defense mechanisms. Moreover, once a SNP was placed on a scaffold, we were able to identify putative pathogen recognition genes localized in the same chromosome region, allowing us to focus on a restricted set of genes. A functional R gene map based on the co-localization of pathogen recognition putative proteins with R loci linked markers on scaffold 2494 is shown in figure 3.18. The map shows the candidate pathogen recognition genes localized on scaffold and the distribution of ‘381e’ genotype SNPs .



**Figure 3.18:** Maps of putative R genes on the scaffold 2494 and their related SNPs of ‘381e’ genotype, meanwhile there was not found ‘TFR’ related SNPs in this scaffold.

The visualization of information linked to a locus is a fundamental step in interpreting data and in suggesting correlations between genetic and genomic data. Indeed, 100 SNPs of ‘381e’ are falling in only three R genes located on scaffold 2494 (figure 3.19 A), the same scaffold to which belongs the SNP1 associated to major gene *Zym-1*. In parallel, out of 251 SNPs discovered in the ‘TFR’ genotype, 19 falling in three R genes of scaffold 1174 (figure 3.19 B), that holds also SNP3 putatively associated to *Zym-3*. While, we not found SNPs falling in R genes located on scaffold 1377 where SNP2 was identified. In addition, it is important to note that the distribution of SNP markers allowed us to identify a peak of SNPs in both genotypes in scaffold1518 with a higher number in the genotype ‘381e’ (159 SNPs) versus genotype ‘TFR’ (34 SNPs).



**Figure 3.19:** R genes-related SNPs distribution on genome scaffolds (SNPs vs scaffolds genome). On the scaffold2494 (green bar in panel A) 100 SNPs of ‘381e’ falling on only three R genes. On the scaffold1174 (red bar in panel B) 19 SNPs of ‘TFR’ falling on three R genes.

#### 4. DISCUSSIONS

The use of resistant varieties in agriculture provides both economic and ecological benefits, considering the reduced use of pesticides needed for their cultivation. The process of introduction of resistance to pathogens in crop species, particularly those polygenic, takes long time and are not always stable over time. Molecular genetics and modern biotechnology techniques can be very useful to identify pathogen resistance genes in related wild species and to speed up the introgression in cultivated varieties. Several studies rely on use of advanced segregating populations for accelerating the introgression of genes of interest into a genotype. This process is also called gene pyramiding, defined as a method aimed to assembling multiple desirable genes from multiple parents into a single genotype (Ye and Smith, 2008).

Main goal of this work was to study the association of molecular markers with resistance genes to ZYMV, in order to make easier, faster and economically sustainable the introgression of these genes in traditional Campania variety 'San Pasquale'. The identification of genetic resource containing the pathogen resistance genes and the knowledge of mode of inheritance is a very important task for carrying out a breeding program. For this purpose, several *in vivo* assays were conducted to evaluate parental lines behavior and to estimate segregation ratio of genes involved in resistance to ZYMV in progenies.

The phenotypic *in vivo* assays, performed on two populations 'F<sub>2</sub> TRF' and 'F<sub>2</sub> SPQ' and three derived 'F<sub>2</sub> SPQ' backcross populations, allowed us to screen resistant and susceptible samples to a ZYMV breaking strain found in the major growing regions of the crop. *In vivo* tests was conducted in environmental controlled conditions in order to ensure a high stress condition. The F<sub>2</sub> segregation populations displayed the 45:19 expected ratio, obtained assuming three complementary dominant genes involved in resistance to ZYMV, reported in previous segregating populations (Paris and Cohen, 2000). However, the phenotypic assays performed on BC<sub>1</sub>-F<sub>2</sub> populations not respected the expected ratio based on Paris and Cohen (2000) assumptions. It is important to note that the genotypes carrying AAbbcc and Aabbcc alleles results resistant. If we hypothesized a major gene control the resistance the observed ratio fitted with the 1:1 expected ratio in the BC<sub>1</sub>-F<sub>2</sub> populations. Following this hypothesis we recalculated the segregation in the F<sub>2</sub> populations showing to be in agreement also with 3:1 ratio. This finding highlighted a very important role of *Zym-1* in conferring resistance. The two complementary *Zym-2* and *Zym-3* genes could be nevertheless indispensable for ensuring the expression of tolerance in late stages of development. In the study of resistance to TMV and *Phytophthora capsici* in pepper, important epistatic effects between a major and three minor QTLs (with or without additive effects) were detected (Lefebvre and Palloix, 1996).

The main role of *Zym-1* gene was also confirmed by the fact that the percent of co-segregation with the phenotype was lower when the genotypes having only the *Zym-1* was considered susceptible especially in the BC<sub>1</sub>-F<sub>2</sub> populations (Table 3.2) and by Mantel test (Mantel, 1967) conducted on the (dis)similarity or distance matrices obtained from two F<sub>2</sub> segregating populations datasets.

Molecular analysis performed with the High Resolution Melting Real-Time PCR, based on each single SNP marker linked to the three resistance genes, allowed us to predict the allelic configurations of the considered loci. Moreover, statistical analysis based on the SNPs dataset showed a match between the observed and expected ratio of segregation in four segregating populations, confirming a higher co-segregation of *Zym-1* with the resistant phenotype respect to minor genes. The CAPS marker analysis discriminating susceptible from resistant subjects, confirmed the results obtained from the HRM analysis showing a stronger co-segregation of SNP1 with resistant phenotypes.

The SNPs and CAPS markers obtained in the study enriched molecular markers of *C. pepo*. CAPS markers are widely used in molecular assisted selection programs and in diagnostic assays of vegetable crops. The relevance of a CAPS marker for MAS was demonstrated in different breeding programs (Langella *et al.*, 2004; Suh *et al.*, 2013; Joshi *et al.*, 2013). The rapidity, accuracy and the cost efficiency of our CAPS analysis, makes it a useful tool to improve the traditional selection methods of breeding.

Molecular markers develop in this thesis can help future genetic linkage and quantitative trait loci analysis, and will be essential to speed up the process of breeding new and better adapted squash varieties (Blanca *et al.*, 2011). This work highlighted three SNP markers linked to three genes putatively responsible of the ZYMV resistance. The ZYMV resistance genetic segregation studied both at phenotypic and genotypic level not showed a clear framework of action of genes involved. Therefore, a different mode of inheritance or the presence of additional genes involved in resistance process could be required. For example, Pachner *et al.* (2014) theorizes a combined dosage effect of the genes that conferring determinate phenotypes. The dosage effect of genes could explain the recovery cases after 25<sup>th</sup> dpi of same analyzed samples. Gomez *et al.* (2009) stated that 51% of resistance traits to plant viruses were dominant, 35% were recessive, and the remainder were more complex (incomplete dominance or dose-dependent).

Marker-based selection reduces/eliminates extensive phenotyping, reducing the selection times, excessive costs due to the large spaces needed for plants and pathogens growth, but also eliminating the errors owed to visual assessments often wrong because made subjectively by the operator. The rate of success is likely to increase due to developments in gene-based marker development, more

efficient quantitative trait locus mapping procedures, and lower cost genotyping systems (Xu and Crouch, 2008). However, in many cases, particularly regarding complex traits, it is difficult to assign a genotype to specific symptoms. Therefore, the most reasonable approach towards merging the various genes for resistance into one genotype is a step-wise combination of phenotypic and marker-assisted selection (Pachner *et al.*, 2014). Development of high throughput precision phenotyping systems for resistance gene mapping, improved understanding of genotype by environment interaction and epistasis, and development of publicly available computational tools tailored to the needs of molecular breeding programs are required (Xu and Crouch, 2008).

Genomic analysis appears to be increasingly useful in the associated markers discovery and resistance genes study. In order to deeply investigate genes-markers association several modern bioinformatics tools were used. To obtain a better overview of R genes co-localized next to our markers we performed a protein prediction and domain search through InterProScan tool (Andolfo *et al.*, 2014). Our attention was focused to the research of protein belonging to resistance gene classes. Proteins belonging to CC, NB-ARC, LRR, ATP-dependent RNA helicase and other typical R-domains were found nearby the SNP1 markers, suggesting the SNP1 association with one of three resistance genes, most probably the major gene *Zym-1*. R genes and/or gene silencing mediated active resistance mechanisms. The NB-ARC system minimizes the cost of defense for the plant, as multiple NB-ARC-LRR proteins can be maintained at a low level in the absence of a pathogen and rapidly induced under pathogen attack through miRNA regulation (Li *et al.*, 2011). Pathogen-encoded suppressors of RNA-silencing mechanisms might result in the induction of multiple NB-ARC-LRR defense proteins (Shivaprasad *et al.*, 2012). The RNA helicase are enzymes that use energy derived from the hydrolysis of a nucleotide triphosphate to unwind double-stranded RNAs (de la Cruz *et al.*, 1999). RNA helicases have been shown to be involved in every step of RNA metabolism, including nuclear transcription, pre-mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay, and organellar gene expression (Tanner *et al.*, 2001). Xu *et al.* (2013) reported the involvement of RNA helicases in response to stress and in plant growth and development.

Plant-expressed molecules during pathogen challenging can give insights into the underlying defense mechanisms. Plants have evolved a defense system against pathogens that involves the regulation of gene expression, cascade signaling activation, hormone balancing and synthesis of defensive metabolites (Mithöfer and Boland, 2012). The advent of next-generation sequencing (NGS) has revolutionized genomic and transcriptomic approaches to biology. These new



sequencing tools are also valuable for the discovery, validation and assessment of genetic markers in populations (Davey *et al.*, 2011).

In order to discovery new genes and their associated markers involved in defense mechanisms to ZYMV in *Cucurbita pepo*, the transcriptomes of one resistant and one susceptible genotype collected in two different time post-inoculation with ZYMV were sequenced. Through the RNA-seq analysis, transcriptional changes were monitored in zucchini upon inoculation with virus. Several researches in this field were conducted in order to identify genes connected to resistance (Li *et al.*, 2013; Sookruksawong *et al.*, 2013; Xue *et al.*, 2013; Zeng *et al.*, 2013). RNA-seq technology, it has also been proving to be an extremely useful tool for identifying single nucleotide polymorphisms, which may be also used to develop high density SNPs chips for studies concerning the genome wide association (GWAS) and to build high density linkage maps (Baranski *et al.*, 2014; Yu *et al.*, 2014). Furthermore, SNPs can be used as markers in order to distinguish allelic transcripts whilst studying the allele-specific expression (Bell and Beck, 2009).

At our knowledge, this is the first work actually available reporting study of zucchini genotypes response to ZYMV attacks through RNA-seq assay. The major limit to analyze transcriptomic data is due to the very fragmented and poorly annotated reference transcriptome of *C. pepo* currently available. Therefore, in this work has been carried out the functional annotation of reference transcriptome version 3, allowing us to obtain more detailed information useful for B2GO annotation of differentially induced genes in our analysis. In genomes that have been poorly studied, sequence availability is the limiting factor for the discovery of SNPs (Blanca *et al.*, 2011). The functional annotation of unique genes expressed in the two genotypes (1006 and 1956), allowed us to identify several transcripts coding for LRR domains and Serine/threonine protein kinase in both genotypes, meanwhile a higher presence of ATP-dependent helicases was detected in the resistant genotype, enlarging the number of SNPs available for this species .

In order to discovery new markers and R genes, a SNP calling and Matrix-R analysis were performed on the sequenced transcriptome of resistance and susceptible genotypes, detecting the genotype-specific SNPs and R genes related SNPs. Merely in the 2011, Blanca *et al.* conducted the first work of SNPs discovery using Expressed sequenced tags (ESTs). SNP calling in the context of NGS data analysis is performed against reference genome, typically including an associated confidence score or statistical evidence metric (Li *et al.*, 2008).

In our work SNP calling was performed on the transcriptome of the two genotypes taking into account only the expressed genes in a given time, excluding from the assessment all the SNPs belongs the not expressed genes and the intronic SNPs. This procedure excluded intron alleles that can affect disease such as transcription regulators, leaving analysis to genome elements of large

effect. While some correlation exists between exon to intron variation, only whole genome sequencing would be able to capture the source of all relevant SNPs. The SNP calling allowed us to detect unique SNPs comparing the two transcriptomes against the reference genome of *Cucurbita pepo*. Stringent quality criteria were used for distinguish sequence variations from sequencing errors and mutations introduced during the cDNA synthesis step. By applying these criteria were identified unique SNPs both for ‘TFR’ and ‘381e’, enhancing our possibility to find new resistance genes associated markers.

R genes prediction with Matrix-R assay, based on the reference transcriptome B2GO annotation, allowed to predict genes putatively involved in defense response. Merging information obtained from SNP calling and Matrix-R assay, we extracted the unique SNPs related to R-gene, discovering a high SNPs concentration (peak) in correspondence of genome portion where are located *Zym-1* and *Zym-3* genes. Detected SNPs could be defined “putatively ZYMV resistance associated”, considering that the analysis was performed on resistant and susceptible isogenic lines dataset. Interestingly a further peak of SNPs in both genotypes was identified in scaffold1518 although in higher numbers in the genotype ‘381e’.

Gene expression differences highlighted in this work between the resistant and susceptible isogenic lines was used for a comprehensive study of expressed genes and for discover expressed SNPs in zucchini in response to ZYMV. Comparing the lists of genes containing SNP markers with lists of genes of ‘TFR’ and ‘381e’ induced or repressed by the virus attack, were highlighted genes differentially expressed in the two genotypes having one or more polymorphisms. The identification and functional annotation of polymorphic genes identified represent a useful resource to comprehend mechanisms determining complex traits and, consequently, to develop programs aiming the genetic improvement (Baranski *et al.*; Yu *et al.*, 2014).

## 5. CONCLUSIONS

Resistant varieties produce significant economic and ecological benefits by reducing inputs of pesticides and, therefore, their use is recommended in integrated and biological cultivation systems. The process of disease resistance transfer by classical breeding in the cultivated varieties requires long times, some resistance genes are difficult to introduce and are not always stable over time. The use of molecular markers associated to the genes controlling resistance characters may facilitate the selection.

In this work a study of ZYMV resistance response in zucchini at genetic and genomic level was performed.

- Phenotypic and genotypic analysis, performed on five segregating populations, highlighted a complex inheritance of resistance investigated with the involvement of at least three genes (*Zym-1*, *Zym-2* and *Zym-3*) in the defense mechanisms. The analyses showed that *Zym-1* has a main role in the zucchini response to virus and that *Zym-2* and *Zym-3* could be required for displaying a durable resistance in later stages of plants development.
- Molecular segregation analysis on 3 SNP loci strongly suggested a relationship between these SNPs and resistance genes *Zym-1*, *Zym-2* and *Zym-3*. A CAPS marker developed on SNP1 validated the association between the SNP1 and the *Zym-1* resistance gene and it is available to improve selection methods employed at breeding companies.
- Transcriptomic analysis allowed to observe above 3000 differentially expressed genes between zucchini isogenic lines for resistance to ZYMV highlighting additional resistance gene candidates. Thanks to the functional annotation performed on the DEGs and comparison of the differentially expressed genes an explanatory model of the response process of zucchini to ZYMV could be build up.
- Transcript SNP calling enlarged the number of marker available for *C. pepo* and reconfirmed the presence of SNPs, putatively associated to genes *Zym-1*, *Zym-2* and *Zym-3*, in the two isogenic lines ‘True French’ and ‘381e’. SNPs localized near R genes evidenced SNPs associated to typical resistance domain expressed in differential mode in resistant and susceptible genotypes.

- The R protein domains prediction showed a close relationship of the SNP1 with a CNL and an ATP-dependent RNA helicase, suggesting the SNP1 association with one of three resistance genes, most probably the major gene *Zym-1*.
- Considering that the local variety ‘San Pasquale’, susceptible to ZYMV, has a good marketability in Campania region, results obtained in this thesis can be useful for the genetic improvement of this variety.

## 6. REFERENCES

- Akhunov, E., Nicolet, C., & Dvorak, J. (2009). Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theoretical and Applied Genetics*, 119(3), 507-517.
- Albar, L., Bangratz-Reyser, M., Hébrard, E., Ndjiondjop, M. N., Jones, M., & Ghesquière, A. (2006). Mutations in the eIF (iso) 4G translation initiation factor confer high resistance of rice to Rice yellow mottle virus. *The Plant Journal*, 47(3), 417-426.
- Ali, M. A. (1950). Genetics of resistance to the common Bean mosaic virus (Bean virus 1) in the Bean (*Phaseolus vulgaris* L.). *Phytopathology*, 40(1), 69-79.
- Allard, I. (1960). *Principles of Plant Breeding*, Chapter 6 through Chapter 9, University of California, Davis.
- Andolfo, G., Sanseverino, W., Aversano, R., Frusciante, L., & Ercolano, M. R. (2014). Genome-wide identification and analysis of candidate genes for disease resistance in tomato. *Molecular breeding*, 33(1), 227-233.
- Anido, F. L., Cravero, V., Asprelli, P., Firpo, T., García, S. M., & Cointy, E. (2004). Heterotic patterns in hybrids involving cultivar-groups of summer squash, *Cucurbita pepo* L. *Euphytica*, 135(3), 355-360.
- Baggett, J. R. (1972). Open growth habit in summer squash. *HortScience*.
- Bagnaresi, P., Biselli, C., Orrù, L., Urso, S., Crispino, L., Abbruscato, P., ... & Valè, G. (2012). Comparative transcriptome profiling of the early response to *Magnaporthe oryzae* in durable resistant vs susceptible rice (*Oryza sativa* L.) genotypes. *PloS one*, 7(12), e51609.
- Baranski, M., Gopikrishna, G., Robinson, N. A., Katneni, V. K., Shekhar, M. S., Shanmugakarthish, J., ... & Ponniah, A. G. (2014). The development of a high density linkage map for black tiger shrimp (*Penaeus monodon*) based on cSNPs. *PloS one*, 9(1).
- Bell, C. G., & Beck, S. (2009). Advances in the identification and analysis of allele-specific expression. *Genome Med*, 1(5), 56.
- Bennett, C. D., Campbell, M. N., Cook, C. J., Eyre, D. J., Nay, L. M., Nielsen, D. R., ... & Bernard, P. S. (2003). The LightTyper: high-throughput genotyping using fluorescent melting curve analysis. *Biotechniques*, 34(6), 1288-1295.

- Blackman, R. L., & Eastop, V. F. (2000). *Aphids on the world's crops an identification and information guide*. 2000. England, John Wiley & Sons.
- Blanca, J., Cañizares, J., Roig, C., Ziarolo, P., Nuez, F., & Picó, B. (2011). Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC genomics*, 12(1), 104.
- Borneman, A. R., Forgan, A. H., Pretorius, I. S., & Chambers, P. J. (2008). Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS yeast research*, 8(7), 1185-1195.
- Buttigieg, P. L., & Ramette, A. (2014). A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS microbiology ecology*, 90(3), 543-550.
- Campbell, P. J., Stephens, P. J., Pleasance, E. D., O'Meara, S., Li, H., Santarius, T., ... & Futreal, P. A. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature genetics*, 40(6), 722-729.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., ... & Gariboldi, M. (2005). The transcriptional landscape of the mammalian genome. *Science*, 309(5740), 1559-1563.
- Chelkowski, J., & Stepien, L. (2001). Molecular markers for leaf rust resistance genes in wheat. *J. Appl. Genet*, 42(2), 117-126.
- Chen, X., Levine, L., & Kwok, P. Y. (1999). Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Research*, 9(5), 492-498.
- Chen, J., Kim, Y. C., Jung, Y. C., Xuan, Z., Dworkin, G., Zhang, Y., ... & Wang, S. M. (2008). Scanning the human genome at kilobase resolution. *Genome research*, 18(5), 751-762.
- Ching, A. D. A., Caldwell, K. S., Jung, M., Dolan, M., Smith, O. S., Tingey, S., ... & Rafalski, A. J. (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC genetics*, 3(1), 19.
- Cho, M. H., Ciulla, D., Klanderma, B. J., Raby, B. A., & Silverman, E. K. (2008). High-resolution melting curve analysis of genomic and whole-genome amplified DNA. *Clinical chemistry*, 54(12), 2055-2058.
- Chomczynski, P., & Mackey, K. (1995). Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide-and proteoglycan-rich sources. *Biotechniques*, 19(6), 942-945.

- Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., ... & Grimmond, S. M. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature methods*, 5(7), 613-619.
- Close, T. J., Bhat, P. R., Lonardi, S., Wu, Y., Rostoks, N., Ramsay, L., ... & Waugh, R. (2009). Development and implementation of high-throughput SNP genotyping in barley. *BMC genomics*, 10(1), 582.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674-3676.
- Croxford, A. E., Rogers, T., Caligari, P. D., & Wilkinson, M. J. (2008). High-resolution melt analysis to identify and map sequence-tagged site anchor points onto linkage maps: a white lupin (*Lupinus albus*) map as an exemplar. *New Phytologist*, 180(3), 594-607.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499-510.
- de la Cruz, J., Kressler, D., & Linder, P. (1999). Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends in biochemical sciences*, 24(5), 192-198.
- Desbiez, C., Wipf-Scheibel, C. & Lecoq, H. (2002). Biological and serological variability, evolution and molecular epidemiology of Zucchini yellow mosaic virus (ZYMV, Potyvirus) with special reference to the Caribbean islands. *Virus Res* 85, 5–16.
- Deulvot, C., Charrel, H., Marty, A., Jacquin, F., Donnadiou, C., Lejeune-Hénaut, I., ... & Aubert, G. (2010). Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC genomics*, 11(1), 468.
- Duchesne, A. N. (1786). *Essai sur l'histoire naturelle des courges*. Paris: Panckoucke, 7–11, 15.
- Esteras, C., Gómez, P., Monforte, A. J., Blanca, J., Vicente-Dólera, N., Roig, C., ... & Picó, B. (2012). High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait loci mapping. *BMC genomics*, 13(1), 80.
- Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome research*, 8(3), 186-194.

- Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome research*, 8(3), 175-185.
- Fan, J. B., Oliphant, A., Shen, R., Kermani, B. G., Garcia, F., Gunderson, K. L., ... & Chee, M. S. (2003, January). Highly parallel SNP genotyping. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 68, pp. 69-78). Cold Spring Harbor Laboratory Press.
- Fraser, R. S. S., & Van Loon, L. C. (1986). Genes for resistance to plant viruses. *Critical reviews in plant sciences*, 3(3), 257-294.
- Fulton, T.F., Chunwongse, J., Tanksley, S.D. 1995. Micropep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol. Biol. Rep.* 13:207-209
- Gal-On, A. (2007). Zucchini yellow mosaic virus: insect transmission and pathogenicity – the tails of two proteins. *Mol Plant Pathol* 8, 139–150.
- Gao, L., Tu, Z. J., Millett, B. P., & Bradeen, J. M. (2013). Insights into organ-specific pathogen defense responses in plants: RNA-seq analysis of potato tuber-*Phytophthora infestans* interactions. *BMC genomics*, 14(1), 340.
- Gold, M. V. (1994). Sustainable agriculture: definitions and terms. *Special reference briefs (USA)*. no. 94-05.
- Gómez, P., Rodríguez-Hernández, A. M., Moury, B., & Aranda, M. A. (2009). Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. *European journal of plant pathology*, 125(1), 1-22.
- Grafton-Cardwell, E. E., Smith, R. F., Valencia, J. and Farrar, C. A. (1996) Occurrence of mosaic viruses in melons in the central valley of California. *Plant Dis.* 80, 1092–1097.
- Grassi, F., Labra, M., & Minuto, L. (2006). Molecular diversity in Ligurian local races of common bean (*Phaseolus vulgaris* L.). *Plant Biosystems*, 140(1), 17-20.
- Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., ... & Celniker, S. E. (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature*, 471(7339), 473-479.
- Gunderson, K. L. (2009). Whole-genome genotyping on bead arrays. In *DNA Microarrays for Biomedical Research* (pp. 197-213). Humana Press.



- Gupta, P. K., Rustgi, S., & Mir, R. R. (2008). Array-based high-throughput DNA markers for crop improvement. *Heredity*, 101(1), 5-18.
- Hall, T. J. (1980). Resistance at the Tm-2 locus in the tomato to tomato mosaic virus. *Euphytica*, 29(1), 189-197.
- Hansen, J. W. (1996). Is agricultural sustainability a useful concept? *Agricultural systems*, 50(2), 117-143.
- Hansen, J. W., & Jones, J. W. (1996). A systems framework for characterizing farm sustainability. *Agricultural Systems*, 51(2), 185-201.
- Hermoso Pulido, T., Contaldi, F., Palomba, F., Nieri, D., Paytuví, A., Anzar, I., Aiese Cigliano, R., Ercolano, M. R., Sanseverino, W. (2014). Matrix Reloaded: a new tool for plant resistance genes annotation.
- Herrmann, M. G., Durtschi, J. D., Wittwer, C. T., & Voelkerding, K. V. (2007). Expanded instrument comparison of amplicon DNA melting analysis for mutation scanning and genotyping. *Clinical chemistry*, 53(8), 1544-1548.
- Hoffmann, M., Hurlbaeus, J., & Weilke, C. (2007). High resolution melting curve analysis on the LightCycler (R) 480 PCR system. *Nat. Methods*, AN17-AN18.
- Horrigan, L., Lawrence, R. S., & Walker, P. (2002). How sustainable agriculture can address the environmental and human health harms of industrial agriculture. *Environmental health perspectives*, 110(5), 445.
- Hung, C. C., Lee, C. N., Chang, C. H., Jong, Y. J., Chen, C. P., Hsieh, W. S., ... & Lin, W. L. (2008). Genotyping of the G1138A mutation of the FGFR3 gene in patients with achondroplasia using high-resolution melting analysis. *Clinical biochemistry*, 41(3), 162-166.
- Hyten, D. L., Song, Q., Choi, I. Y., Yoon, M. S., Specht, J. E., Matukumalli, L. K., ... & Cregan, P. B. (2008). High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theoretical and Applied Genetics*, 116(7), 945-952.
- International HapMap Consortium I: The International HapMap Project. *Nature* 2003, 426(6968):789-96.
- Johansen, E., Edwards, M. C. & Hampton, R. O. (1994). Seed transmission of viruses: current perspectives. *Annu Rev Phytopathol* 32, 363–386.

- Joshi, B. K., Panthee, D. R., Louws, F. J., Yencho, G. C., Sosinski, B., & Arellano, C. (2013). RAPD Markers Linked to Late Blight Resistance in Tomato. *Nepal Journal of Science and Technology*, 14(1), 1-14.
- Kanazin, V., Talbert, H., See, D., DeCamp, P., Nevo, E., & Blake, T. (2002). Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). *Plant Molecular Biology*, 48(5-6), 529-537.
- Kang, B. C., Yeam, I., & Jahn, M. M. (2005). Genetics of plant virus resistance. *Annu. Rev. Phytopathol.*, 43, 581-621.
- Katis, N. I., Tsitsipis, J. A., Lykouressis, D. P., Papapanayotou, A., Kokinis, G. M., Perdakis, D. C. & Manoussopoulos, I. N. (2006). Transmission of Zucchini yellow mosaic virus by colonizing and non- colonizing aphids in Greece and new aphid vectors of the virus. *J Phytopathol* 154, 293–302.
- Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., ... & International Sheep Genomics Consortium. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PloS one*, 4(3), e4668.
- Kim, P. M., Lam, H. Y., Urban, A. E., Korbel, J. O., Affourtit, J., Grubert, F., ... & Gerstein, M. B. (2008). Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome research*, 18(12), 1865-1874.
- Komori, T., & Nitta, N. (2005). Utilization of the CAPS/dCAPS method to convert rice SNPs into PCR-based markers. *Breeding Science*, 55(1), 93-98.
- Konieczny, A., & Ausubel, F. M. (1993). A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *The Plant Journal*, 4(2), 403-410.
- Korbel, J. O., Urban, A. E., Affourtit, J. P., Godwin, B., Grubert, F., Simons, J. F., ... & Snyder, M. (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science*, 318(5849), 420-426.
- Labra, M., Miele, M., Ledda, B., Grassi, F., Mazzei, M., & Sala, F. (2004). Morphological characterization, essential oil composition and DNA genotyping of *Ocimum basilicum* L. cultivars. *Plant Science*, 167(4), 725-731.

- Lanfermeijer, F. C., Dijkhuis, J., Sturre, M. J., de Haan, P., & Hille, J. (2003). Cloning and characterization of the durable tomato mosaic virus resistance gene Tm-22 from *Lycopersicon esculentum*. *Plant molecular biology*, 52(5), 1039-1051.
- Langella, R., Ercolano, M. R., Monti, L. M., Frusciante, L., & Barone, A. (2004). Molecular marker assisted transfer of resistance to TSWV in tomato elite lines. *Journal of horticultural science & biotechnology*, 79(5), 806-810.
- Lecoq, H., Pitrat, M. and Cle'ment, M. (1981) Identification et caractérisation d'un potyvirus provoquant la maladie du rabougrissement jaunedu melon. *Agronomie*, 1, 827–834.
- Lefebvre, V., & Palloix, A. (1996). Both epistatic and additive effects of QTLs are involved in polygenic induced resistance to disease: a case study, the interaction pepper—*Phytophthora capsici* Leonian. *Theoretical and Applied Genetics*, 93(4), 503-511.
- Lepoittevin, C., Frigerio, J. M., Garnier-Géré, P., Salin, F., Cervera, M. T., Vornam, B., ... & Plomion, C. (2010). In vitro vs in silico detected SNPs for the development of a genotyping array: what can we learn from a non-model species?. *PloS one*, 5(6), e11034.
- Li, S., Zhang, X., Sun, Z., Li, F., & Xiang, J. (2013). Transcriptome analysis on Chinese shrimp *Fenneropenaeus chinensis* during WSSV acute infection. *PloS one*, 8(3), e58627.
- Li, H., Ruan, J., & Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research*, 18(11), 1851-1858.
- Li, J., Zhang, W., Wu, H., Guo, T., Liu, X., Wan, X., ... & Wan, J. (2011). Fine mapping of stable QTLs related to eating quality in rice (*Oryza sativa* L.) by CSSLs harboring small target chromosomal segments. *Breeding science*, 61(4), 338.
- Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E., & Wittwer, C. (2004). Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical chemistry*, 50(7), 1156-1164.
- Lisa, V., Boccardo, G., D'Agostino, G., Dellavalle, G., & d'Aquilio, M. (1981). Characterization of a Potyvirus That Causes Zucchini Yellow Mosaic. *Phytopathology*, 71(7), 667-672.
- Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W., & Deetz, K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *Genome Research*, 4(6), 357-362.

- Lupas, A., Van Dyke, M., & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science*, 252(5009), 1162-1164.
- Madden, J. P., & Chaplowe, S. G. (1997). For ALL generations: making world agriculture more sustainable. World Sustainable Agriculture Association Publications.
- Malhi, R. S., Trask, J. S., Shattuck, M., Johnson, J., Chakraborty, D., Kanthaswamy, S., ... & Smith, D. G. (2011). Genotyping single nucleotide polymorphisms (SNPs) across species in Old World Monkeys. *American journal of primatology*, 73(10), 1031-1040.
- Mane, B. G., Tanwar, V. K., Girish, P. S., & Dixit, V. P. (2006). Identification of species origin of meat by RAPD-PCR technique. *Journal of Veterinary Public Health*, 4(2), 87-90.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1), 209-220.
- Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, 470(7333), 198-203.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bembien, L. A., ... & Volkmer, G. A. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376-380.
- Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research*, 18(9), 1509-1517.
- Massman, J., Cooper, B., Horsley, R., Neate, S., Dill-Macky, R., Chao, S., ... & Smith, K. P. (2011). Genome-wide association mapping of *Fusarium* head blight resistance in contemporary barley breeding germplasm. *Molecular breeding*, 27(4), 439-454.
- Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560-564.
- McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., ... & Moore, S. S. (2008). An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *Bmc Genetics*, 9(1), 37.
- Mesarich, C. H., Griffiths, S. A., van der Burgt, A., Ökmen, B., Beenen, H. G., Etalo, D. W., ... & de Wit, P. J. (2014). Transcriptome Sequencing Uncovers the *Avr5* Avirulence Gene of the Tomato Leaf Mold Pathogen *Cladosporium fulvum*. *Molecular Plant-Microbe Interactions*, 27(8), 846-857.

- Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., ... & Finn, R. D. (2014). The InterPro protein families database: the classification resource after 15 years. *Nucleic acids research*, gku1243.
- Mithöfer, A., & Boland, W. (2012). Plant defense against herbivores: chemical aspects. *Annual review of plant biology*, 63, 431-450.
- Morin, R. D., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T. J., ... & Marra, M. A. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques*, 45(1), 81.
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7), 621-628.
- Muchero, W., Diop, N. N., Bhat, P. R., Fenton, R. D., Wanamaker, S., Pottorff, M., ... & Close, T. J. (2009). A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. *Proceedings of the national academy of sciences*, 106(43), 18159-18164.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881), 1344-1349.
- Naudin, C. (1856). *Nouvelles recherches sur les caractères spécifiques et les variétés des plantes du genre Cucurbita*.
- Nijman, I. J., Kuipers, S., Verheul, M., Guryev, V., & Cuppen, E. (2008). A genome-wide SNP panel for mapping and association studies in the rat. *BMC genomics*, 9(1), 95.
- Notarnicola, B., Hayashi, K., Curran, M. A., & Huisinigh, D. (2012). Progress in working towards a more sustainable agri-food industry. *Journal of Cleaner Production*, 28, 1-8.
- Olivier, M. (2005). The Invader® assay for SNP genotyping. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 573(1), 103-110.
- Pachner, M., Paris, H. S., Winkler, J., & Lelley, T. (2014). Phenotypic and marker-assisted pyramiding of genes for resistance to zucchini yellow mosaic virus in oilseed pumpkin (*Cucurbita pepo*). *Plant Breeding*.
- Paris, H. S. (1986). A proposed subspecific classification for *Cucurbita pepo*. *Phytologia (USA)*.

- Paris, H. S. (2001). Characterization of the Cucurbita pepo collection at the Newe Ya'ar Research Center, Israel. *Plant Genetic Resources Newsletter*, 41-45.
- Paris, H. S. (2008). Summer squash. In *Vegetables I* (pp. 351-379). Springer New York.
- Paris, H. S., & Cohen, S. (2000). Oligogenic inheritance for resistance to zucchini yellow mosaic virus in Cucurbita pepo. *Annals of applied biology*, 136(3), 209-214.
- Paris, H.S., and Brown, R.N. (2005). The genes of pumpkin and squash. *HortScience* 40:1620-1630
- Park, J., & Seaton, R. A. F. (1996). Integrative research and sustainable agriculture. *Agricultural Systems*, 50(1), 81-100.
- Pasqualone, A., Montemurro, C., Summo, C., Sabetta, W., Caponio, F., & Blanco, A. (2007). Effectiveness of microsatellite DNA markers in checking the identity of protected designation of origin extra virgin olive oil. *Journal of Agricultural and Food Chemistry*, 55(10), 3857-3862.
- Pavy, N., Pelgas, B., Beauseigle, S., Blais, S., Gagnon, F., Gosselin, I., ... & Bousquet, J. (2008). Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC genomics*, 9(1), 21.
- Pearson, B. M., Gaskin, D. J., Segers, R. P., Wells, J. M., Nuijten, P. J., & van Vliet, A. H. (2007). The complete genome sequence of Campylobacter jejuni strain 81116 (NCTC11828). *Journal of bacteriology*, 189(22), 8402-8403.
- Pelham, J. (1966). Resistance in tomato to tobacco mosaic virus. *Euphytica*, 15(2), 258-267.
- Piron, F., Nicolai, M., Minoia, S., Piednoir, E., Moretti, A., Salgues, A., ... & Bendahmane, A. (2010). An induced mutation in tomato eIF4E leads to immunity to two potyviruses. *PloS one*, 5(6), e11313.
- Pitrat, M., & Lecoq, H. (1980). Inheritance of resistance to cucumber mosaic virus transmission by Aphis gossypii in Cucumis melo. *Phytopathology*, 70(10), 958-961.
- Quinn, N. L., Levenkova, N., Chow, W., Bouffard, P., Boroevich, K. A., Knight, J. R., ... & Davidson, W. S. (2008). Assessing the feasibility of GS FLX Pyrosequencing for sequencing the Atlantic salmon genome. *BMC genomics*, 9(1), 404.

- Rai, M. (2006). Refinement of the Citrus tristeza virus resistance gene (Ctv) positional map in *Poncirus trifoliata* and generation of transgenic grapefruit (*Citrus paradisi*) plant lines with candidate resistance genes in this region. *Plant molecular biology*, 61(3), 399-414.
- Robaglia, C., & Caranta, C. (2006). Translation initiation factors: a weak link in plant RNA virus infection. *Trends in plant science*, 11(1), 40-45.
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, 11(3), R25.
- Robinson, R. W., Provvidenti, R. & Shail, J. W. (1993). Tests for seedborne transmission of Zucchini yellow mosaic virus. *HortScience* 28, 694–696.
- Robinson, R. W., Munger, H. M., Whitaker, T. W., & Bohn, G. W. (1976). Genes of the Cucurbitaceae. *HortScience*, 11(6), 554-568.
- Ruffel, S., Gallois, J. L., Lesage, M. L., & Caranta, C. (2005). The recessive potyvirus resistance gene *pot-1* is the tomato orthologue of the pepper *pvr2-eIF4E* gene. *Molecular Genetics and Genomics*, 274(4), 346-353.
- Sanchez, J. J., Børsting, C., Hallenberg, C., Buchard, A., Hernandez, A., & Morling, N. (2003). Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. *Forensic science international*, 137(1), 74-84.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467.
- Santos, C. A., Blanck, D. V., & de Freitas, P. D. (2014). RNA-seq as a powerful tool for penaeid shrimp genetic progress. *Frontiers in genetics*, 5.
- Satkoski, J. A., Malhi, R. S., Kanthaswamy, S., Tito, R. Y., Malladi, V. S., & Smith, D. G. (2008). Pyrosequencing as a method for SNP identification in the rhesus macaque (*Macaca mulatta*). *BMC genomics*, 9(1), 256.
- Schrijnwerkers, C. C. F. M., Huijberts, N., & Bos, L. (1991). Zucchini yellow mosaic virus; two outbreaks in the Netherlands and seed transmissibility. *Netherlands Journal of Plant Pathology*, 97(3), 187-191.
- Schubert, T. S., & Ritchie, J. J. (1984). Zucchini Yellow Mosaic Virus. *Plant Pathology Circular*, (259).

- Schultz, J., Milpetz, F., Bork, P., & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences*, 95(11), 5857-5864.
- Seo, Y. S., Rojas, M. R., Lee, J. Y., Lee, S. W., Jeon, J. S., Ronald, P., ... & Gilbertson, R. L. (2006). A viral resistance gene from common bean functions across plant families and is up-regulated in a non-virus-specific manner. *Proceedings of the National Academy of Sciences*, 103(32), 11856-11861.
- Shivaprasad, P. V., Chen, H. M., Patel, K., Bond, D. M., Santos, B. A., & Baulcombe, D. C. (2012). A microRNA superfamily regulates nucleotide binding site–leucine-rich repeats and other mRNAs. *The Plant Cell Online*, 24(3), 859-874.
- Shukla, D. D., Frenkel, M. J. & Ward, C. W. (1991). Structure and function of the potyvirus genome with special reference to the coat protein coding region. *Can J Plant Pathol* 13, 178–191.
- Slinger, R., Bellfooy, D., Desjardins, M., & Chan, F. (2007). High-resolution melting assay for the detection of *gyrA* mutations causing quinolone resistance in *Salmonella enterica* serovars Typhi and Paratyphi. *Diagnostic microbiology and infectious disease*, 57(4), 455-458.
- Smith, B. D. (1997). The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago. *Science*, 276(5314), 932-934.
- Smith, G. D., Chadwick, B. E., Willmore-Payne, C., & Bentz, J. S. (2008). Detection of epidermal growth factor receptor gene mutations in cytology specimens from patients with non-small cell lung cancer utilising high-resolution melting amplicon analysis. *Journal of clinical pathology*, 61(4), 487-493.
- Sookruksawong, S., Sun, F., Liu, Z., & Tassanakajon, A. (2013). RNA-Seq analysis reveals genes associated with resistance to Taura syndrome virus (TSV) in the Pacific white shrimp *Litopenaeus vannamei*. *Developmental & Comparative Immunology*, 41(4), 523-533.
- Soosaar, J. L., Burch-Smith, T. M., & Dinesh-Kumar, S. P. (2005). Mechanisms of plant resistance to viruses. *Nature Reviews Microbiology*, 3(10), 789-798.
- Spaniolas, S., May, S. T., Bennett, M. J., & Tucker, G. A. (2006). Authentication of coffee by means of PCR-RFLP analysis and lab-on-a-chip capillary electrophoresis. *Journal of agricultural and food chemistry*, 54(20), 7466-7470.



- St. Clair, D. A. (2010). Quantitative disease resistance and quantitative resistance loci in breeding. *Annual review of phytopathology*, 48, 247-268.
- Stephens, A. J., Inman-Bamber, J., Giffard, P. M., & Huygens, F. (2008). High-resolution melting analysis of the spa repeat region of *Staphylococcus aureus*. *Clinical Chemistry*, 54(2), 432-436.
- Suh, J. P., Jeung, J. U., Noh, T. H., Cho, Y. C., Park, S. H., Park, H. S., ... & Jena, K. K. (2013). Development of breeding lines with three pyramided resistance genes that confer broad-spectrum bacterial blight resistance and their molecular analysis in rice. *Rice*, 6(1), 5.
- Sustainable Agriculture Network. Exploring Sustainability in Agriculture: Ways to Enhance Profits, Protect the Environment and Improve Quality of Life. Available: <http://www.sare.org/htdocs/pubs/explore/index.htm>.
- Takahashi, H., Miller, J., Nozaki, Y., Takeda, M., Shah, J., Hase, S., ... & Dinesh-Kumar, S. P. (2002). RCY1, an *Arabidopsis thaliana* RPP8/HRT family resistance gene, conferring resistance to cucumber mosaic virus requires salicylic acid, ethylene and a novel signal transduction mechanism. *The Plant Journal*, 32(5), 655-667.
- Tanner, N. K., & Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Molecular cell*, 8(2), 251-262.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., ... & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, 28(5), 511-515.
- University of California Sustainable Agriculture Research and Education Program. What is Sustainable Agriculture? Available: <http://www.sarep.ucdavis.edu/concept.htm>.
- Urcuqui-Inchima, S., Haenni, A. & Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Res* 74, 157–175.
- van der Biezen, E. A., & Jones, J. D. (1998). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Current Biology*, 8(7), R226-R228.
- Varshney, R. K., Terauchi, R., & McCouch, S. R. (2014). Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS biology*, 12(6), e1001883.

- 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.
- Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., ... & Ye, J. (2008). The diploid genome sequence of an Asian individual. *Nature*, 456(7218), 60-65.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63.
- Wang, Y., Gaba, V., Yang, J., Palukaitis, P., & Gal-On, A. (2002). Characterization of synergy between Cucumber mosaic virus and potyviruses in cucurbit hosts. *Phytopathology*, 92(1), 51-58.
- Weiland, J. J., & Yu, M. H. (2003). A cleaved amplified polymorphic sequence (CAPS) marker associated with root-knot nematode resistance in sugarbeet. *Crop science*, 43(5), 1814-1818.
- Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., ... & Rothberg, J. M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *nature*, 452(7189), 872-876.
- Whitaker, T. W. (1947). American Origin of the Cultivated Cucurbits. I. Evidence from the Herbals. II. Survey of Old and Recent Botanical Evidence. *Annals of the Missouri Botanical Garden*, 101-111.
- Whitaker, T. W., & Robinson, R. W. (1986). Squash breeding. *Breeding vegetable crops*, 209-242.
- Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G., & Pryor, R. J. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical chemistry*, 49(6), 853-860.
- Wright, W. T., Heggarty, S. V., Young, I. S., Nicholls, D. P., Whittall, R., Humphries, S. E., & Graham, C. A. (2008). Multiplex MassARRAY spectrometry (iPLEX) produces a fast and economical test for 56 familial hypercholesterolaemia-causing mutations. *Clinical genetics*, 74(5), 463-468.
- Xu, R., Zhang, S., Huang, J., & Zheng, C. (2013). Genome-wide comparative in silico analysis of the RNA helicase gene family in *Zea mays* and *Glycine max*: A comparison with *Arabidopsis* and *Oryza sativa*. *PLoS one*, 8(11), e78982.
- Xu, Y., & Crouch, J. H. (2008). Marker-assisted selection in plant breeding: from publications to practice. *Crop Science*, 48(2), 391-407.

- Xue, S., Liu, Y., Zhang, Y., Sun, Y., Geng, X., & Sun, J. (2013). Sequencing and de novo analysis of the hemocytes transcriptome in *Litopenaeus vannamei* response to white spot syndrome virus infection. *PloS one*, 8(10), e76718.
- Yan, J., Shah, T., Warburton, M. L., Buckler, E. S., McMullen, M. D., & Crouch, J. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One*, 4(12), e8451.
- Ye, G., & Smith, K. F. (2008). Marker-assisted gene pyramiding for inbred line development: Basic principles and practical guidelines. *International Journal of plant breeding*, 2(1), 1-10.
- Yoshii, M., Nishikiori, M., Tomita, K., Yoshioka, N., Kozuka, R., Naito, S., & Ishikawa, M. (2004). The *Arabidopsis* cucumovirus multiplication 1 and 2 loci encode translation initiation factors 4E and 4G. *Journal of Virology*, 78(12), 6102-6111.
- Yu, Y., Wei, J., Zhang, X., Liu, J., Liu, C., Li, F., & Xiang, J. (2014). SNP discovery in the transcriptome of White Pacific Shrimp *Litopenaeus vannamei* by next generation sequencing. *PloS one*, 9(1), e87218.
- Zdobnov, E. M., & Apweiler, R. (2001). InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, 17(9), 847-848.
- 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.
- Zeng, D., Chen, X., Xie, D., Zhao, Y., Yang, C., Li, Y., ... & Chen, X. (2013). Transcriptome analysis of Pacific white shrimp (*Litopenaeus vannamei*) hepatopancreas in response to Taura syndrome Virus (TSV) experimental infection. *PloS one*, 8(2), e57515.
- Zhao, M. F., Chen, J., Zheng, H.-Y., Adams, M. J. & Chen, J.-P. (2003). Molecular analysis of Zucchini yellow mosaic virus isolates from Hangzhou, China. *J Phytopathol* 151, 307–311.
- Zhou, L., Vandersteen, J., Wang, L., Fuller, T., Taylor, M., Palais, B., & Wittwer, C. T. (2004). High-resolution DNA melting curve analysis to establish HLA genotypic identity. *Tissue antigens*, 64(2), 156-164.
- Zhou, L., Wang, L., Palais, R., Pryor, R., & Wittwer, C. T. (2005). High-resolution DNA melting analysis for simultaneous mutation scanning and genotyping in solution. *Clinical Chemistry*, 51(10), 1770-1777.

Zhu, Y. L., Song, Q. J., Hyten, D. L., Van Tassell, C. P., Matukumalli, L. K., Grimm, D. R., ... & Cregan, P. B. (2003). Single-nucleotide polymorphisms in soybean. *Genetics*, 163(3), 1123-1134.

## 7. HOST COMPANY

This Ph.D. project was realized thanks to the collaboration between the Department of Agricultural of the University of Naples Federico II and “La Semiorto Sementi S.r.l.” company. For this reason, the research activities were performed in both laboratories.

## 8. STAGE

The stage was conducted from 17 November 2014 to 18 December 2014 at the “SEQUENTIA BIOTECH” laboratory, Campus UAB - Edifici CRAG - Bellaterra - Cerdanyola del Vallès – BARCELONA. During this stage, it was performed research about RNA-seq analysis and functional gene annotation with particular attention to the SNPs discovery and R-genes analysis.

## 9. PUBLICATONS

**Capuozzo C., Andolfo G., Formisano G., Ercolano M.R. Identification of a codominant CAPS marker associated to Zym-1 gene.** International Conference Plant Diseases and Resistance Mechanisms Vienna, Austria 20-22 February 2013

**Capuozzo C., Manzo D., Formisano G., Rrcolano M.R. CAPS and HRM markers for rapid detection of ZYMV resistance loci in zucchini.** 57° Congresso annuale SIGA – Foggia 16-19 September 2013

**Capuozzo C., Iovieno P., Barbella M., Frusciante L., Tomassoli L., Formisano G., Ercolano M.R. Towards a deep transcriptomic analysis of zucchini response to ZYMV attack for rapid detection of virus resistance loci in zucchini.** 58° Congresso annuale SIGA – Alghero 15-18 September 2014

**Andolfo G., Capuozzo C., Di Donato A., Iovieno P., Manzo D., Nieri D., Ercolano M.R. Integrated genomic approaches for investigating the plant R genes machinery.** XVI International Congress on Molecular Plant-Microbe Interactions – Rhodes, Greece 6-10 July 2014