Exploration of the Effector Triggered Immunity (*ETI*) in *S. lycopersicum* using genomic approaches

ANDOLFO GIUSEPPE

University of Naples "Federico II"

Department of Agricultural Sciences

Division of Plant Genetics and Biotechnology

Ph.D. Agrobiology and Agrochemistry

Academic year 2013-2014

Tutor: **Dott.ssa Maria R. Ercolano**[1]

Co-tutor: **Prof. Jonathan D.G. Jones**[2]

Coordinator: **Prof. Domenico Carputo**[1]

[1]Department of Agriculture Sciences, University of Naples 'Federico II', Via Universita` 100, 80055 Portici, Italy.

[2]The Sainsbury Laboratory, Norwich Research Park, NR4 7UH, Norwich, UK.

Exploration of the Effector Trigged Immunity (ETI) in *S. lycopersicum* using genomic approaches

Andolfo Giuseppe

University of Naples "Federico II", Department of Agricultural Sciences - Division of Plant genetics and biotechnology, Italy.

ABSTRACT

In order to investigate the genome-wide spatial arrangement of R loci, a complete catalogue of tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) nucleotide-binding site (NBS) NBS, receptor-like protein (RLP) and receptor-like kinase (RLK) gene repertories was generated. Candidate pathogen recognition genes were characterized with respect to structural diversity, phylogenetic relationships and chromosomal distribution. NBS genes frequently occur in clusters of related gene copies that also include RLP or RLK genes. This scenario is compatible with the existence of selective pressures optimizing coordinated transcription. A number of duplication events associated with lineage-specific evolution were discovered. These findings suggest that different evolutionary mechanisms shaped pathogen recognition gene cluster architecture to expand and to modulate the defense repertoire. Analysis of pathogen recognition gene clusters associated with documented resistance function allowed the identification of adaptive divergence events and the reconstruction of the evolution history of these loci. Most candidate pathogen recognition gene orthologues were distributed at less than perfectly matching positions, suggesting an ongoing lineage-specific rearrangement. Taken together, these findings have implications for improved understanding of the mechanisms of molecular adaptive selection at *Solanum* R loci. An updated *Solanaceae* RenSeq bait library to reannotate the full NB-LRR gene complement in tomato (*Solanum lycopersicum*) Heinz 1706 and to identify novel sequences that were not picked up by the semi-automated gene prediction software. Using 250-bp MiSeq reads after resistance gene enrichment sequencing (RenSeq) on genomic DNA of Heinz 1706, we identified 105 novel NB-LRR sequences. Reannotation included the splitting of gene models, combination of partial genes to a longer sequence and closing of assembly gaps. Within the draft *S. pimpinellifolium* LA1589 genome, RenSeq enabled the annotation of 355 NB-LRR genes. Phylogenetic analyses show a high conservation of all NB-LRR classes between Heinz 1706, LA1589 and the potato clone DM, suggesting that all sub-families were already present in the last common ancestor. A phylogenetic comparison to the *Arabidopsis thaliana* NB-LRR complement verifies the high conservation of the more ancient CC_{RPW8}-type NB-LRRs. Use of RenSeq on cDNA from uninfected and late blight-infected tomato leaves allows the avoidance of sequence analysis of non-expressed paralogues. cDNA RenSeq enables for the first time next-generation sequencing approaches targeted to this very low-expressed gene family without the need for normalization. Moreover in this thesis, we show that information on the tomato genome can be used predictively to link resistance function with specific sequences. An integrated genomic approach for identifying new resistance (*R*) gene candidates was developed. An *R* gene functional map was created by co-localization of candidate pathogen recognition genes and anchoring molecular markers associated with resistance phenotypes. In-depth characterization of the identified pathogen recognition genes was performed. Such methodology can help to better direct positional cloning, reducing the amount of effort required to identify a functional gene. The resulting candidate loci selected are available for exploiting their specific function. Finally, in order to identify a set of genes of interest in tomato plants infected with *F. oxysporum f. sp. lycopersici* (Fol) and *Tomato Mosaic Virus* (ToMV) a transcriptional analysis was performed. Differentially expressed tomato genes upon inoculation with Fol and ToMV were identified at 2 days post-inoculation. A large overlap was found in differentially expressed genes throughout the two incompatible interactions. However, GO enrichment analysis evidenced specific categories in both interactions. Response to ToMV seems more multifaceted, since more than 70 specific categories were enriched versus the 30 detected in Fol interaction. In particular, virus stimulated the production of an invertase enzyme that is able to redirect the flux of carbohydrates, whereas Fol induced homeostatic responses that prevent the attempt of fungus to kill cells. Genomic mapping of transcripts suggested that specific genomic regions are involved in pathogen resistance response. Coordinated *R* gene machinery could have an important role in prompt the response, since the 60% of pathogen receptor genes were differentially expressed during both interactions. Assessment of gene expression patterns could help for tracing a genomic model of *R* gene mediated resistance response.

LIST OF PUBLICATIONS

This thesis is based on the following papers which are referred to by their Roman numerals in the text:

I. **Andolfo G.**, Sanseverino W., Frusciante L., Ercolano M. R.. An integrated genomic approach for isolating R-genes in tomato. Minerva Biotecnologica, 2011 June; 23 (2 Suppl 1): 4-6.

II. Sanseverino W., Hermoso A., D'Alessandro R., Vlasova A., **Andolfo G.**, Frusciante L., Lowy E., Roma G., Ercolano M.R.. PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants. Nucleic Acids Research, 2012 October; 1-5.

III. **Andolfo G.**, Sanseverino W., Rombauts S., Van de Peer Y., Bradeen J.M., Carputo D., Frusciante L., Ercolano M.R.. Overview of tomato (Solanum lycopersicum) candidate pathogen recognition genes reveals important Solanum R locus dynamics. New Phytologist, 2012 September; 197(1):223-237.

IV. **Andolfo G.**, Sanseverino W., Aversano R., Frusciante L., Ercolano M.R.. Genome-wide identification and analysis of candidate genes for disease resistance in tomato. Molecular breeding, 2013 July; 9928:7.

V. **Andolfo G.**, Ferriello F., Tardella L., Ferrarini A., Sigillo L., Frusciante L., Ercolano M.R.. Genome-wide transcriptional responses to Fusarium oxysporum and Tomato mosaic virus in Tomato. PLOSONE accepted.

VI. **Andolfo G.**, Jupe F., Witek K., Etherington G.E., Ercolano M. R., Jones J.D.G.. Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. Manuscript.

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Plant innate immune system design

Multicellular organisms defend themselves against diseases with a so-called innate immune system. Vertebrates have acquired in addition the so-called adaptive immune system, which depends on antibodies. Innate immunity consists of both non-mobile cell types and mobile cell types that roam the circulatory systems in search for invaders (Kumar et al. 2011). The innate immune system is complemented by a second line of defense, known as specific or acquired immunity. This highly adaptive type of immunity consists of cells (B- and T-lymphocytes) carrying receptors that are produced by somatic recombination of gene segments and leads to an almost infinite antigen-specificity (Iwasaki et al. 2010). In this way, a molecular memory is created that primes the immune system for future attacks from the same organism. Innate immunity has evolved in plants, resulting in a wide diversity of defense mechanisms adapted to specific threats. The 30.000 described plant-diseases are just a little fraction of the possible plant pathogen interactions. The plant immune system has been modeled by the vast complexity in feeding behaviors of microbial pathogens through co-evolution over millions of years. Mainly, plant pathogens can be divided into three classes accordingly to lifestyle. Biotrophs rely on live host cells either completely or partially in their life cycle. Hemi-biotrophic pathogens that only depend on living cells for the initial stage of the infection and then switch to a necrotrophic lifestyle. Necrotrophic microorganisms kill their hosts during the infection and feed from the dead matter.

Plants do not have an elaborate immune system since they lack both a circulatory system and specialized immune cells. In plants, each individual living cell of the plant body must be able to respond autonomously to a wide range of potential pathogens. Nevertheless, plants are capable of perceiving a wide range of antigens with high specificity, avoiding reactivity to self, and to ward off potentially dangerous microbes. Plants are also capable of immunizing tissue distal to an infection, creating a “memory” that can be maintained over generations (Spoel et al. 2012). It thus appears that the immune systems of vertebrates and plants must at least partly use alternative mechanisms to deal with infectious disease.

1.2. First defense line: PAMP-Triggered Immunity (PTI)

The first line of inducible defense in plants is mediated through surface-localized pattern recognition receptors (PRRs). PRRs perceive pathogens directly via “non-self” molecules or indirectly through the detection of danger associated “self” signals. The direct sensing of

microbes is activated by the recognition of general elicitors, Pathogen-Associated Molecular Patterns (PAMPs) which are characteristic for entire groups of microbes, such as bacterial peptidoglycan or fungal chitin. Attempted infections may lead to the indirect recognition of the pathogen through host derived danger signals, Danger-Associated Molecular Patterns (DAMPs), that arise from wounding or injury. These include plant peptides released from the cell wall. PAMP-binding and DAMP-binding activate the PRRs and induce multiple defense responses in the plant cells resulting in PAMP-triggered immunity (PTI) (Monaghan et al. 2012). PAMPs are highly conserved molecules that are shared among several classes of microbes. They include lipopolysaccharides and flagellin from gram-negative bacteria, peptidoglycans from gram-positive bacteria, chitin, ergosterol and β-glucans from oomycetes and fungi. As many of the PAMPs represent vital components for microbial life, they are not per se important for pathogenicity. PAMPs serve as molecular cues for surface localized pattern-recognition receptors (PRRs) that relay the signal of an attack to the plant cell interior. Principally, the PRRs identified so far can be divided into receptor-like kinases (RLKs) or receptor-like proteins (RLP).

1.3. Second defense line: Effector-Triggered Immunity (ETI)

To evade PTI, adapted pathogens secrete effector molecules into the plant cells that interfere with PRR signaling and suppress pattern-triggered responses (Zipfel et al. 2008). Effectors may also enforce metabolic shifts on the host plant which are beneficial for the attacker (Chen et al. 2010; Ward et al. 2010). In turn, plants express intracellular resistance (R) proteins that directly interact with the effectors or sense their presence through perturbation of endogenous effector targets (Liu et al. 2009; Jia et al. 2000). The resulting Effector-triggered immunity (ETI) is a much faster and stronger immune reaction then those triggered by PAMPs (Spoel et al. 2008). ETI and PTI responses are often overlapping although distinct differences exist. For example, the hypersensitive response (HR), a type of localized programmed cell death, most often follows R-mediated resistance; while callose deposition and cell wall fortification are commonly associated with PRR-triggered resistance. As an evolutionary twist to the system, pathogens have developed effectors that render the R proteins useless. These effectors may in turn be sensed by another set of R proteins, reflecting an evolutionary arms race between the plant and the microbe (the "zigzag model", Jones and Dangl 2006).

1.3.1. Structure of principal pathogen receptor family

Pathogen recognition genes can be categorized according to the presence and organization of protein domains, such as Toll/Interleukin-1 receptor (TIR), coiled coil (CC), the nucleotide-binding site (NBS), leucine-rich repeats (LRRs), receptor domains and kinase domains. Deduced proteins from cloned path rec genes can be classified into four typical families: the TIR-NBS-LRRs (TNLs; e.g. N; Erickson et al., 1999), the CC-NBS-LRRs (CNLs; e.g. I2; Simons et al., 1998), the receptor-like kinases (RLKs; e.g. SlFls2; Robatzek et al., 2007) and the receptor-like proteins (RLPs; e.g. Cf4; Thomas et al., 1997).

The majority of *R* genes encode immune receptors containing a nucleotide-binding and a leucine-rich repeat domain (NB-LRR). Plant NB-LRR proteins (also called NLR, NBS-LRR or NB-ARC-LRR proteins) can be categorized into TIR and non-TIR classes based on the identity of the sequences that precede the NBS domain (Meyers et al. 2005). The TIR class of plant NB-LRR proteins (TNLs) contains a Toll interleukin 1 receptor homology (TIR) protein-protein interaction domain at the amino terminus. The non-TIR class (CNLs) is less well defined, but some members of this class contain helical coiled-coil–like sequences in their amino-terminal domain (Pan et al 2013). The LRR domain seems to have a dual function, namely, as a sensor of pathogen stimuli and as an intramolecular signal transducer. In the inactive state, the NB domain interacts with the LRR domain and forms a closed nucleotide binding pocket. During activation, the NB domain is released allowing exchange of ADP for ATP, alternatively ATP hydrolysis (Fenyk et al. 2012), and enables the protein to assume an open conformation (Williams et al. 2011).

RLKs consist of a ligand-binding extracellular region, a single membrane spanning domain, and a cytoplasmic kinase domain. RLPs differ from RLKs in that they lack the kinase domain and only have a short cytoplasmic tail. The extracellular region of RLKs/RLPs shows great diversity and more than 20 structurally distinct domains have been identified (Cock et al. 2012). This large versatility in amino acid sequence has been ascribed the need for plants to quickly adapt to the ever-changing structures of microbial elicitors (Shiu et al. 2001).

1.3.2. Function of NB-LRR genes

NB-LRR proteins are involved in the recognition of specialized pathogen effectors (also called avirulence (Avr) proteins) that are thought to provide virulence function in the absence of the cognate R gene (Chisholm et al. 2006). In the absence of a corresponding R protein, the pathogen is able to colonize its host. In contrast, the activation of NB-LRR receptors results in a fast and strong response: effector trigged immunity (ETI), that has been also called gene-for-gene resistance (Flor et al. 1971). *R* proteins perceive effectors by direct physical interaction or indirectly through effector modification of host targets (Heidrich et al. 2012). Some Avr proteins are virulence factors that facilitate pathogen growth or interfere with basal plant defenses (PTI) (Nimchuk et al.2000).

Recently, Takken and Tameling have proposed a model for *R*-protein activation. In the absence of a pathogen, NB-LRR R proteins reside in an autoinhibited, ADP-bound "OFF" state. Effector-perception by the LRR domain open the conformation of the R protein that is prone to nucleotide exchange. ADP/ATP exchange triggers a second conformational change, resulting in the "ON" state. In the activated state, the NB subdomain becomes exposed to initiate defense signalling (Takken and Tameling 2009).

Numerous studies have identified components involved also in PRR signaling as effector targets. AvrPphB, a cysteine protease from *P. syringae*, and AvrAC, an uridylyl transferase from *X. campestris pv. campestris*, both target the BIK1 kinase of the FLS2/EFR/CERK1 signalosome (Feng et al. 2012; Zhang et al. 2010). AvrPto ligase that promotes degradation of the FLS2 receptor by catalyzing polyubiquitination of the kinase domain in Arabidopsis (Gohre et al. 2008). Similarly, the MAP kinase pathways downstream of PRR activation are targeted by multiple effectors (Zhang et al. 2012; Cui et al. 2010). Also later events in the plant defense reaction have been identified as effector targets; the HopZ effector was found to enhance pathogenicity by degrading an enzyme involved in isoflavonoid biosynthesis in soybean (Zhou et al. 2011). However, not all effector molecules associates interfere with protein function. TAL effectors from *Xanthomonas* bacterial pathogens contain domains that are characteristic for eukaryotic transcription activators. TALEs bind host DNA with high sequence specificity and induce expression of target genes, also termed disease susceptibility genes (Kay et al. 2007). Target genes for transcriptional reprogramming by TAL effectors include transcription factors and SWEET sugar transporters (Kay et al. 2007; Antony et al. 2010). SWEET proteins mediate glucose transport and up-regulation of the encoding genes may help the pathogen to fulfill its nutritional needs (Chen et al. 2010).

1.3.3. R gene models of action

Direct gene-for-gene relationship (Flor, 1971), how observed between *XA21* in rice or *FLS2* in Arabidopsis with microbial peptides (Song et al. 1995; Gomez-Gomez et al. 2000), has been described in few cases. For most of NBS-LRR genes a direct interaction has not been shown. It appears that the *R* proteins act as guards of proteins known as the "guardee" which is the target of the Avr protein. Indeed, the Guard Model clarifies how numerous functionally unrelated effectors can be recognized by a single NB-LRR if they share a common target and how a relatively low number of NB-LRR proteins, 150 in Arabidopsis and about 300 in tomato (Meyers et al. 2003; Andolfo et al 2013), can confer resistance to endless number of pathogen effectors. Recently, a new model has been proposed and it suggested that some effector targets act as plant baits for effector detections by *R* proteins (van der Hoorn et al. 2008). The Decoy Model states that in a plant population polymorphic for *R* genes, there are two opposing selection forces that act on guarded effector targets. This evolutionarily unstable could be relaxed from evolution of a host protein, termed "decoy". This latter specialized in perception of effector by *R* protein but uninvolved in the development of disease or resistance.

A molecularly well characterized guarded effector target is the protein *Pto* of *Solanum lycopersicum*. *Pto* is a kinase inhibitor that block the PAMP signaling that is under the surveillance of two receptor-like kinases, *FLS2* and *EFR* (Xing et al., 2007; Xiang et al., 2008). *Pto* confers resistance to *P. syringae*, interaction that requires the presence of *R* gene Prf. AvrPto contributes to virulence on tomato and Arabidopsis but not on Arabidopsis lacking FLS2, indicating that FLS2 is an operative virulence target of AvrPto (Xiang et al., 2008).The tomato *Pto*/*Prf*/AvrPto interaction showed that the avirulence protein, AvrPto, interacted directly with *Pto*. This makes *Pto* the guardee protein, which is protected by the NBS-LRR protein *Prf*. Emerging evidence suggests that other effector targets, as the phosphorylation of RIN4 by the *P. syringae* effectors AvrRpm1 and AvrB, can act as pathogen bait and prevent virulence (Mackey et al. 2002).

1.4. Pathogen recognition genes in an evolutionary perspective

Selection pressure exercised form pathogen recognition genes on effector genes is very strong. This force produces an equally highly polymorphic in effector sequences. It is well known from several years, that *R* gene are organized, within the plant genomes, in cluster of

vary size (Hulbert et al. 2001; Meyers et al. 2005; Andolfo et al. 2013) and that the pathogen recognition genes clusters have been originated by gene duplications. The clustered arrangement of pathogen recognition gene family may be a critical attribute allowing the generation of novel resistance specificities via recombination or gene conversion (Hulbert et al., 2001). These diversification events have facilitated the *R* gene evolution and a faster arms race of plants against the pathogens. Interestingly, recent studies described *R* gene-rich regions with higher rates of mutations and recombination events than the genome average (Bakker et al. 2006; Andolfo et al. 2013). In addition, analyses of individual clusters provided evidence of diversifying selection in the majority of plant R genes studied, suggesting that variation may be concentrated within predicted binding surfaces (Cooley et al., 2000; Luck et al., 2000; Mondragon-Palomino et al., 2002). Allowing greater specificity and effectiveness among the R proteins and effector-targets. Lately, it has been observed that the pathogen pressure can stimulate the development of new *R* genes epigenetic destabilization of the genome (Alvarez et al. 2010). The genomic flexibility, detected in the cluster, outfit plant with new *R* gene that can serve as guards for unencountered epitopes of microbial pathogens. Additional NB-LRR genes identified in on going plant genomics projects are contributing to our understanding of the mechanisms that generate sequence diversity in these genes.

1.5. Hormones activation

In all phases of plant life (plant growth, development, and reproduction), plant hormones, also known as phytohormones, play essential roles. Phytohotmones are signal molecules that regulate cellular processes locally but also systemically. Additionally, they emerged as substances with key functions in the regulation of immune responses to microbial pathogens. The complex network that interconnected the pathways of signal provide the plant of an enormous regulatory potential to rapidly adapt to biotic environment.

The phytohormone triad: salicylic acid (SA), jasmonates (JAs) and ethylene (ET), has a well documented role in responses to biotic stresses (Loake and Grant 2007). In addition, recent studies have identified abscisic acid (ABA), auxin, brassinosteriods (BR), cytokinin (CK), gibberellic acid (GA) and peptide hormones as important regulators of immune responses, they emerged as important players on the battle field (Bari and Jones 2009). Current evidences in plant suggest that salicylic acid is involved in resistance against (hemi-) biotrophic pathogens; by contrast resistance to nectrotrophic pathogens is controlled by jasmonic acid and ethylene signaling pathways. Mutations leading to either reduced

hormones-production or impaired hormones-perception exhibit enhance susceptibility to virulent pathogens (Tsuda et al. 2009). In many plants SA and JA/ET defence pathways interact antagonistically (Glazebrook 2005; Shah 2003). Evidence for both synergistic and antagonistic interactions between SA and JA are reported (Schenk et al. 2000; Mur et al. 2006).

1.6. Systemic acquired resistance

The systemic acquired resistance (SAR) is a "whole-plant" resistance response that occurs following an earlier localized exposure to a pathogen and helps plants to withstand secondary infections. Plant receptors recognize the conserved microbial signatures triggered an hypersensitive response that usually induces the SAR but it is not obligatory required for the production of SAR signals (Alvarez ey al. 1998; Mishina and Zeier 2007). SAR, also named “broad spectrum”, is associated with the induction of a wide range of genes (so called PR proteins or "pathogenesis-related" proteins), and the activation of SAR requires the accumulation of endogenous salicylic acid (SA) (Tsuda et al. 2008). Mutant and transgenic plants that have a compromised SA signaling are incapable of developing SAR (Durrant and Dong 2004), which indicates that SA is a necessary intermediate in the SAR signaling pathway. The regulatory protein NPR1 emerged as an important transducer of the SA signal; upon activation by SA, NPR1 acts as a transcriptional co-activator of PR gene expression (Dong 2008). Recent studies point to a role for methyl-SA, JAs, a plastid glycerolipid-based factor, and a lipid-transfer protein (Vlot et al. 2008). Whereas SAR is predominantly effective against biotrophic pathogens that are sensitive to SA-dependent defenses (Ton et al 2002; Van Oosten et al. 2008).

1.7. Genomic approach for *R* gene investigation

The genomic approach for exploring resistance genes dataset could be useful for shed light in molecular and evolutionary mechanisms of this gene family. The use of such technologies will facilitate to design of diagnostic tests, to conduct comparative and functional analysis and to perform breeding by in *silico* design. In modern science, DNA sequencing technologies are being brought up to date at an intense pace. Genomic approaches offer new insights into the plant immune system. It is possible to start from an unprocessed data to select a specific set of candidate genes putatively involved in biotic stress response. The

RNA-Seq experiments make it possible to precisely determine the expression levels of specific genes, differential splicing, and allele-specific expression of transcripts. The next generation sequencing platforms produce a large amount of available data with an unprecedented level of sensitivity. Thus, the NGS platforms provide clear advantages as well as new challenges and issues. Tomato represents one of best-explored model plants for studying defense response systems. Global information on tomato defense responses can create a body of knowledge regarding the frequency of relevant sequences, their evolution and possible functions. Hence, it would be of high importance to develop tools to pool information obtained through different systems, to connect and to compare information in molecular biology and biochemistry; this would help to delineate a systems biology approach in order to understand the plant-defense mechanism, thereby allowing the design of new breeding methods. A multiple omic approaches that use different platforms integrating these results, is therefore a modern and high effective strategy for explaining integral molecular systems for plant improvement.

1.8. Scientific aims

The work presented in this thesis is my contribution to a deepened understanding of the plant immune system. The first section of this thesis introduces the basic structural design of the plant innate immune system. The systems by which plants recognize pathogens directly or indirectly through so-called pathogen recognition genes are described in detail. The second part is dedicated to describing the outcome of four overall aims : I) the functional and structural genomic investigation of pathogen recognition gene family in *Solanum lycopersicum* and *Solanum tuberosum*, II) the improvement of a system to identify novel NB-LRR genes that were not picked up by the semi-automated gene prediction in *Solanum lycopersicum* and *Solanum pimpinellifolium*, III) the developing of an integrated genetic and genomic system to reduce the complexity for functional *R* gene identification, and IV) the exploration of tomato transcriptional machinery in incompatible interactions with a foliar (*Tomato mosaic virus*) and a vascular (*Fusarium oxysporum f. sp. lycopersici*) pathogens.

Specifically, the following questions are addressed:

- What is the composition and the distribution of pathogen recognition gene family in tomato and potato genomes? **(Papers I, II and III)**
- How evolution mechanisms modulated the defence gene repertories? **(Papers II and III)**
- Is the tomato pathogen recognition genes annotation a model for functional genomic comparison between *Solanum spp.*? **(Papers II and III)**
- How NGS-technology can help to improve the detection of resistance gene family complexity? **(Paper VI)**
- How new generated sequencing data speed up selection program for disease resistance? **(Papers IV and VI)**
- How is the transcriptional machinery in tomato reprogrammed after the inoculation with key pathogens? **(Paper V)**
- What are the major metabolic perturbation activated in tomato during the interaction with Fol and ToMV? **(Paper V)**
- What is the correlation between genomic organization and transcriptional response to the two biotic stresses? **(Papers V)**

2. Overview of tomato pathogen recognition genes and potato comparison

2.1. Introduction

All pathogen recognition genome-wide analyses to date have focused on NBS genes, showing that a large fraction of these genes are arranged in clusters. These analyses also suggest that gene duplications have played an important role in the establishment and expansion of NBS-LRR gene clusters, resulting in broad allelic diversity at NBS-LRR loci (Meyers et al., 2005; Liu et al., 2007). Deciphering the evolutionary history of a gene cluster is important not only to understand the functional specificity of each gene inside the cluster, but also to provide new insights into gene duplication mechanisms and to answer questions regarding other evolutionary events. Defense response is also thought to be a primary biological function of transmembrane receptor (RLK and RLP) proteins. A number of mixed clusters of genes were identified in Arabidopsis species (Guo et al., 2011). Several studies have shown that genes are not distributed randomly across eukaryotic chromosomes, but rather occur in functional neighbourhoods (Al-Shahrour et al., 2010). Therefore, discovering the functional significance of clustered gene arrangements may provide insights into gene evolution. Comparative study of pathogen recognition genes is poised to yield fundamental understanding of processes of pathogen detection and plant evolution. Phylogenetic and orthology analyses of pathogen recognition genes have been previously employed to draw inferences about evolutionary processes such as speciation, gene duplication and gene conversion (Chen et al., 2010). Several authors provided evidence of rapid evolution and diversification of pathogen recognition genes resulting from either positive or stabilizing selection (Yang et al., 2008; Chen et al., 2010; Li et al., 2010; Zhang et al., 2010). Indeed, recent genomic investigations indicated that polyploidy is ubiquitous among angiosperms and have provided evidence of several ancient genome-doubling events (Soltis et al., 2009; Jiao et al., 2011). Such events may have played a role in the expansion of pathogen recognition gene numbers in some species (Yang et al., 2008), but the fate of most duplicated pathogen recognition genes is poorly understood. This is due to a result in part of the difficulty in identifying orthologous pathogen recognition loci across species because of plant lineage-specific pathogen recognition gene amplifications and deletions. Importantly, given the accumulated knowledge of disease resistance in these two species, whole-genome comparative analysis of the pathogen recognition gene component should facilitate rapid advancement of disease resistance breeding. In this study, we investigated the physical arrangement of candidate pathogen recognition genes in the tomato genome and explored how gene duplication has influenced candidate pathogen recognition gene cluster composition and evolution. We also deciphered the evolutionary history of pathogen

recognition gene loci associated with documented resistance function. Finally, we conducted genome-wide comparisons of the candidate pathogen recognition components of tomato and potato.

2.2.Materials and methods

2.2.1. Pathogen recognition protein prediction analyses

A script developed in-house was employed to identify tomato (Solanum lycopersicum L.) and potato (Solanum tuberosum L.) pathogen recognition proteins. First, deduced protein sequences from 76 previously cloned pathogen recognition genes were assigned to structural groups (e.g. CNL, TNL, RLK, RLP, and kinase-like) through phylogenetic analysis. Protein sequences belonging to a single group were aligned using MUSCLE 3.6 (Edgar, 2004) with manual editing. The resulting alignments for each group were used as a base for the creation of hidden Markov models (HMMs) using the HMMER v2 package (http://hmmer.janelia.org/). The HMM profiles were next used to screen the tomato and potato proteomes (34.727 and 35.004 proteins, respectively; iTAG (International Tomato Annotation Group) annotation) to identify putative pathogen recognition proteins. The result of this screen is a numeric matrix that represents the similarity score of every single protein with each HMM profile. The set of predicted pathogen recognition proteins identified via HMM profiling was further analysed using INTERPROSCAN software version 4.8 (Quevillon et al., 2005) to verify the presence of conserved domains and motifs characteristic of pathogen recognition 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, HMMPIR, 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 pathogen recognition 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) was used to detect CC domains in CNL proteins using a threshold = 0.9 followed by visual inspection.

2.2.2. Sequence alignments

MUSCLE version 3.6 (Edgar, 2004) was used to align amino acid and nucleotide sequences. MAFFT version v 6.814b (Katoh et al., 2002) was used to align the C3-F region of RLP proteins and the kinase domain of RLK proteins. MAUVE package 2.2.0 (http:// asap.ahabs.wisc.edu/mauve) was used to align homologous regions among two or more genome sequences. To determine a reasonable value, for the Min Locally Collinear Blocks (LCBs) we performed an initial alignment at the default value and then used the LCB weight slider in the MAUVE GUI to fix parameters empirically, eliminating all spurious rearrangements. The sequences were then realigned using the manually determined weight value.

2.2.3. Physical genome locations

A pipeline was developed in-house to locate predicted pathogen recognition genes on tomato and potato genome sequences. The latest version of the tomato and potato gff3 annotation files was parsed to extract the genomic coordinates of the predicted pathogen recognition genes. Moreover, sequences of 75 chromosome markers associated with genetically defined pathogen recognition loci (Foolad, 2007) were mined from the Sol Genomics Network (solgenomics.net) database. The physical genome position of each marker was determined by mapping marker nucleotide sequences to the tomato genome sequence using BLASTn (E-value 1e-1 and 100% match length). A database detailing pathogen recognition gene features, tomato genome locations and pathogen recognition gene marker information was built to draw physical maps.

2.2.4. Duplication and phylogenetic analysis

To identify duplicated pathogen recognition gene pairs, we defined a gene duplication according to criteria reported in Gu et al. (2002). Evolutionary analyses were conducted using MEGA5 (Tamura et al., 2011). The phylogenetic relationships of predicted tomato pathogen recognition proteins and of pathogen recognition proteins encoded by previously cloned pathogen recognition genes were inferred separately for each structural class (e.g. CNL, TNL, RLK, RLP, and kinase-like) using the maximum likelihood method based on the Whelan and Goldman + Freq. model (Whelan &

Goldman, 2001). For nucleotide sequences, the general time reversible model was used. The bootstrap consensus tree inferred from 100 replicates was taken to represent the evolutionary history of the sequences analysed (Felsenstein, 1985). Phylogenetic trees of RLP proteins were estimated using the conserved C3-F region as input. The trees were drawn to scale, with branch lengths measured in the number of substitutions per site.

2.2.5. Orthology analysis

Orthologous relationships among tomato and potato candidate pathogen recognition genes were established on the basis of reciprocal best BLAST hits and a maximal E-value of 1e-5 with a minimal match length of 50%. Orthologous gene pairs were subjected to phylogenetic analysis as described in the previous paragraph. Putative orthologous genes belonging to the same protein class were filtered for sequence identity. Proteins with a global alignment identity of at least 90% in the NBS and TIR domains or 80% for other protein classes were selected. Putative orthologues were also filtered for physical co-localization (synteny). Based on these criteria, a set of tomato and potato orthologous proteins was generated.

2.2.6. Evolution rates at codon sites

Selective pressures acting on the MI1.2 and I2 super-clusters (chromosomes 6 and 11, respectively) were investigated by determining the nonsynonymous to synonymous nucleotide substitution ratios (dN/dS) (KaKs_Calculator; http://code.google.com/p/kaks-calculator/downloads/list). Tests were conducted to estimate the evolution of each codon: positive ($x > 1$); neutral ($x = 1$); and negative ($x < 1$). These values were obtained using the joint maximum likelihood reconstructions of ancestral states with a GY method (Goldman & Yang, 1994) of codon substitution and general time reversible model (Nei & Kumar, 2000) of nucleotide substitution. P values of < 0.05 were considered significant. Then, the average nucleotide diversity (p) of orthologous pairs within each gene cluster was calculated using a bootstrap procedure (100 replicates). Analyses were conducted using the maximum composite likelihood model (Tamura et al., 2004). Differences in the composition bias among sequences were considered in

evolutionary comparisons (Tamura & Kumar, 2002). Codon positions included were first. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011).

2.3. Results

2.3.1. Chromosomal locations of candidate pathogen recognition genes in tomato

The genome-wide distribution of CNL, RLP and RLK genes (Supporting Information Table S1), based on chromosome size, calculated at $P < 0.001$ was nonrandom (CNL $\chi 2 = 60$; RLP $\chi 2 = 70$; RLK $\chi 2 = 92$;). On a per chromosome basis, the distribution of the CNL class ranged from 3 (chromosomes 3 and 7) to 26 (chromosome 11), while that of the TNL class ranged from 1 (chromosomes 2 and 7) to 5 (chromosome 1). Some chromosomes showed a predominance of specific classes. Chromosomes 1 and 12 were rich in genes of the RLP class, RLK genes were concentrated on chromosomes 2, 3 and 4, and CNL genes were mainly grouped on chromosomes 4 and 5. The TNL class was not present on chromosomes 3, 6 and 10, while on chromosome 12, three TIR domains and a TIRNBS gene were present (Figure 1). Genes encoding proteins of the RLP and CNL classes tended to cluster more than those encoding proteins belonging to other classes. Based on Holub's (2001) definition of a gene cluster (a region that contains four or more genes within 200 kb or less), we found that 227 candidate pathogen recognition genes (c. 30% of the total) resided in clusters. Forty-two pathogen recognition gene clusters with an average number of 5.6 genes per cluster were found (Table 1). Ten clusters of four to seven members comprised genes belonging to the same class (six RLK cluster; four RLP clusters). Twenty-six of the remaining 32 mixed clusters contained at least one NBS gene (data not shown). Sixteen clusters included one or more CNL genes and four included one or more TNL genes. Mixed clusters usually occupied a larger genome region and contained a larger number of gene copies. In particular, 12 clusters of four genes, seven with five, two with six, four with seven, three with eight, one with nine, one with ten, and one with 13 genes were discovered (Figure 2). The two largest clusters were both located on chromosome 4 (Solyc04g011890.1.1–Solyc04g012190.1.1 in one cluster and Solyc04g009070.1.1–Solyc04g009290.1.1 in the other). Besides gene clusters, 79 tandem arrays and 30 triplets were identified. Therefore, 65% of all tomato candidate pathogen recognition genes resided either in a gene cluster or in an array of two or three genes. In our survey were identified 18 genome regions, within 500 kb, showing aggregates of clusters or arrays. Of these, 13 included at least one cluster and were classified as super-clusters. Chromosomes 4, 5, 7 and 12 contained 15% of the entire data set organized in clusters or arrays, spanning regions ranging from 2.5 to 5.5 Mb (Figure 2).

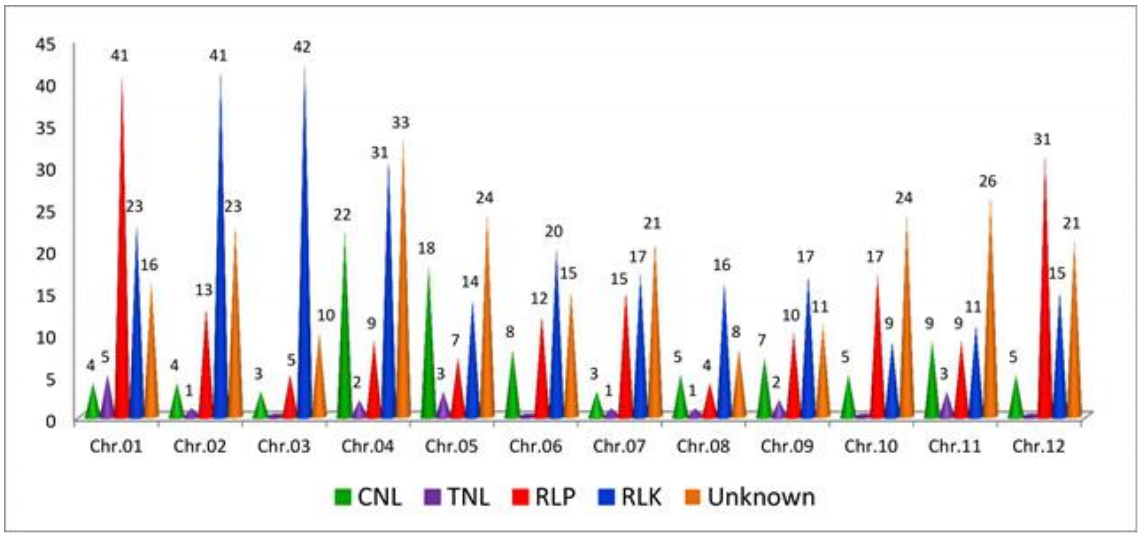

Figure 1. Distribution of R-gene classes across tomato chromosomes. R-genes within the tomato genome are differentiated into structural classes by color: RLP Receptor-Like Protein; RLK Receptor-Like Kinase; CNL CC-NB-LRR protein; TNL TIR-NB-LRR protein; Unknown genes that encode novel domain associations or single domains; Pto/like genes encoding the typical serine theoinine domain characteristic of the Pto protein.

Cluster composition	Cluster N.	Average N. genes	Average size (Kb)
CNL-RLK-Unknown	4	6.5	129.9
CNL-Unknow	11	5	98.1
CNL-TNL- Unknown	1	5	102.6
TNL-Unknow	1	4	30.9
TNL -RLK-RLP	1	4	134.2
TNL-RLK-RLP- Unknown	1	9	167
RLK	6	4.8	94
RLK- Unknown	2	4.5	112.7
RLK-RLP	2	8	173.9
RLP	4	5.5	52.4
RLP- Unknown	5	4.6	63.6
Unknown	4	5	82.7
Total clusters	42		
Mean cluster size	5.6		
Median cluster size	5		

Table 1. R-gene clusters found in the tomato genome. CNL CC-NB-LRR protein; TNL TIR-NB-LRR protein; Unknow: genes that present new domain associations or single domains; RLP Receptor-Like Protein; RLK Receptor-Like Kinase.

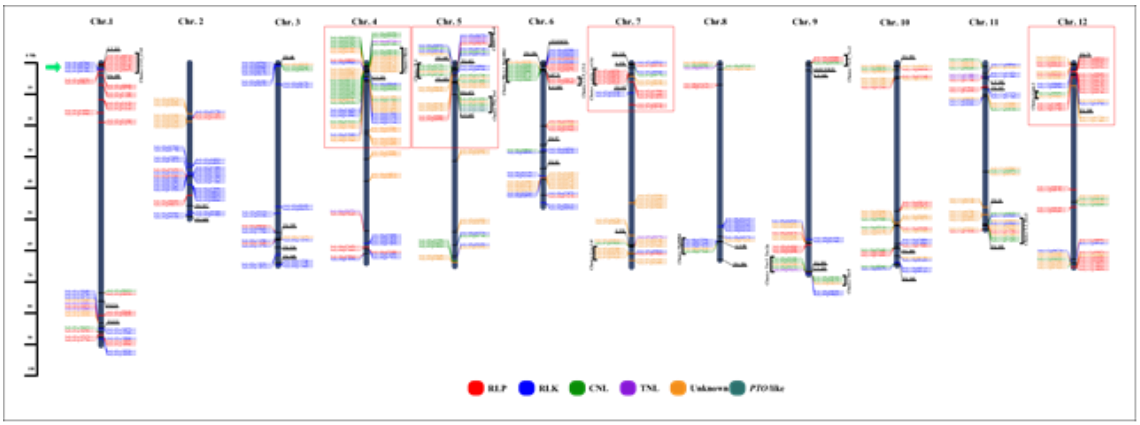

Figure 2 Physical genome locations of tomato (Solanum lycopersicum) candidate pathogen recognition gene clusters. Tomato chromosomes are represented as blue bars, the approximate location of each candidate pathogen recognition gene cluster and tandem array is designated by associated marker locations; marker codes are underscored with horizontal lines to the left or right of each chromosome. Candidate pathogen recognition genes residing in either a cluster or a tandem array are indicated with a gene ID. The colour used for each ID indicates the structural class to which the encoded protein has been assigned. RLP, receptor-like protein; RLK, receptor-like kinase; CNL, CC-NB-LRR protein; TNL, TIR-NB-LRRCC-NB-LRR TIR-NB-LRR protein; Unknown, genes that encode novel domain associations or single domains; PTO-like, genes encoding the typical serine theoinine domain characteristic of the Pto protein. Tandem arrays and gene clusters containing a previously cloned pathogen recognition gene are indicated with the cloned gene name and denoted with vertical lines. The green arrow on chromosome 1 indicates the location of a duplication involving two Gnk2-RLK genes. The red boxes indicate large aggregates of clusters or arrays.

2.3.2. Reconstructing candidate pathogen recognition gene duplication events

In the tomato genome, a total of 19 tandem duplicated candidate pathogen recognition gene pairs and 25 blocks of duplication were detected. Chromosomes 1, 4 and 12 showed a high number of duplication events. For example, a tandem array with two RLK-Gnk2 genes on chromosome 1 was duplicated on the same chromosome, generating a cluster of four genes. Interestingly, a cluster composed of 13 genes (eight CNL, one NL, one CN and three N genes), on chromosome 4, appeared to have experienced several duplication events. From the analysis conducted it is clear that the block of duplication (Figure 3d) determined the expansion of the gene cluster. The phylogenetic analysis, tree topology and genomic alignment (see Figure 3a–c) show the homology level between and within each of two gene subclusters. On chromosome 12, we found a block of duplication involving four candidate pathogen recognition genes located in a cluster composed of seven receptor-like proteins. Another interesting duplication involved Solyc10g054970.1.1 (a CCNBS gene) and Solyc10g054940.1.1 (an NBS gene) on chromosome 10 (overall pairwise identity 95.4%). Out of 44 gene duplication events, nine were related to the duplication of gene segments or

rearrangement of coding regions. The duplication of Solyc01g005730.1.1 generated an apparent pseudogene (Solyc01g005760.1.1) encoding a protein disrupted by a premature stop codon. The CC-NBS-LRR gene Solyc09g098130.1.1, located on chromosome 9, was generated by the duplication of Solyc09g098100.1.1. This gene is c. 11kb long, comprising two exons and a large intron of c. 7 kb. Solyc09g098130.1.1 lacks the intron. Four tandem duplication pairs and seven blocks of duplication impacted 15 loci containing pathogen recognition genes similar to cloned functional genes: Bs4, Cf2, Cf5, Cf9, Cf9b, Hero, I2, SlEix1, SlEix2, Mi1.2, Pto, Prf, Sw5, Tm2, Tm2a, Ve1, Ve2, Gpa2, Gro1.4, R1, R3a, Rx1, Rx2, Ry1, Rpi_blb1 and Rpi_blb2 (Figure 2 and Table 2). Most of these clusters/arrays are < 100 kb in size (median cluster size: 98 kb), with a mean size of 134 kb. Duplication events involving the two super-clusters (I2 and Mi) were analysed in more detail. Figure 4 presents the region of chromosome 11 flanked by markers Tg36 and Tg26 (c. 1.4Mb) that includes the I2 super-cluster (c. 387kb). Gene members of the I2 super-cluster have an average nucleotide diversity of 0.372. The organization of this super-cluster revealed inconsistencies with previously reported data (Simons et al., 1998). The first subcluster consists of the homologous I2C6 (Solyc11g069920.1.1), I2C5 (Solyc11g069990.1.1), and I2C4 (Solyc11g070000.1.1) and the receptor-like kinase RLK-1 (Solyc11g069960.1.1), distributed along a 54-kb region. The second subcluster includes the I2 gene (Solyc11g071430.1.1), which confers resistance to Fusarium oxysporum f. sp. lycopersici and is closely associated with the Tg105 marker, and the homologues I2C2 (Solyc11g071420.1.1), I2C3 (Solyc11g071400.1.1) and I2C7 (Solyc11g071390.1.1). This subcluster encompasses 28 kb. A receptor-like kinase, RLK-2 (Solyc11g071880.1.1), is also located between markers Tg105 and Tg26. The analysis of orthology (described in the section on Synteny between tomato and potato candidate pathogen recognition genes) revealed that I2C5 was the ancestral gene of the I2 cluster, because of its orthologous relationship to Sotub11g027150.1.1 in potato (nucleotide identity: 73.7%; amino acid identity: 85.013%). The functional I2 gene may have originated from a reverse duplication of a fragment of the I2C5 gene (identical sites (i.s.): 88.2%; overall pairwise identity (p.i.): 80.9%). The I2C2 gene may have originated by gene duplication of the functional I2 gene (i.s.: 78.3%; p.i.: 77.8%). I2C1 and I2C4 resulted from a reverse duplication of a fragment of I2 (i.s.: 96.2%; p.i.: 76%) and I2C2 (i.s.: 96.8% p.i.: 85.4%), respectively. Six tomato CNL genes (Solyc06g008400.1.1, Solyc06g008 450.1.1, Solyc06g008480.1.1, Solyc06g008770.1.1, Solyc06g008 790.1.1 and Solyc06g008800.1.1) and one NL gene (Solyc06g008380.1.1) form the Mi super-cluster, with an average nucleotide diversity of

0.066. More than 432 kb of sequence was involved in the intra-chromosomal duplications that created this super-cluster (Figure 5). The cluster Mi1.2, associated with the Tg216 marker, comprises two subclusters (Sl06-1 and Sl06-2). Sl06-1 consists of the homologues Mi1.4 (Solyc06g008380.1.1), Mi1.7 (Solyc06g008480.1.1), Mi1.5 (Solyc06g008400.1.1), and Mi1.6 (Solyc06g008450.1.1), an orthologue to Rpi-blb2, which confers resistance to Phytophthora infestans in potato. Sl06-1 is distributed along a region of c. 59 kb. Sl06-2 includes the Mi1.2 gene (Solyc06g008770.1.1; which confers resistance to Meloidogyne incognita and Paratrichodorus minor), and the homologues Mi1.3 (Solyc06g08790.1.1) and Mi1.1 (Solyc06g008800.1.1), and is distributed across a region of c. 47 kb. Analysis of orthology revealed that Mi1.2 was the ancestral gene of the super-cluster (nucleotide identity: 84%; amino acid identity: 76.4%). The identity between the Mi1.2 functional gene copy and its homologues was on average 84.8%. Phylogenetic analysis also allowed us to reconstruct the gene duplication events that occurred in this region on chromosome 6. The Mi1.1 gene may have originated by gene duplication of Mi1.2. (i.s.: 95.9%; p.i.: 95.8%). Mi1.5, Mi1.6 and Mi1.7 originated from two duplication events involving Solyc06g 008400.1.1 and Solyc06g008450.1.1 (i.s.: 95.8%; p.i.: 95.9%), Solyc06g008400.1.1 and Solyc06g008480.1.1 (i.s.: 92.7%; p.i.: 91.9%), and Solyc06g008480.1.1 and Solyc06g008450.1.1 (i.s.: 93.1%; p.i.: 92.4%), respectively.

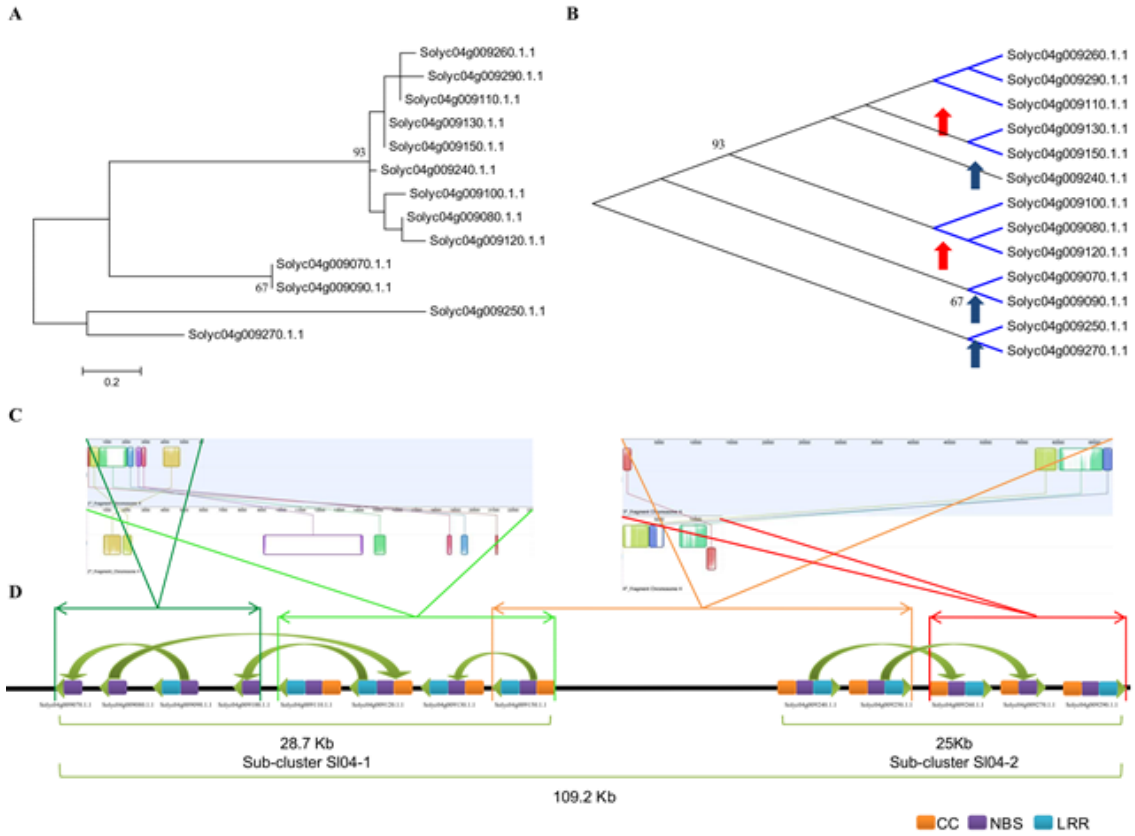

Figure 3. Block of duplication involving 13 R-genes (8 CC-NBS-LRR, 1 CC-NBS, 1 NBS-LRR and 3 NBS) on chromosome 4 analyzed with MAUVE. (A) The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model in MEGA5. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. (B) Topology of phylogenetic tree. The blue arrow indicates the location of a duplication involving two genes, the red arrows indicates two ancestral gene duplication among genes located in the sub-cluster Sl04-1 and Sl04-2. (C) Alignments of two pairs of fragments of S. lycopersicon L. var. Heinz 1706 chromosome 4. The region aligned whit MAUVE, were evidenced in gene cluster schematization (D) by light green and dark green for sub-cluster Sl04-1, red and orange for sub-cluster Sl04-2. Similar locally collinear blocks are labeled with the same color and connected by fine lines. The boundaries of colored blocks indicate the breakpoints of genome rearrangement, unless sequence has been gained or lost in the breakpoint region. (D) Schematization of gene duplication occurred in the sub-clusters Sl04-1 and Sl04-2. The most probable gene duplication events are highlighted by arrows.

Name of cluster or array	Chromosme	N. R-genes for cluster (protein class)	Markers associated	Size cluster Kb
BS4	5	2 (TNL); 1 (RLK); 1 (RLP)	TG432	134.2
CF2_CF5	6	2 (RLP)	GP79; CT119	25.1
CF9_CF4	1	6 (RLP)	CT223; TG236	97.9
GPA2	12	2 (CNL)	TG79; TG100	11.2
GRO1.4	7	4 (Unknown)	CP56; GP27	157.3
HERO	4	2 (CNL); 6(Unknown)	CT229	80.9
I2_RX2	11	2 (CNL); 4 (Unknown); 1 (RLK)	TG36; TG105	387.1
LE EXI-1_LE EXI-2	7	5 (RLP)	TG131; TG61	32.2
MI1.2_RPI BB-2	6	7 (CNL)	TG216; Ctof8f19	431.9
PTO_PRF	5	2(CNL); 3(Unknown)	TG475; CT229	69 Kb
R1	5	3 (CNL); 1 (Unknown)	TG441; TG431	182.6 Kb
RPI BLB1	8	3 (RLK); 2(CNL)	CT88	40.2 Kb
SW5	9	2 (CNL); 1 (Unknown)	CT220	79.3 Kb
TM2_TM2A	9	3 (CNL); 1 (Unknown); 1 (TNL)	TG591	102.5 Kb
VE1_VE2	9	2 (RLP); 1 (CNL)	CT143;	174.7 Kb

Table 2. List of clusters and arrays that showed the presence of one or more proteins with high similarity to functional proteins. For each clusters/array, the chromosome, the number and the class of R-genes retrieved within the cluster , the markers associated and the size is reported.

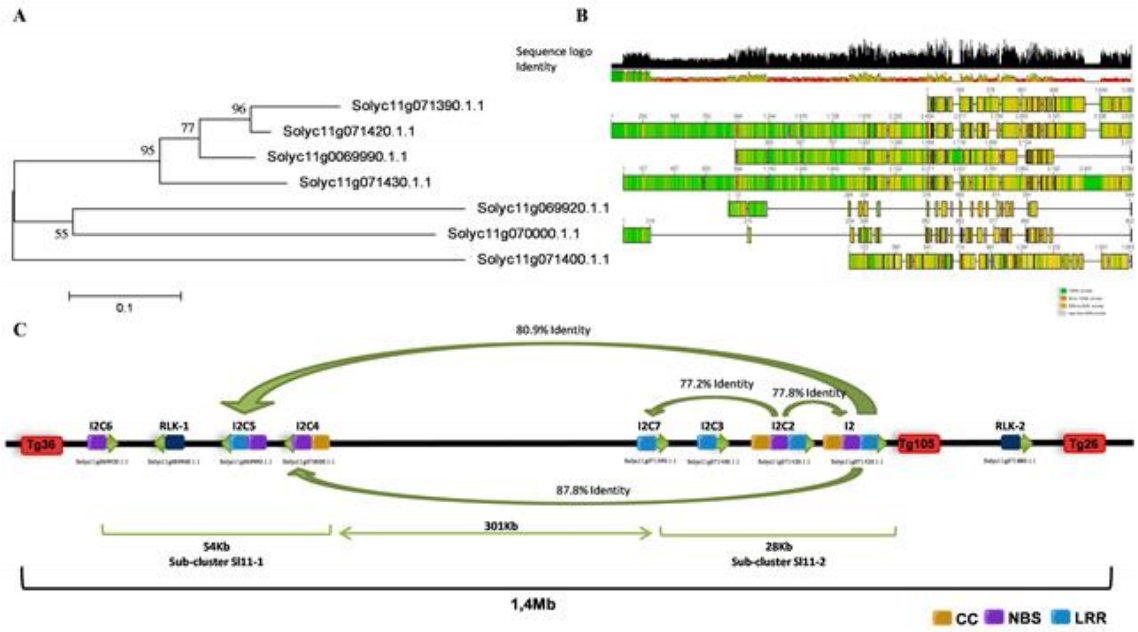

Figure 4. Reconstruction of gene duplication history of I2 cluster. (A) Phylogenetic analysis obtained by Maximum Likelihood method, based on the Jukes-Cantor model, for homologue sequences of I2 cluster. Reliability of internal branches was evaluated using the bootstrap method, with 100 replicates. (B) Multiple alignment of the I2 gene and 6 I2 homologues (C) Schematic representation of gene duplication occurred in the genomic region included betweenTg36 and Tg26 markers.

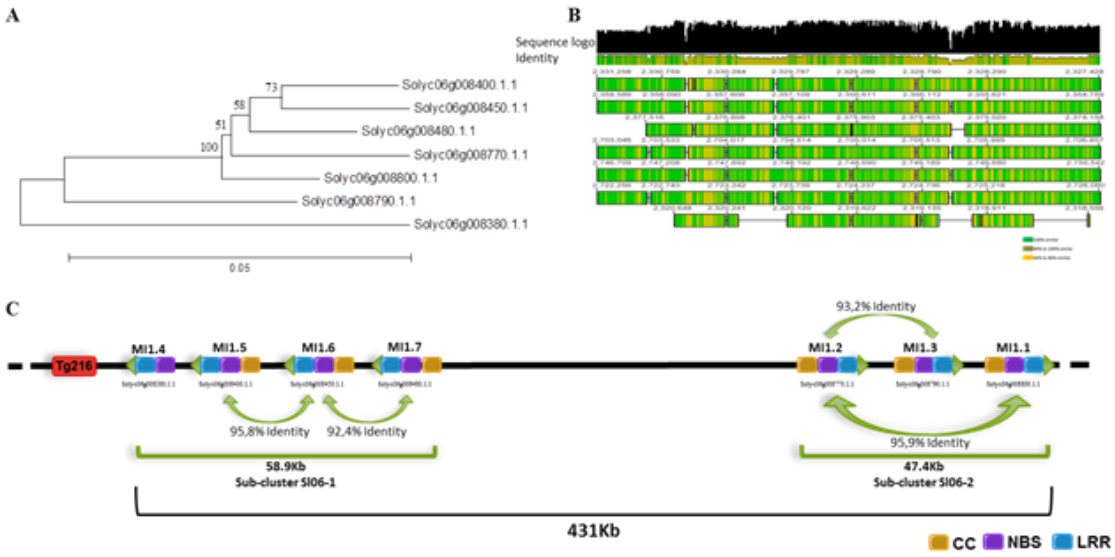

Figure 5. Reconstruction of the gene duplication history of the tomato MI1.2 cluster. (A) Phylogenetic analysis obtained using the Maximum Likelihood method, based on the General Time Reversible model, for homologous sequences of MI cluster. Bootstrap values (100 replicates) are indicated above branches. (B) Multiple alignment of the MI1.2 gene and six MI homologues. (C) Schematic representation of postulated gene duplication events that occurred in the genomic region associated with the Tg216 marker.

2.3.3. Phylogenetic analysis

For phylogenetic analyses, predicted tomato protein sequences were aligned by structural class. Relevant reference protein sequences originating from cloned pathogen recognition genes (www.prgdb. org) were included in each alignment. Reference sequences were deduced from pathogen recognition gene sequences originating from tomato, potato, rice, wheat (Triticum aestivum), lettuce (Lactuca sativa), barley (Hordeum vulgare), maize (Zea mays), Arabidopsis thaliana, Capsicum chacoense, Aegilops tauschii, Helianthus annuus and Glycine max. The number of proteins analysed per class varied greatly. A phylogenetic tree containing 119 CNL and NL tomato protein sequences and 42 relevant reference sequences was generated (Figure 6). Clustering of the resultant protein sequences into robust clades supported by bootstrap values ≥ 60% allowed definition of clusters A to M (Figure 6). The placement of clusters A to F suggests that differentiation of the encompassed CNL genes occurred after Solanaceae speciation, as CNL proteins from other plant families are not included. Cluster A comprises two subclusters. The first includes Mi1.2, Rpi_blb2 and homologous proteins encoded on chromosome 6. The second subcluster includes proteins encoded on chromosome 5. These too share homology with Mi1.2 and Rpi_blb2, suggesting that a translocation event could have resulted in genes of a common evolutionary origin occupying two distinct genome locations. It is striking to note that cluster A shares a common ancestor with clusters B and C, which comprise members of the Hero family

encoded on chromosome 4 and the Sw5 family on chromosome 9, respectively. Ancestral to the A, B, and C clusters are proteins assigned to cluster D, including R1 and Prf and other proteins encoded on chromosome 5. In this phylogenetic reconstruction it is difficult to establish the history of cluster E, which comprises three subclusters. Two the third (E2) encompasses Bs2, Rx1, Rx2 and Gpa2. The E3 subcluster shares a more recent common ancestor with cluster E2 than cluster E1. The E3 cluster contains proteins encoded on nine chromosomes, while E1 contains proteins encoded on chromosomes 1, 3, 4, and 10. In the small cluster E2, four tomato homologues of the Bs2 protein were identified, but no close homologues of Rx1, Rx2 and Gpa2 are evident. Finally, cluster F includes the reference protein Tm2a and close homologues encoded on chromosome 9. Notably, a number of proteins similar to A. thaliana reference proteins were found. Cluster G includes homologues of Rcy1, Rpp8, Hrt and Rpp13 while cluster H includes homologues of Rpm1. Tomato homologues to cereal proteins were not identified in cluster H. The small cluster I included tomato proteins lacking homology to previously described CNL proteins. A large number of proteins are included in cluster L. This group of proteins might be of a more ancestral origin that predates Solanaceae speciation, as it includes members belonging to both the Solanaceae and several other taxa. A homologue of the Triticum PM3 protein and several homologues of the Rpi_blb1, I2 and R3a proteins are included in cluster L. Analysis of amino acid sequences of these proteins revealed a very conserved motif of 12 amino acids in the LRR domains (Fig. S1). The proteins Cre1, Cre3, Rp1d, Xa1, Rps1k1, Rps1k2 and P18 do not have close homologues in tomato and are separated from the rest of cluster. Finally, cluster M contains proteins similar to the Dm3, Rps2 and Rps5 proteins. Phylogenetic analysis of TNL proteins was conducted on 18 tomato sequences and 13 reference sequences encoded by previously cloned genes reported in the PRG database (Figure 7). Three main clusters were identified. Cluster A includes all previously described Solanaceous TNL proteins of known function. Six bifurcating nodes with high bootstrap values were observed in cluster A. Three of these subclusters are composed of Bs4, Ry1 and Gro1.4 and their respective homologous proteins. The Gro1.4 subcluster contains two closely related members (one encoded on chromosome 2 and one encoded on chromosome 4). Tomato protein sequences similar to N were not identified. Cluster B includes sequences similar to the L6 and M proteins and is split into four subclusters comprising pairs of highly similar proteins. Cluster C lacks tomato proteins. The phylogenetic relationships of 169 RLPs proteins based on an alignment of the C3-F region of tomato and 15 characterized RLPs. The phylogenetic tree revealed seven distinct RLPs clades with an average identity of

50% and bootstrap support of 60% or greater. Each of these clades contains at least one functionally characterized member, allowing us to infer possible functions for the proteins in each clade. The phylogenetic analysis of the RLKs, based on alignment of the kinase sequences of all tomato RLKs and eight functional RLK reference sequences. The phylogenetic tree revealed seven distinct RLKs clades encompassing subclusters (E1 and E3) lack proteins of known function while several proteins with similarity to pathogen recognition proteins of known disease resistance function.

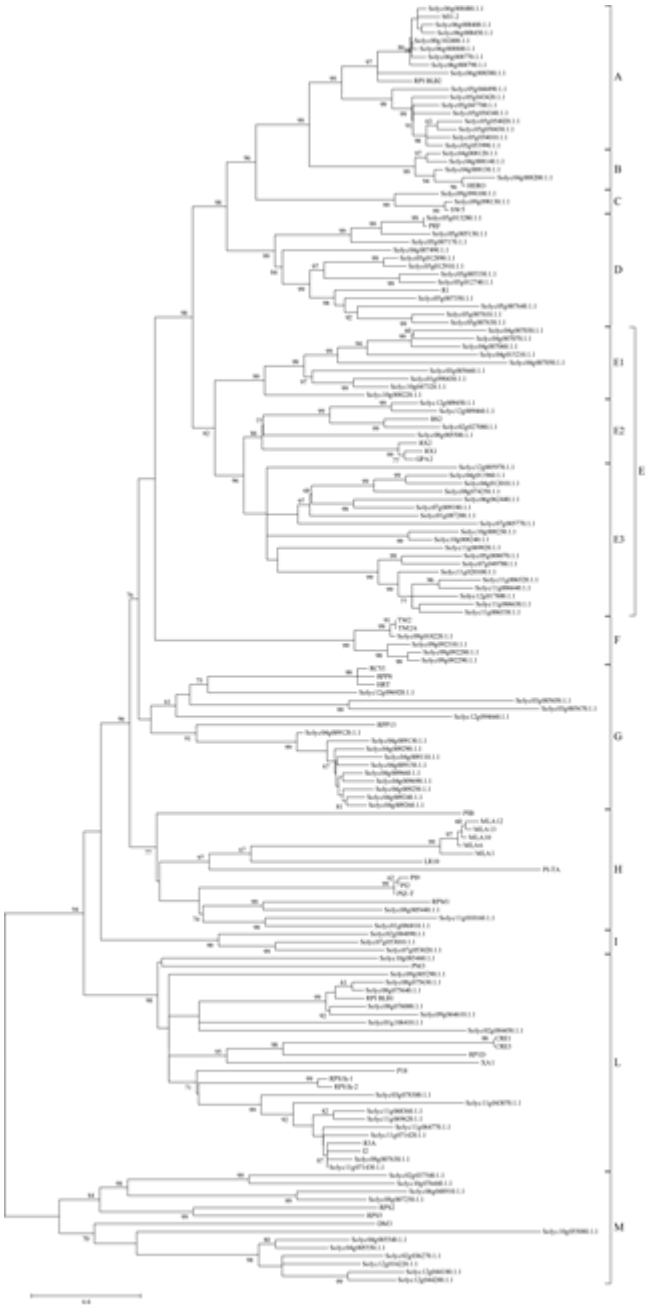

Figure 6. Evolutionary history of tomato CNL genes. The CNL evolutionary history was inferred by analysis of 159 aligned amino acid sequences using the Maximum Likelihood method based on the Whelan And Goldman + Freq. model. Evolutionary analyses were conducted in MEGA5. Bootstrap values >60% are indicated above branches. The tree is drawn to scale, with branch lengths measured in the number of

substitutions per site (note scale). Identified clades are indicated by letter from A to M and delineated by vertical lines.

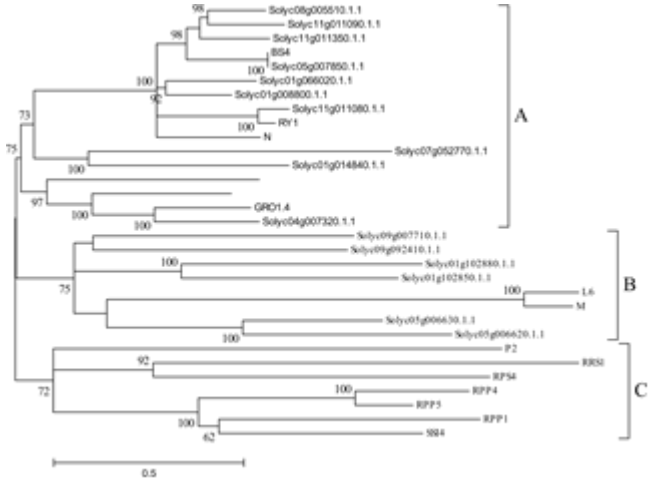

Figure 7. Evolutionary history of tomato TNL genes. The TNL evolutionary history was inferred by analysis of 31 aligned amino acid sequences using the Maximum Likelihood method based on the Whelan And Goldman + Freq. model. Evolutionary analyses were conducted in MEGA5. Bootstrap values >60% are indicated above branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (note scale). Identified clades are indicated by letter from A to C and delineated by vertical lines.

2.3.4. Potato annotation

Following the procedures used for tomato, 1331 putative pathogen recognition proteins were identified from the potato iTAG proteome. The chromosomal distribution of CNL, RLP and RLK genes (based on chromosome size calculated at $P < 0.001$) was nonrandom (CNL v2 = 89; RLP v2 = 109; RLK v2 = 97). Comparing the number of putative pathogen recognition genes on each chromosome of tomato and potato revealed similar distribution trends. All classes, except RLK, were predominant in potato. Few chromosomes showed a predominance of species-specific classes in one or the other species. The TNL class was not detected on chromosome 3 or 10 in either species. In potato, 777 genes (c. 58%) reside in 91 clusters, with an average number of 8.5 genes per cluster.

2.3.5. Synteny between tomato and potato candidate pathogen recognition genes

A Best-BLAST approach was employed to identify candidate NBS, RLP and RLK protein pairs of maximal homology and possible orthology between potato and tomato. In defining probable orthologues, we required that orthologous proteins belong to the same structural class and have a pairwise identity of at least 90% for NBS and TIR proteins or 80% for other proteins. Based on this analysis, we identified 319 putative orthologous pairs. Orthology was

further explored by phylogenetic analysis. Corresponding candidate orthologous gene pairs were also filtered for chromosome location as, in some documented cases, orthologous pathogen recognition genes have been shown to occupy corresponding genome locations in tomato and potato (Grube et al., 2000; Huang et al., 2005). Based on these criteria, 304 candidate pathogen recognition protein pairs were selected as bona fide orthologues (Figure 8). The corresponding orthologous pathogen recognition gene pairs were distributed over c. 97% of the tomato genome and 93% of the potato genome. The genomic region analysed in tomato was 1.2 Mb, while the corresponding genes together covered a region of 1.3 Mb of the potato genome. Genomic rearrangements are highlighted on the candidate pathogen recognition gene physical synteny map (Figure 8). In some cases, genomic rearrangements associated with speciation of tomato and potato from a common ancestor have impacted the location and organization of pathogen recognition gene homologues. For instance, in tomato, I2 homologues form a single super-cluster. However, in potato, the corresponding genes comprise two clusters of three and 15 genes. Evolutionary patterns at codon sites of six gene members belonging to the tomato Mi1.2 and I2 super-clusters indicated that all gene pairs are under a general process of purifying selection.

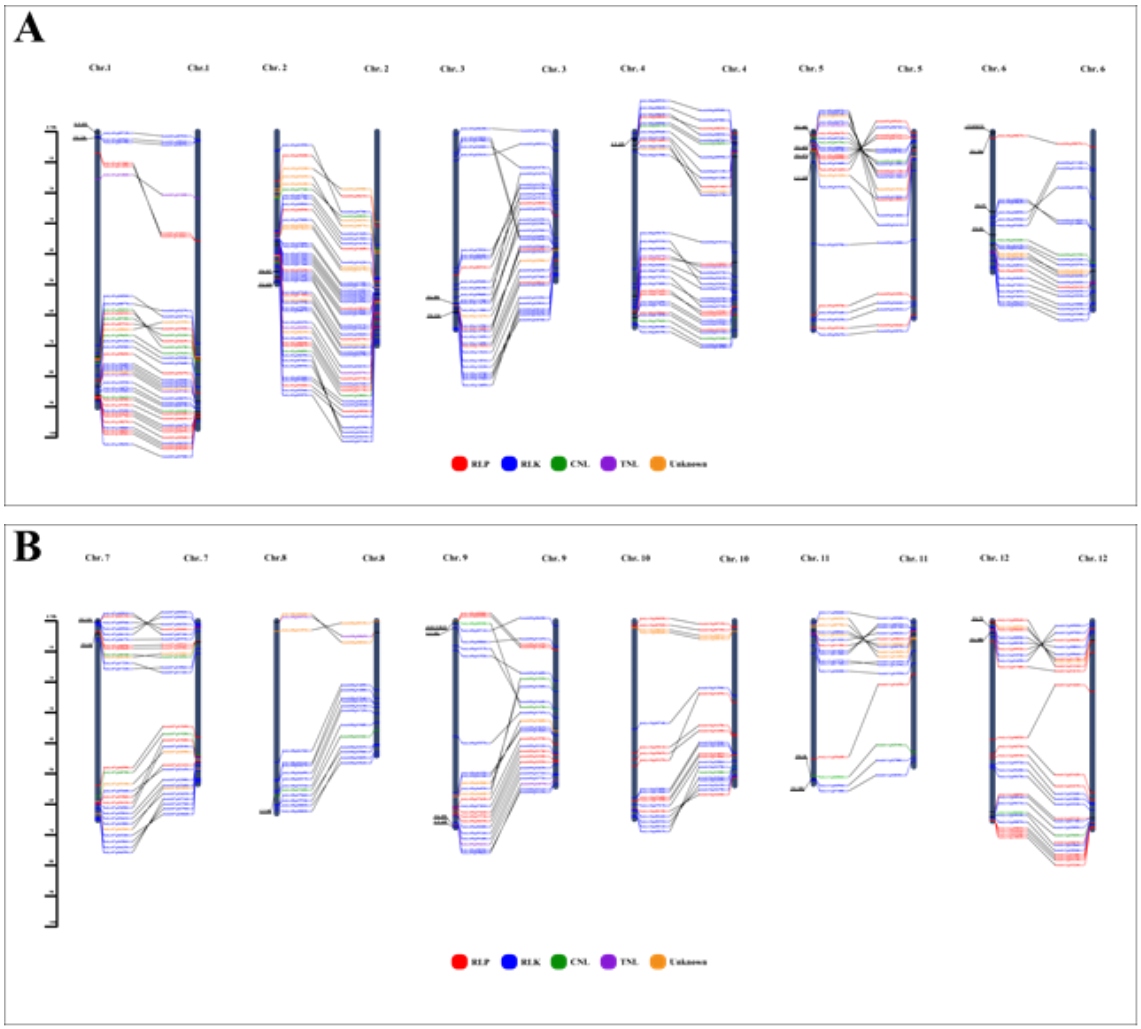

Figure 8 Physical map of orthology between tomato and potato. Comparative genome positions of orthologous genes identified between two species for the first six pairs of chromosomes (A) and the following six pair of chromosomes (B).

2.4. DISCUSSION

2.4. Discussion

We have undertaken an R-domain-based annotation screening of the entire tomato proteome. In total, 769 diverse proteins containing domains similar to those found in known pathogen recognition proteins were identified. A previous survey indicated that the tomato genome encodes approximately the same number of NBS or TIR genes as the A. thaliana genome, but fewer than the potato, rice, and grape (Vitis vinfera) genomes (The Tomato Genome Consortium, 2012). Genomes also varied substantially in the number of RLP and RLK genes encoded (Fritz-Laylin et al., 2005; Wang et al., 2008a; The Tomato Genome Consortium, 2012). Collectively, these studies confirm that the pathogen recognition gene component of plant genomes is large and diverse and under pressure for high rates of change. Indeed, gene evolution and diversification provide unique pathogen recognition gene patterns. We performed a Solanum whole-genome survey of gene classes reported to be involved in disease resistance response (NBS, RLP and RLK), confirming data already reported (The Tomato Genome Consortium, 2012), and conducted deep analyses of the individual protein classes, providing an unprecedented view of candidate pathogen recognition gene dynamics. Genome distribution and arrangement: impact on pathogen recognition gene function and evolution The distribution of genes, in general, is nonrandom in plant genomes (Sidhu & Gill, 2005). Consistently, our analyses showed that the various classes of candidate pathogen recognition genes are differentially represented on each of the tomato chromosomes. Most of the tomato NBS genes are organized in tandem arrays, clusters, and super-clusters, confirming observations reported previously (Grube et al., 2000; Pan et al., 2000). Our analyses indicate that CNL and RLP genes are most likely to be clustered in the tomato genome. However, we also observed that some candidate pathogen recognition gene clusters comprise genes of different evolutionary origin and structure. Specifically, NBS genes frequently occur in clusters of related gene copies and these gene clusters often also contain RLP or RLK genes. The origin of structurally heterogeneous gene clusters is not clear. Importantly, the predominance of specific gene classes in any genome region might correlate with their biological role (Singer et al., 2005; Yi et al., 2007). Thus, pathogen recognition genes from different classes residing together within a common chromosome region could, in some instances, jointly exert the same function, such as conferring resistant to a specific pathogen. The Prf/Pto locus provides an illustrative example. Prf, an NBS-LRR gene, lies within the Pto gene cluster comprising five kinase genes, and Prf function is required for Pto-mediated resistance to P. syringae pv. tomato strains carrying avrPto (Salmeron et al., 1996). Intriguingly, clustering of tomato NBS and

RLP or RLK genes is also compatible with the existence of selective pressures optimizing coordinated gene transcription of functionally related genes (Al-Shahrour et al., 2010). The identification of CNL, RLP and RLK genes located in close proximity in this study allows future testing of hypotheses of coregulation of structurally distinct but physically linked pathogen recognition genes. Pathogen recognition gene evolution has both a vertical component across generations and a horizontal component throughout the genome, and each is likely to be shaped by natural selection for resistance function (Bergelson et al., 2001). Pathogen recognition gene diversification results from the generation of pathogen recognition gene sequence variation and subsequent selection for effective sequence combinations. Thus, observed pathogen recognition gene frequency in a plant species varies as a function of pathogen pressures and genotypes. Fluctuations in these factors over time result in maintenance of pathogen recognition gene allele frequency as a dynamic polymorphism (Stahl et al., 1999). Larger gene aggregates provide more potential for the generation of novel sequence variation through recombination. In other words, pathogen recognition gene clustering facilitates the generation of new pathogen recognition genes upon which pathogen selective pressures act. The significance of pathogen recognition gene clustering in their adaptation is further underscored by the fact that, to date, all cloned tomato pathogen recognition genes with known resistance function have been found to exist in clusters or gene arrays within the genome. The pattern of tomato CNL gene evolution suggests that few ancestral genes underwent local amplification. Indeed, a portion of the CNL gene contingent has experienced recent but dramatic gene rearrangements, with duplications of individual loci being a significant source of sequence variation. Tomato pathogen recognition gene clusters can be both large, occupying large chromosomal regions, and functionally diverse, harbouring diverse genes encoding resistance to several different pathogens (Gebhardt & Valkonen, 2001). In some cases, structurally divergent pathogen recognition genes may have been brought together by chromosomal rearrangements (Hulbert et al., 2001). Investigation of the functional and spatial arrangement of pathogen recognition gene clusters in the tomato genome can be useful in reconstructing the history of chromosomal rearrangements that shaped the genome architecture of the extant species. Importantly, analysis of CNL and TNL phylogenetic relationships revealed that most NBS gene clades contain one or more cloned genes of known resistance function. This finding suggests that identifying new sources of resistance will be aided by ancestral reconstructions in order to understand mechanisms of adaptive selection and functional divergence in their evolutionary past. At many complex loci, most gene copies comprised a putative open

reading frame. In some instances, intact paralogous or truncated pathogen recognition gene copies may possess as-yet-undetermined disease resistance function. Identification of these genes through analysis of the whole genome sequence provides a treasure trove for future efforts aimed at identifying tomato genes of novel resistance function. Our analyses revealed that some tomato chromosomes have experienced a substantial number of duplication events involving candidate pathogen recognition genes. This observation is consistent with previous reports from other plant species (Nobuta et al., 2005; Yu et al., 2005; Ameline-Torregrosa et al., 2008; Yang et al., 2008). Our analyses also revealed that segmental duplication is associated with some of the largest gene clusters. The identification of duplication events provided a basis for understanding lineage-specific pathogen recognition gene development in an evolutionary context. Duplication events involving tandem candidate pathogen recognition gene pairs appear to be the predominate source of larger clusters or novel cluster loci in tomato. CNL and RLP genes have been more frequently involved in tandem duplication events than other classes. Generation of larger pathogen recognition gene clusters can lead to functional specialization and acquisition of novel disease resistance functions. Together, these findings confirm that tandem duplication plays an important role in the expansion of pathogen recognition gene clusters in plant genomes. However, our analyses suggest that duplication has not been uniformly important in the evolution of all types of NBS-LRR genes in tomato. Among Solanaceae CNL genes, some cloned genes originate from clades showing patterns of significant gene expansion (e.g. Mi, R1 and Rx1), while others do not (e.g. Tm2a and Sw5). In contrast, TNL genes show a very low expansion capacity. In contrast to the duplication patterns observed for CNL genes, our analyses indicate that duplication of tomato TNL loci typically entailed a single gene event, leading to phylogenetically isolated tandem gene pairs, while other TNL genes have been maintained as single loci. Candidate pathogen recognition genes of high sequence similarity distributed to distant genome regions and mixed pathogen recognition gene clusters were also observed in the tomato genome. These patterns of distribution are thought to result from ectopic duplications and large-scale segmental duplications with or without subsequent rearrangements. Some of these rearrangements may have been associated with segmental inversions of entire chromosomal regions, events that seem to have occurred frequently throughout speciation in the Solanaceae (The Tomato Genome Consortium, 2012). We investigated specific, well-known pathogen recognition gene clusters to reconstruct the main duplication events and to trace the evolution history of these loci. Importantly, combining gene proximity analyses and phylogenetic profiling yielded a revised view of the

organization of the I2 and Mi loci. These analyses also suggested mechanisms that led to lineage-specific diversification, providing insights into adaptive processes. Based on our analysis, the I2 cluster also contains RLK members and the tomato and potato I2 orthologues had different duplication fates. In particular, while tomato I2 gene copies belong to a single super-cluster, potato I2 orthologues are assigned to two clusters. By contrast, Mi homologues display lower sequence divergence among tomato gene cluster members and similar genomic organization in tomato and potato. The availability of genome sequences from both tomato and potato enabled the first step to be taken in genome-wide evolutionary analysis of candidate pathogen recognition genes in the Solanaceae. Several surveys of potato NBS-LRR-encoding sequences have already been performed (Xu et al., 2011; Jupe et al., 2012; Quirin et al., 2012). Each of these is based on a different prediction system which makes comparison difficult (Jupe et al., 2012). Based on parameters used in our analysis, using iTAG proteome annotations, the number of candidate pathogen recognition genes (genes encoding at least one pathogen recognition domain) varies considerably between tomato and potato. Specifically, our analyses reveal that tomato encodes fewer candidate pathogen recognition genes than potato (769 in tomato vs 1331 in potato). Polyploidization, genome size variation, natural selection, artificial selection including domestication, breeding and cultivation, and gene family interactions have probably influenced pathogen recognition gene evolution in Solanum. NBS gene lineages tested in sequence and cross-hybridization experiments in Solanum species were shown to be of an ancient origin that predated speciation events in the Solanaceae (Quirin et al., 2012). Nevertheless, large- and small-scale rearrangements in the tomato and potato genomes over time could have hidden exact orthologous gene relationships in many cases, leading to pairs of genome regions or gene clusters between the two species that share a number of homologous genes, but in which neither gene order nor content is strictly conserved. The sequencing of the genomes of additional Solanum species, especially wild relatives of tomato and species that diverged from the common tomato– potato progenitor > 7 million yr ago, will yield data that will be very informative in exploring patterns of pathogen recognition gene evolution in tomato. Synteny comparison between the tomato and potato genomes indicates that deletion of roughly one-third of the redundant proteome and transposable elements in euchromatic regions has occurred in tomato (The Tomato Genome Consortium, 2012). Genome-wide deletion of redundant NBS-encoding genes has been observed in several plant genomes. No or few NBS-encoding genes have been identified in the WGD blocks in rice (Yu et al., 2005), grape (Yang et al., 2008), A. thaliana (Nobuta et al., 2005), and poplar (Populus

trichocarpa) (Yang et al., 2008), indicating that most duplicated NBS-encoding genes were lost soon after WGD events. Loss of duplicated gene copies from the tomato genome probably accounts for the stark difference in the number of total pathogen recognition genes discovered in the tomato and potato genomes. Analysis of DNA sequence data suggests that the last common ancestor of Solanum, Petunia, and Nicotiana existed c. 30 million yr ago (Wang et al., 2008b), while divergence of tomato and potato has occurred in the last c. 7 million yr (The Tomato Genome Consortium, 2012). Thus, loss of duplicated pathogen recognition loci from the tomato genome appears to have occurred relatively rapidly, in evolutionary terms. Functional redundancy or artificial selection during domestication could have increased the speed of deletion of redundant candidate pathogen recognition genes following WGD in Solanum species. Indeed, a higher average number and a different arrangement of gene members were found in potato candidate pathogen recognition gene clusters. The residual copies probably provide a sufficient number of genes for specific pathogen defence in tomato. Differences or similarities in gene order in the genomes of related plant species allow prediction of functional relationships between genes and phylogenetic relationships between genomes (Stoye & Wittler, 2009). However, the study of processes governing evolutionary histories of genes involved in pathogen recognition across related plant taxa is currently incomplete. To date, only a few studies have focused on the variation and evolution of gene family sizes in plants (Li et al., 2010; Zhang et al., 2010; Guo et al., 2011). More thorough, systematic surveys of NBS, RLP and RLK gene content and genome-wide organization will provide insights into the impact genome arrangement has on gene and species evolution. Most identified tomato/potato NBS, RLP and RLK orthologues are distributed at less than perfectly matching positions across the two genomes, suggesting that conserved gene order on chromosomes is a mere vestige of common ancestry. While tomato is a diploid, the cultivated potato is an autotetraploid. Importantly, gene loss in polyploid genomes is known to be nonrandom (Blanc & Wolfe, 2004; Seoighe & Gehring, 2004), and polyploidy followed by the genome-wide removal of some, but not all, redundant genomic material can result in species-specific differences in the function of homologous genes (Adams & Wendel, 2005). The distribution of NBS, RLP and RLK genes in the genomes of Solanum species and the discovery of redundant genes provide insight into how novel pathogen-resistant genes can evolve via subsequent gene duplication events, ectopic recombination, unequal crossing over, and diversifying selection (Michelmore & Meyers, 1998). The tetraploid nature and biological characteristics of potato could have also accelerated the candidate pathogen recognition gene mutational processes. Lineage-specific

amplification of NBS-encoding genes in plant genomes is an ongoing process, as emphasized by local expansion of TNL genes in the potato genome. The detection of potato-specific TIR genes indicates that domain acquisition has resulted in biological innovation. Together, genome duplication events, autotetraploidization of potato, adaptation to agricultural practices, and differing pathogen pressures could have impacted the pathogen recognition gene contingents of the tomato and potato genomes, resulting in differences in both the total number of candidate pathogen recognition genes harboured in these two species and the functional divergence of candidate pathogen recognition gene homologues. In conclusion, comparisons of gene arrangements between related plant genomes offer insights into a number of questions regarding how complex biological systems evolve and function: spatial analyses of orthologous genomes focus on elucidating evolutionary processes and history, and comparative physical and genetic mapping facilitates the transfer of knowledge between organisms (Murphy et al., 2004). The complete sequencing of the tomato and potato genomes permitted an unprecedented view of Solanum NBS, RLP and RLK genomic architecture that will have implications for understanding the evolutionary history of pathogen recognition genes in these and other Solanaceous species. This will also impact molecular breeding efforts.

3. REYDEFINING THE FULL TOMATO NB-LRR RESISTANCE GENE REPERTOIRE

3.1. Introduction

The first draft tomato genome assembly revealed a large size of the NB-LRR gene family, and thus awide potential R gene repertoire (Andolfo et al. 2013). A first tomato R gene annotation (Andolfo et al. 2013) was reported based on the existing automated gene and protein predictions of the Tomato Genome Consortium (TGC 2012).

Recently, it was demonstrated that the automated gene and protein predictions for the potato reference sequence failed to reveal over 300 potential NB-LRR genes in potato, using the Resistance gene enrichment and sequencing (RenSeq) approach (Jupe et al. 2012). The RenSeq method facilitates obtaining sufficient sequence depth to sequence the many NB-LRR genes that exist in multigene families (Jupe et al. 2012). However, even when RenSeq data was used to map the resistance to specific loci, it is still challenging to define the sequence of each paralogue in a multigene family.

In this study, we adopted an improved version of the RenSeq approach in combination with Illumina MiSeq 250bp paired-end sequencing on genomic DNA (gDNA) and on cDNA of the two sequenced tomato genomes S. pimpinellifolium LA1589 and S. lycopersicum Heinz 1706. RenSeq on gDNA allowed us to correct about 25 % of the previously described tomato NB-LRR genes and to identify 105 novel genes from previously unannotated regions. We further report the first comprehensive study of the phylogenetic relationship between the individual NB-LRR genes in S. pimpinellifolium LA1589, S. lycopersicum Heinz 1706 and Arabidopsis thaliana. An important result for future applications of RenSeq was the reduction of sequence data complexity by enriching NB-LRR genes from cDNA, thus avoiding sequence analysis of non-expressed paralogues.

3.2. Materials and methods

3.2.1 Plant material and preparation of RenSeq libraries

Fully expanded leaves of S. lycopersicum Heinz 1706 and S. pimpinellifolium LA1589 were detached from 3-week old glasshouse grown plants. Three leaves were inoculated with two 20 µl-drops per leaflet of water, or a suspension of P. infestans isolate 2006_3928A (50,000 zoospores/ml). One inoculation spot per leaflet was harvested 24 hours post-inoculation as leaf discs with 10 mm in diameter, and frozen in liquid nitrogen. The remaining spots were observed at 6-dpi for successful colonisation with P. infestans. Leaf discs of both treatments were mixed and RNA was extracted using the TRI-reagent (Sigma-Aldrich) and Directzol RNA Mini-prep (Zymo Research), following manufacturers recommendations. First-strand cDNA was made using oligo-dT and random hexamer primers and First-Strand Superscript II (Sigma-Aldrich). The second strand was made as described in (Park et al. 2005). gDNA was extracted from young leave tissue of the same plants, using the DNeasy Plant Mini kit (Qiagen), following manufacturers recommendations. Illumina MiSeq libraries were prepared using the NEBNext Ultra DNA library prep kit (NEB) using 2 – 3 µg starting material. Libraries were multiplexed using the NEBNext Multiplex Oligos for Illumina (Index Primers Set I). Up to three libraries were pooled and NB-LRR like sequences were captured as described in Jupe et al. (2013) using Agilent SureSelect kit with an updated bait library comprising 28,787 unique 120-mer oligos. Enriched libraries were amplified up to 1 µg, and sent for MiSeq 250-bp paired end sequencing at The Genome Analysis Center (TGAC, Norwich Research Park, UK).

3.2.2. Identification and annotation of NB-LRR genes in Solanum spp.

All Illumina MiSeq data analysis was carried out using the Sainsbury Laboratory instance of the Galaxy project if not stated otherwise (Okayama and Berg 1982). To identify and annotate NB-LRR loci in the Tomato genome (TGC 2012), NB-LRR enriched paired-end Illumina MiSeq reads were mapped to the twelve chromosomes, using BWA version 1.2.2 (default parameters) (TGC_SL2.40_pseudomolecules.fasta). The mapping information (BAM-format) was imported into Geneious 6.0 and visualized per chromosome (http://www.geneious.com/). The Illumina read coverage over previously identified NB-LRRs was determined as described in Jupe et al. (2013). Potential full-length sequences were determined using the MAST output, and this was further used to identify start and stop positions for each gene. Gaps in the assembly were closed following the method described in Jupe et al. (2013). IDs for novel genes are as per definition in Jupe et al. (2013) for the R

gene discovery consortium (RDC) and include the species code RDC0002 (Heinz 1706) and RDC0003 (LA1589).

3.2.3. Analysis of cDNA RenSeq libraries

Raw high-quality MiSeq reads were mapped to the reannotated NB-LRR gene complement using Bowtie version 1.1.2 under stringent conditions (reads mapping more than once are omitted). The resulting SAM-file was filtered for mapped reads and the number was counted per NB-LRR gene. No cut-off was applied to the number of mapping reads.

3.2.4. Phylogenetic analysis

To identify the NB domain sequences used for the phylogenetic analysis, amino acid sequences of the NB domain of the reference R genes, were used to search in a BLASTx analysis with an expected value of <1e-3. Sequences with less than 50% of the full-length NB-ARC domain (Pfam database ID: PF00931) were excluded. Evolutionary analyses were conducted using MEGA5 (Tamura et al. 2011). The phylogenetic relationships of mapped NB-LRR genes were inferred separately (e.g., *S. lycopersicum* Heinz 1706 and *S. pimpinellifolium* LA1589 groups) using the maximum likelihood method based on the WAG model (Whelan and Goldman 2001). 162 Arabidopsis thaliana NB-LRR gene sequences were extracted from the TAIR database (http://www.arabidopsis.org/).

3.3. Results

3.3.1. Design and application of a tomato and potato RenSeq bait-library

In an effort to reannotate the NB-LRR gene complements of the sequenced tomato genomes Solanum lycopersicum Heinz 1706 and S. pimpinellifolium LA1589 (hence referred to as Heinz 1706 and LA1589, respectively), we designed an updated version of our customized RenSeq bait-library for NB-LRR gene targeted sequence enrichment (Jupe et al. 2013). This version of the bait-library comprises 28,787 unique 120-mer baits designed from the 260 and 438 NB-LRR-like sequences that were previously described from the tomato and potato genomes (Andolfo et al. 2013; Jupe et al. 2012). The RenSeq experiment was carried out on genomic DNA, to facilitate the reannotation of the full NB-LRR complement, and in addition on double-stranded cDNA, to test if the complexity of sequencing data for this multigene family can be further reduced by only sequencing the expressed genes. Up to five barcoded samples were combined in one SureSelect reaction, and further pooled to up to 12 single samples prior sequencing.

The resulting RenSeq libraries with an average insert size of 700 bp were sequenced on a MiSeq platform (250-bp reads). For Heinz 1706, 9,395,874 reads were produced from gDNA. Of these, 50% (4,867,603) could be mapped to the 12 (plus ch00) reference tomato chromosomes, respectively (Table 1). Similarly, for LA1589, 4,980,032 reads were derived from the MiSeq run and 34% (1,680,734) mapped to the superscaffolds. Analysis of un-mapped gDNA derived reads revealed some sequence contamination from mitochondrial and chloroplast DNA.

	Mapping Heinz 1706 reads	Andolfo et al. 2012 annotation	Novel NB-LRR	Total NB-LRR
Ch00	823,314	3 (2)	1	3
Ch01	369,154	17 (14)	7	21
Ch02	383,004	24 (16)	7	23
Ch03	334,034	8 (6)	3	9
Ch04	430,876	55 (40)	16	56
Ch05	495,739	39 (34)	11	45
Ch06	361,718	19 (17)	3	20
Ch07	202,113	21 (11)	8	19
Ch08	276,354	13 (11)	2	13
Ch09	230,882	16 (14)	9	23
Ch10	247,821	23 (19)	8	27
Ch11	451,661	34 (22)	20	42
Ch12	260,933	22 (15)	10	25
Total	**4,867,603**	**294 (221)**	**105**	**326**

Table 1. Identification of novel NB-LRR genes from MiSeq RenSeq data. BWA mapping of NB-LRR-enriched Illumina PE 250-bp MiSeq-reads to the reference S. lycopersicum Heinz 1706 aided the verification

of previously reported NB-LRR genes (Andolfo et al. 2013) (verified genes in brackets), as well as the identification of novel NB-LRR encoding sequences.

3.3.2. RenSeq data enables NB-LRR gene reannotation in HEINZ 1706 and LA1589

To locate all potential NB-LRR encoding regions, gDNA RenSeq reads were mapped to the corresponding reference genome. Sequences with read coverage higher than 20× over a minimum of 45 nucleotides were identified, and resulted in a total of 7,290 and 6,465 genomic fragments from Heinz 1706 and LA1589, respectively, that were extracted with a 500 bp extension to both ends. Overlapping sequences were concatenated and used in a MAST search to identify amino acid motif compositions that are similar to NB-LRR genes . This resulted in a total of 326 and 355 potential NB-LRR sequences from Heinz 1706 and LA1589, respectively (Table 2).

Using the available MAST motifs, genes could be classified as TNL or CNL, and presence/absence of motifs allowed conclusions to whether the identified gene is partial or full-length. In comparison to previous efforts (Andolfo et al. 2013; Xiaoxun et al. 2013), the RenSeq approach established 105 and 126 additional NB-LRRs within the Heinz 1706 and the LA1589 genome. About 70% (221) of all Heinz 1706 NB-LRR genes are potentially full length, while in S. pimpinellifolium LA1589 only 37% (124) of the total NB-LRR complement (Table 1 and Table 2) encodes the minimal domain structure (NB-ARC and LRR) necessary for a full length gene. Positional information of the motifs that are either associated with an N-terminal domain or the beginning of the NB-ARC were further used to predict the putative start codon, and the last LRR specific motif and reading frame information to establish the stop codon for potentially full length sequences (Table 2).

	Protein domains	***S. pimpinellifolium* LA1589**	***S. lycopersicum* Heinz 1706**
Full-length	CC-NB-LRR	110	195
	TIR-NB-LRR	14	26
Total full-length		124	221
Partial	CC-NB	33	14
	TIR-LRR	1	1
	TIR-NB	7	3
	NB	122	57
	TIR	12	10
	LRR	56	20
Total partial		231 (124)*	102
Total		355	326

Table 2. Numbers of S. pimpinellifolium LA1589 and S. lycopersicum Heinz 1706 genes that encode domains similar to plant R proteins as identified in this study. *Partial S. pimpinellifolium LA1589 NB-LRR genes were considered fragmented, and thus part of a full not yet combined gene, when they are located within 500 bp of the beginning or end of the contig.

3.3.3. Correction of NB-LRR gene models in HEINZ 1706

Our results identified 72 mis-annotated NB-LRR sequences compared to a previous study (Andolfo et al. 2013) in which an semi-automated annotation was used (Table 1). To fully reannotate the NB-LRR complement, we manually analysed all identified loci to correct erroneous start and stop codons, missing or additional exons, as well as erroneously fused or split genes . In Figure 1A and 1B we present two examples of genes that were corrected using RenSeq data. Although the tomato genome is of high quality it still contains a number of regions with unknown sequence content, and among the annotated NB-LRR genes we found eight with stretches of N's of varying length (between 97 and 7,851 bp). This number is significantly smaller than the 39 gaps found in potato NB-LRR sequences (Jupe et al. 2013). These gaps were filled by creating arches of sequence reads from both sides using the long 250 bp RenSeq reads, and the corresponding paired end information. An example is shown in Figure 2, where four sequence gaps were identified (Gap1–Gap 4, Figure 2B in violet) within a gene cluster on chromosome 4 that originally comprised three partial and four full-length NB-LRR genes (Andolfo et al. 2013). Solyc04g008130 (CC-NB-LRR) had a gap at the expected stop codon position, which was then corrected. Two gaps were identified between the four partial NB-LRR genes Solyc04g008160, Solyc04g008170, Solyc04g008180 and Solyc04g008190, and closing of these enabled the reannotation of the partial genes into two full-size CC-NB-LRR genes (RDC0002NLR0020 and RCD0002NLR0021). Solyc04g008200 had a predicted gap of 784 nt in the middle of the sequence, that was corrected to 503 nucleotides. The RenSeq data further identified a novel NB-LRR in this cluster (RDC0002NLR0019, Figure 2B in red), and the final gene models are graphically depicted in Figure 2C.

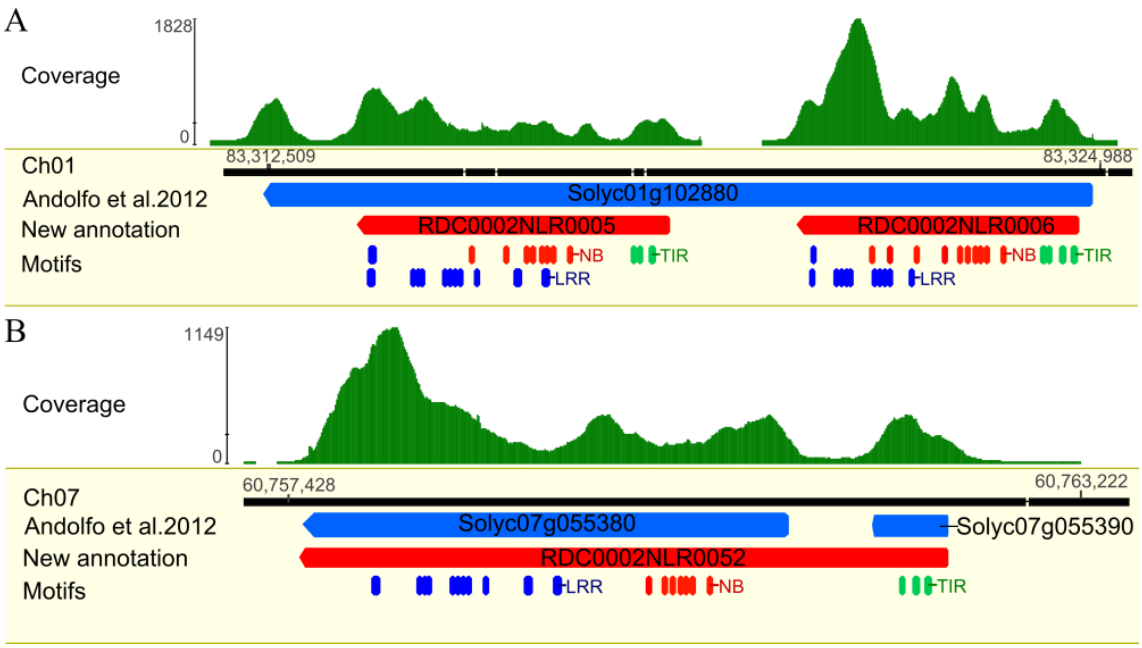

Figure 1. Reannotation of two erroneously fused NB-LRR genes. (A) Mapping of RenSeq reads identified two distinct patterns within Solyc01g102880, suggesting a fusion of two genes (green box); (B) In contrast, Solyc07g055380 and Solyc07g055390 are predicted individual genes (red box), however a gap-free RenSeq read coverage pattern suggested that both are part of one longer sequence. The corrected annotation was confirmed with the NB-LRR gene specific MAST motifs (TIR, NB and LRR motifs are shown in green, red and blue boxes, respectively) and are depicted as boxed arrows (green) for the novel TIR-NB-LRR full-length genes RDC0002NLR0005, RDC0002NLR0006 and RDC0002NLR0052.

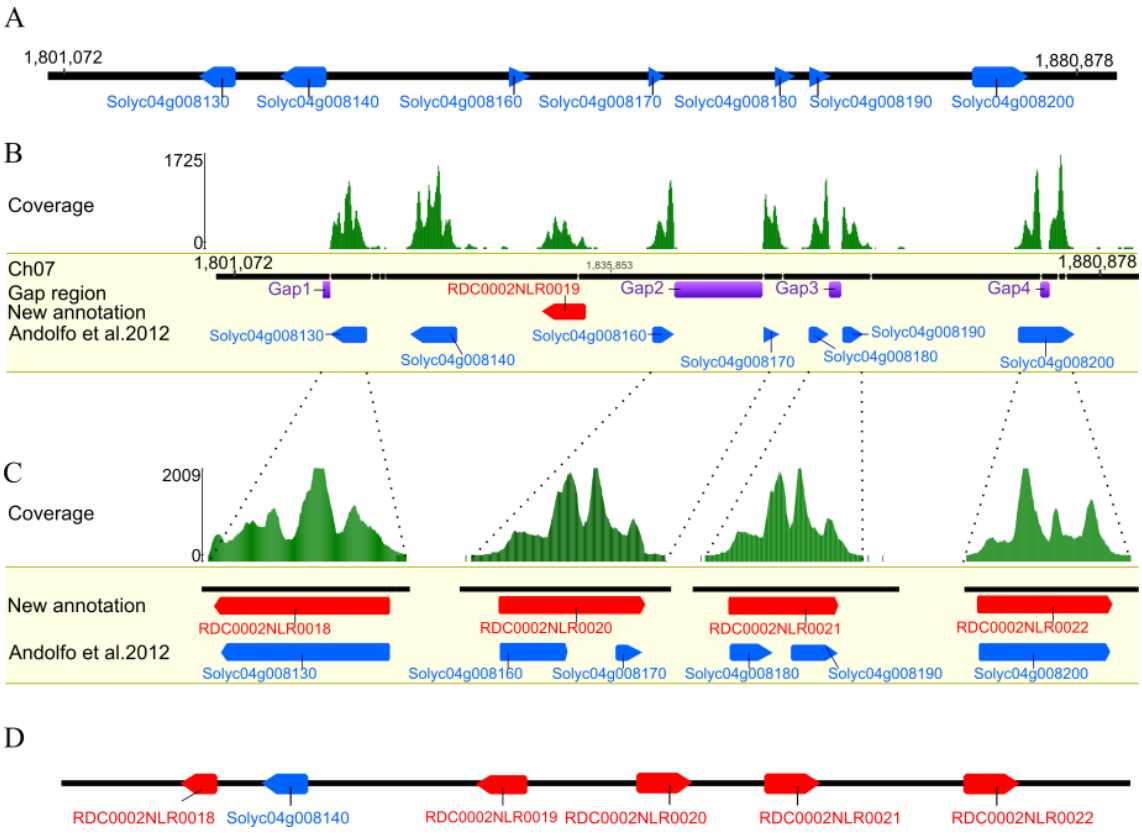

Figure 2. Detailed analysis of a NB-LRR cluster between positions 1.81 - 1.87 Mb on chromosome 4. (A) The Heinz 1706 region with annotations from Andolfo et al 2012. NB-LRR genes are depicted as blue boxes. (B) MiSeq RenSeq read coverage is shown with green peaks and identifies one yet unannotated NB-LRR RDC0002NLR0019 (red box). The purple boxes indicate stretches of N's as unknown genomic sequences (Gap1 to Gap4, in violet). (C) Close-up of the analysed loci in which gaps were closed. Previous gene models (blue boxes), novel models (red boxes) and RenSeq read coverage (green peaks) are shown. (D) Representation of the reannotated NB-LRR gene cluster.

3.3.4. Conservation of the NB-LRR distribution between tomato and potato

The genome-wide distribution of NB-LRR genes, based on the chromosome size, was significantly non-random (χ^2=96, P <0.001) (Figure 3). The greatest numbers of NB-LRR genes are found on chromosomes 4, 5 and 11 (about 45% of the mapped genes), with the smallest number on chromosome 3 (9 genes), which is consistent with other Solanaceae including potato (Jupe et al. 2013). There was a clear difference between the genome distribution of the TNL and CNL genes, and the largest number of TNLs (43%) was found on chromosome 1, while TNLs are absent on chromosomes 3, 6 and 10. CNLs are however present on all chromosomes. The majority (about 66%) of the NB-LRR genes in tomato are organized in clusters (a region that contains four or more genes within 200 kb or less; (Andolfo et al. 2013), including tandem arrays). We found 20 gene clusters that in total carry 107 NB-LRR genes, with on average five, and a maximum of 14 NB-LRR-encoding genes. The largest cluster was located on the short arm of chromosome 4 (Solyc04g009070 to Solyc04g009290) and resides in a ~110-kb-wide region.

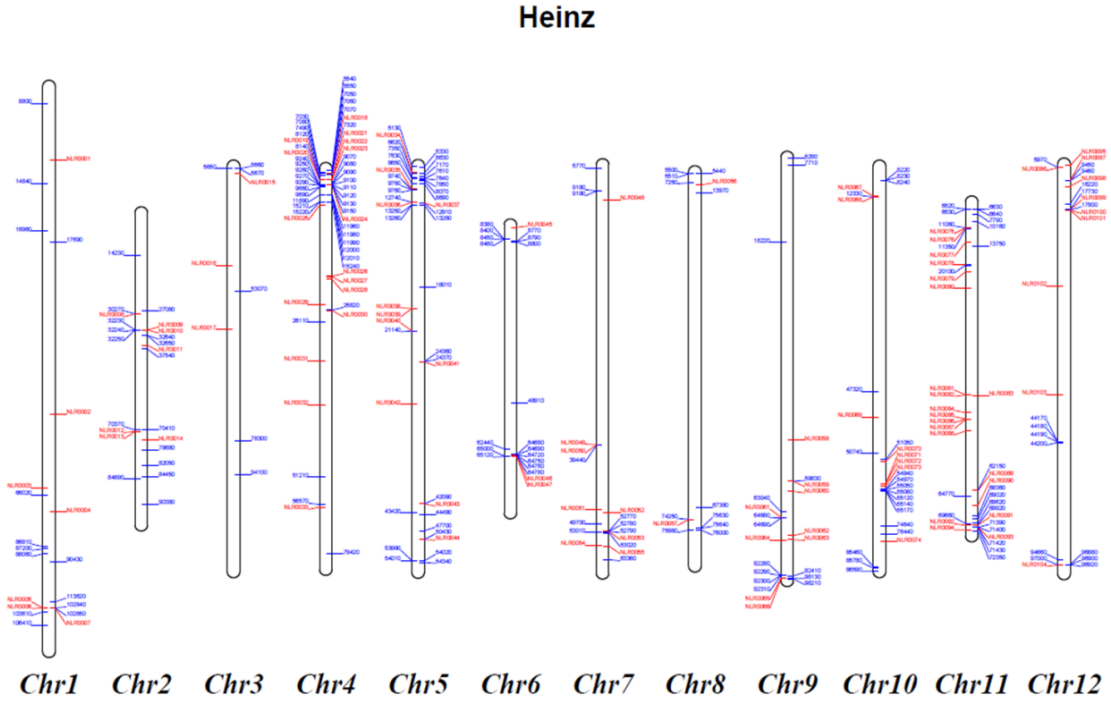

Figure 3. Chromosomal distribution of Heinz 1706 NB-LRR genes. The previously annotated NB-LRR genes (Andolfo et al. 2013) are shown in black and those discovered in this study are blue. Genes depicted to the left of the chromosome are on the forward strand and those on the right are on the reverse strand.

3.3.5. Phylogenetic relationships between tomato NB-LRR genes

The NB-ARC domain of NB-LRR genes has proven to be the most reliable protein domain with which to analyse phylogenetic relationships. Therefore the amino acid sequence of this domain was extracted from each NB-LRR gene with a full NB-ARC domain and used to perform a phylogenetic analysis for Heinz 1706 (Figures 4) and LA1589 separately). For comparative purposes, we included 30 well-characterized cloned reference R genes from eleven different plant species and two out-group genes with a nucleotide-binding domain, the human Apaf1.1 and nematode Ced-4, respectively (green in Figure 4 and 5). A total of 240 and 222 NB-ARC domains of Heinz 1706 and LA1589 were aligned, respectively. The sequences were grouped into robust clades supported by bootstrap values ? 75%, and allowed the definition of 17 and 16 clades that have high sequence similarities in Heinz 1706 (Figure 4) and LA1589, respectively.

The phylogenetic tree presents a clear distinction between TNL, CNLRPW8 and CNLEDVID genes (Figures 4), as reported earlier for potato, and we also found this distinction to be very clear in Arabidopsis (Mayers et al. 2005; Jupe et al. 2012; Collier et al.

2011]. Within the CNL genes, 17 clades were defined in Heinz 1706 and 15 clades in LA1589 (Figures 4;). Clade CNL-1 comprises Mi1.2, Rpi-blb2 and similar sequences on chromosomes 5 and 6. It is interesting to note that clade CNL-1 shares a common ancestor with clades CNL-2 and CNL-3 (supported by 93% bootstrap indexes), which comprise members of the Hero family encoded on chromosome 4 and the Sw5 family on chromosome 9, respectively. Within the LA1589 phylogenetic tree these first three similar clades (CNL-1, CNL-18 and CNL-3) are less well defined, and Hero has only two similar sequences (RDC0003NLR0189 and RDC0003NLR0120) that were not considered a clade. Differences like these are likely due to the poor quality of the LA1589 genome assembly and the fragmented nature of genes annotated from this. CNL-4 shares in both phylogenetic trees similarities with R1 and Prf, and all sequences are located on chromosome 5. Two small clades present in Heinz 1706 and LA1589 are CNL-6 and CNL-7 that share similarity to the characterized genes Rx, Rx2 and Gpa2, and Bs2, respectively. Five individual large clades (CNL-5, CNL-8, CNL-9, CNL-13 and CNL-18) do not have similarity to any functional R gene, and might thus be potential sources of novel R genes. Clade CNL-10 includes the reference protein Tm2 and highly similar sequences encoded on chromosome 9 in both species. 14 and 10 genes similar to the A. thaliana RPP13 were clustered in Heinz 1706 and LA1589, respectively, and can be found in clade CNL-11. Unique to tomato is CNL-12, which includes sequences similar to RPM1. CNL-13 harbours seven and eight genes from Heinz 1706 and LA1589, respectively. The small clade CNL-14 includes homologs of Rpi-blb1 with high homology in both phylogenetic trees. Nine and 13 homologues of the very similar tomato I2 and potato R3a genes are found in clade CNL-15 of Heinz 1706 and LA1589, respectively. Clade CNL-16 is located on an ancestral position between TNL and CNL genes, and harbours the characterized genes PRS2 and RGC2B (Bent et al. 1994; Shen et al. 2002).

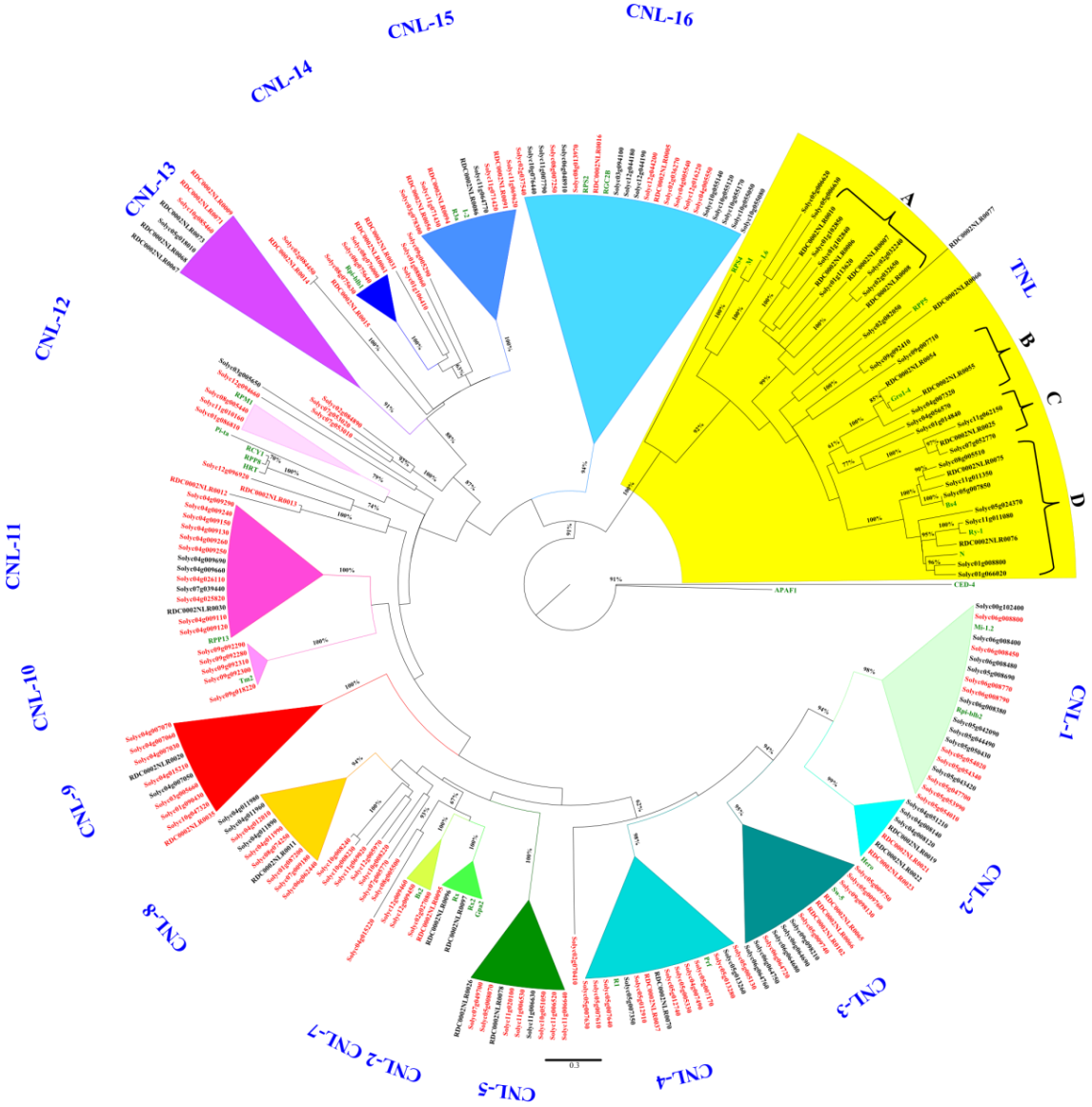

Figure 4. Phylogenetic analysis of the reannotated Heinz 1706 NB-LRR genes. Full NB-ARC domains of 240 reannotated NB-LRR genes were used together with 30 functionally characterized plant R genes (green font) to do a maximum likelihood analysis based on the Whelan and Goldman model. Clades are collapsed based on a bootstrap value over 79 and numerated. The TNL clade is drawn with a yellow background. Expressed genes, as identified by the cDNA RenSeq analysis, are in red font. Evolutionary analyses were conducted in MEGA5. Labels show the gene IDs (red for expressed NB-LRR genes; black for not-expressed genes). Bootstrap values higher than 79 (out of 100), are indicated above the branches. The tree is drawn to scale, with branch lengths proportional to the number of substitutions per site

3.3.6. cDNA RenSeq significantly reduces the complexity of the NB-LRR gene complement

RenSeq was established as a tool to conduct targeted sequencing of the NB-LRR gene complement in order to identify polymorphisms that are linked to disease resistance between resistant and susceptible individuals of a segregating population (Jupe et al. 2013).We tested whether the ability to enrich NB-LRR sequences 500-1000× using RenSeq could provide enough read depth to sequence cDNA of these low-expressed genes. A RenSeq experiment was carried out on double-stranded cDNA from pooled RNA samples of untreated and late blight (*Phytophthora infestans*)-infected Heinz 1706 and LA1589 leaves.

In total 2,882,986 paired-end 250-bp MiSeq reads were recovered from NB-LRR enriched Heinz 1706 cDNA; 65% (1,863,598 reads) of which map to the 12 reference chromosomes. Reads not mapping to the chromosomes, were identified to originate from ribosomal RNA. High-stringency Bowtie mapping, omitting reads that would map to more than one sequence (see Material and Methods), placed 214,050 and 235,656 reads onto 167 Heinz 1706 and 154 LA1589 NB-LRR genes, respectively. On average 1281 and 1560 reads mapped per NB-LRR sequence (Table 4). Several sequences had very low number of mapping reads (minimum of 2) and might be mapping artefacts, but were still considered. Overall, the complexity of the NB-LRR complement was reduced by 51% in Heinz 1706 (Figure 4), and 43% in LA1589. More importantly, this reduction was even over all phylogenetic clades.

3.4 DISCUSSION

3.4. Discussion

The availability of high-quality genome sequences from tomato can help to address functional studies and comparative analysis. Importantly, given the accumulated knowledge of disease resistance in this important crop species, whole-genome analysis of the R gene complement will facilitate rapid advancement of disease resistance breeding. The tomato R gene annotation were solely based on the existing protein predictions of the Tomato Genome Sequencing (TGC 2012).

Automated gene prediction software does not annotate all gene models correctly, and the efforts of genome sequencing consortia do generally not include the detailed verification of individual genes and gene families (Andolfo et al. 2013). Resistance gene enrichment sequencing (RenSeq) was recently introduced as an efficient method to reduce the genome complexity for the refined annotation of the NB-LRR gene family (Jupe et al. 2013). RenSeq facilitates deep sequencing and identification of the complete NB-LRR gene complement in tomato genome. The Illumina MiSeq platform with 250-bp reads reduced error-free closing of gaps in the assembly. In comparison to Jupe et al. (2013) who relied on 76 bp paired read data, we used longer reads that allowed a very rapid closure of the gaps with high confidence, using minimum numbers of reiterative mapping rounds.

Through the use of this technology we were able to identify 105 new NB-LRR genes in Heinz 1706. In LA1589 it was highlighted that only 124 NB-LRR complement encodes the minimal NB-ARC and LRR domain structure necessary for a full length gene. This is unlikely to reflect the true structure and might be due to the fragmented nature of the LA1589 genome, since about 35% (124) of the partial genes are fragments found at the border of contigs, whose missing counterparts are anticipated to lie on other contigs. We anticipate that carrying out RenSeq on other assembled plant genomes would increase the number of annotated NB-LRR sequences and will enable more targeted and specific resistance breeding strategies. It is intriguing that tomato has less than half of the number of NB-LRR genes compared to the doubled-monoploid reference potato. However, those present are found in syntenic chromosomal clusters between both species. Overall, the difference is not due to absence of gene sub-families, but due to a significantly smaller number of single genes within these clusters in tomato. Whole-genome duplication events did not contribute to the expansion in potato (TGC 2012).

The most reliable protein domain with which to analyse phylogenetic relationships have been proven to be the NB-ARC domain of NB-LRR genes. The phylogenetic tree presents a clear distinction between TNL, CNL_{RPW8} and CNL_{EDVID} genes as reported earlier for potato, and found in *Arabidopsis* (Meyers et al. 2005; Jupe et al. 2012; Collier et al. 2011). Distinct from the canonical CNL genes are those with a CC-domain similar to RPW8, that are suggested to have conserved functions and can be found throughout the plant kingdom (Collier et al. 2011). The ancient position in the phylogenetic trees of tomato, potato and Arabidopsis, as well as other reports suggest that this group was present prior to the monocot/dicot split (Collier et al. 2011). Well-characterized members of this clade are N-required gene 1 (NRG1) from *N. benthamiana*, and the *Arabidopsis* Activated Disease Resistance 1 (ADR1) gene family. It is interesting to note that clade CNL-1 shares a common ancestor with clades CNL-2 and CNL-3 (supported by 93% bootstrap indexes), which comprise members of the *Hero* family encoded on chromosome 4 and the *Sw5* family on chromosome 9, respectively. Within the LA1589 phylogenetic tree these first three similar clades (CNL-1, CNL-18 and CNL-3) are less well defined, and Hero has only two similar sequences (RDC0003NLR0189 and RDC0003NLR0120) that were not considered a clade. Differences like these are likely due to the poor quality of the LA1589 genome assembly and the fragmented nature of genes annotated.

While RenSeq on bulked resistant and bulked susceptible plants allows the identification of NB-LRR gene alleles that cosegregate with a resistance phenotype using “quick”-mapping or genotype-specific mapping, the list of candidate genes can further be reduced by cDNA RenSeq that limits the number of R gene candidates expressed. For some gene families, however, it is still challenging to define the many paralogous NB-LRR genes within chromosomal clusters and phylogenetic clades, and to identify the individual paralogue from which a co-segregating SNP derives. NB-LRR genes are not highly expressed, probably to prevent auto-immunity, and thus RNA-seq approaches would be unlikely to recover enough sequence depth. These data however do not allow any conclusions about a correlation between read number and expression level, as a certain bias from the bait-library cannot be excluded (though was not seen after RenSeq on gDNA). Of the expressed genes, 90% are full length and 10% are partial genes. The high number of expressed partial genes is consistent with the hypothesis that these could function as a decoy for miRNA, and are thus important in NB-LRR gene regulation (Li et al. 2012). A combination of these methods will greatly accelerate the recruitment of natural resistance gene biodiversity for crop improvement.

4. Genome-wide identification and analysis of candidate genes for disease resistance in tomato

4.1. Introduction

Currently, there is tremendous interest in the advanced use of genome-wide data for identifying new resistance genes. In order to speed up R loci tagging and pathogen recognition gene identification, several strategies have been explored (Pan et al. 2000; Riely and Martin 2001; Caicedo and Schaal 2004; Mazourek et al. 2009; McHale et al. 2006). Genomic approaches can enhance the identification of genes that encode for resistance traits. After a genomic interval underlying a disease resistance trait has been identified, there are various possibilities for tracking down the gene responsible. Traditional approaches can be extremely costly, tedious, and time-intensive, given the difficulty of marker development and the size and complexity of R gene clusters (McDowell and Simon 2008). Annotation data and genetic map information represent an invaluable resource for performing this task. A better understanding of tomato R gene genomic architecture could streamline cloning efforts. A dataset of pathogen recognition gene tomato dataset was categorized according to the presence and order of protein domains, phylogenetic analysis and physical arrangement within the genome (Andolfo et al. 2013). In this study, we identified strong pathogen recognition gene candidates linking predicted pathogen recognition proteins with previously mapped R loci, characterized in detail the identified pathogen recognition genes, highlighting peculiar pathogen recognition domain arrangements, and finally provided molecular validation of our predictions, both exploring the tomato transcriptome and performing experimental validation. Puzzling information were collected and combined in order to obtain a synergy between different approaches. Our strategy was constructed to reduce the time required for R gene identification and to make easier their cloning, a critical step towards modern genome breeding. In many cases, a predicted protein was narrowed down to a small region, allowing the identification of one or few candidates, now available for exploiting their specific function. Our attempt was conducted in order to capture fundamental aspects of data integration contributing to pinpointing key steps in genetic, genomic, and phenotypic data synthesis for a better R gene isolation.

4.2. Materials and methods

4.2.1. Physical location

To construct the physical maps, the predicted pathogen recognition genes were collected in an SQL database and catalogued with the information on their characteristics and their location. A custom PERL script, connecting the database and transforming the information into vector graphics (SVG) images, was written to design each chromosome. The sequences of the 82 chromosome markers linked with R genes not yet cloned and reported by Foolad (2007) were taken from the SGN database (Supplemental Table S1).

4.2.2. Phylogenetic and gene duplication analysis

Evolutionary analyses were conducted using MEGA5 (Tamura et al. 2011). The phylogenetic relationships of predicted pathogen recognition proteins were inferred separately (e.g., NBS, eLRR-Ser/Thr and KIN groups) using the maximum likelihood method based on the WAGmodel (Whelan and Goldman 2001). The bootstrap consensus tree, inferred from 100 replicates, was taken to represent the evolutionary history of the sequences analyzed (Felsenstein 1985). All the amino acid sequences were aligned using MUSCLE 3.6 (Edgar 2004). For nucleotide sequences, the General Time Reversible Model was used. The bootstrap consensus tree inferred from 100 replicates was taken to represent the evolutionary history of the sequences analysed (Felsenstein 1985). All the sequences were aligned using ClustalW 1.74 (Thompson et al. 1994).

4.2.3. Validation of prediction results

The expression data of S. lycopersicum variety HEINZ 1706 were obtained from the Tomato Genome Consortium (2012) and used to extract a pathogen recognition gene expressed dataset. A pool of 37 predicted R genes was used to perform molecular validation. DNA and RNA were extracted from leaf tissue of genotype S. lycopersicum variety HEINZ 1706 using DNeasy and RNeasy Plant mini kits (Qiagen, Valencia, CA, USA), respectively. PCR was executed with 25 ng of genomic or complementary DNA, 10 pmol primers, 1 U of Taq DNA polymerase Kit (Invitrogen, Carlsbad, CA, USA), 10 pmol dNTPs, and 2 mM MgCl2 in 25 ll reaction volumes. Amplification was performed using the following cycling conditions: 1 min at 94 °C, followed by 30 cycles of 1 min at 94 °C, 1 min 30 s at 60 °C and 2 min at 72 °C, with a final extension for 7 min at 72 °C. Amplicons were separated by electrophoresis on agarose gel (1.5 %), and photographed by a GelDoc apparatus. Primers were designed with Primer3 (http:// frodo.wi.mit.edu), with a length between 18 and 27 bp. The length of the amplified fragments ranged from 300 to 1,000 bp, and the Tm of the

specific primers was 59 °C for all pairs of primers (Online Resource S4). Amplicons were sequenced using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and run on automated DNA sequencers (ABI PRISM 3100 DNA Sequencer, Applied Biosystems). Sequence data deriving from SL2.40 reference were aligned with corresponding sequences originated from amplicons using MUSCLE 3.6 (Edgar 2004).

4.3. Results

4.3.1. Constructing the physical map of tomato R genes

In tomato, several resistance phenotypes have been genetically mapped, delineating genome regions harboring causative resistance loci. Examples include the Py1 gene for resistance to corky root rot (Doganlar et al. 2002) on chromosome 3, the Pto gene conferring resistance to Pseudomonas syringae resistance (Martin et al. 1991) on chromosome 5, the root-knot nematode resistance locus Mi (Kaloshian et al. 1998), the Ty1 gene for tomato yellow leaf curl virus resistance (Hanson et al. 2000) on chromosome 6, and the Sw5 gene for tomato spotted wilt virus resistance (Stevens et al. 1995) on chromosome 9. By performing a literature search we selected a pool of 82 markers flanking the target loci (R genes and quantitative trait loci). The markers used were chosen on the basis of a very important factor, namely, their close association with genes not yet cloned. Once the markers were placed on chromosomes, we were able to select putative pathogen recognition genes, which permitted us to discriminate those which fell between at least one pair of markers, allowing us to focus on a restricted set of genes. Out of 769 predicted pathogen recognition protein sequences, about 368 corresponding genes (48 %) were localized among markers linked with functionally defined and mapped, but uncloned, R genes. The functional R gene map based on the colocalization of pathogen recognition putative proteins with R loci linked markers is shown in figure 1. The map shows that the 368 candidate pathogen recognition genes are distributed on 12 chromosomes. Moreover, in some cases, the chromosomal regions bounded by markers of different R loci overlapped, delimiting a common chromosome area, as in the case of markers Lb3, EB7 and Xv-4.

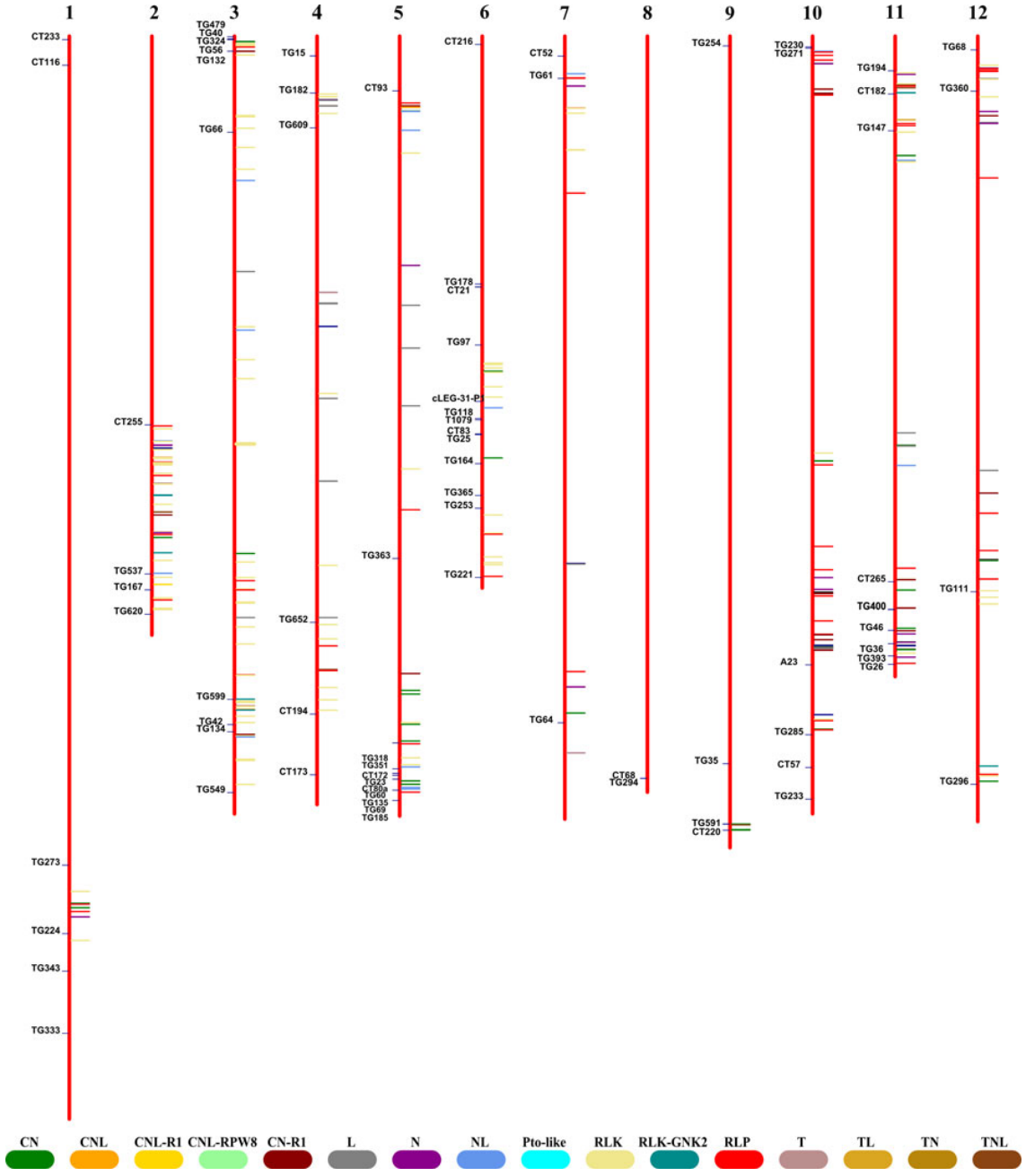

Figure 1 Overview of predicted pathogen recognition genes localized among markers linked with functionally established R loci. The color used for each gene indicates the structural class to which the encoded protein has been assigned. RLP receptorlike protein, RLK receptor-like kinase, CNL (coiled coil/ nucleotide-binding site/leucine-rich repeat) protein, TNL Toll interleukin resistance/nucleotide-binding site/leucine-rich repeat, Gnk2 ginkbilobin-2, RPW8 domain was identified in two proteins isolated in A. thaliana that confer resistance against a broad range of powdery mildew races; R1 domain characteristic of the R1 protein, PTO-like genes encoding the typical serine threonine domain characteristic of the Pto protein.

4.3.2. Characterization of putative tomato pathogen recognition genes

In order to better classify and label tomato proteins in the functional map, phylogenetic investigations were conducted. Three separate phylogenetic trees were produced, designated the Nucleotide-Binding Site (NBS) group, extracellular Leucine-Rich Repeat-Serine/ Threonine (eLRR-Ser/Thr) group, and Kinase (KIN) group. These analyses were useful for obtaining additional information on chromosome regions under study for a specific resistance trait. We defined as a genomic region specific sub-clade a phylogenetic clade containing at least five sequences situated on the same genomic region with bootstrap support greater than 70 %. The NBS phylogenetic tree, containing 86 tomato predicted proteins and 23 reference proteins involved in the resistance process, allowed us to identify orthologs to functional proteins and to detect three interesting genomic region sub-clades. Proteins grouped in the region-specific sub-clade on chromosome 11 showed an average identity of 56 % and those on chromosome 5 an average identity of 32 and 67 %. Interestingly, the Toll-Interleukin-Resistance/Nucleotide-Binding Site (TIR-NBS) clade included a protein with a Resistance to Powdery Mildew (RPW8) domain at the N terminal. This domain was identified in two proteins isolated in Arabidopsis thaliana that confer resistance against a broad range of powdery mildew races (Xiao et al. 2001). The association of the RPW8 domain with NBS-LRR domains could help to shed light on the mechanism of action of both RPW8 and genes of similar architecture in the Solanum species. The eLRR-Ser/Thr group comprised 83 sequences and 15 reference proteins, including receptor-like proteins (RLP), involved in defense as well as in the development process. Phylogenetic analysis highlighted sub-clades that identified specific chromosomal regions with potential candidate genes for resistance. In particular, two sub-clades that included proteins located on chromosomes 12 and 7 were identified. The KIN group contained 143 predicted sequences and 12 reference proteins. The relative phylogenetic tree revealed two interesting super-clades (Pto and Gnk2 superclades). On looking at the genome distribution of pathogen recognition genes linked to R markers, 224 genes were identified (about 60 %) that reside either in a gene cluster or in an array of 2–3 genes. Of these, some resistance loci, inherited in Mendelian fashion, were analyzed in greater detail to perform initial screening of potential resistance genes. Table 1 reports seven loci containing potential pathogen recognition genes. The markers linked to Xv4 discriminated a region on chromosome 3 that comprised nine genes, including a receptor-like kinase (RLK) protein with an extracellular Gnk2 domain. On chromosome 6 the markers of Ol1 identified a single coiled coil/nucleotide-binding site/leucine-rich repeat (CNL) gene that has a peculiar domain

(PTHR23155: SF94-Panther). Eight genes were located on chromosome 6 between the markers of Ty1, and a Pto-like gene was also included between markers of Ty3. On chromosome 9 the markers linked to Ph3 included three CNL and one Toll interleukin resistance/nucleotide- binding site/leucine-rich repeat (TNL) protein. The TNL sequence showed a tyrosine-protein kinase active site (IPR008266-Prosite) corresponding to the LRR domain. The markers linked to the Ty2 locus allowed 14 genes to be discriminated, located on chromosome 11, including genes belonging to cluster I2. Sixteen genes were located on chromosome 12 between the markers of Lv. This analysis allowed us to fine-tune the search for candidate genes.

Locus name	Pathogen	Reference	No. candidate genes/protein class				
			CNL[a]	RLK	RLP	TNL	Unknown
Lv	*Leveillula taurica*	Foolad (2007)	1	2	11	–	2
Ty2	Tomato yellow leaf curl virus	Foolad (2007)	3	1	1	–	8
Ph3	*Phytophthora infestans*	Foolad (2007)	3	–	–	1	–
Ol1	*Oidium lycopersicum*	Foolad (2007)	1	–	–	–	–
Ty3	Tomato yellow leaf curl virus	Foolad (2007)	–	–	–	–	1
Ty1	Tomato yellow leaf curl virus	Foolad (2007)	1	6	–	–	1
Xv4	*Xanthomonas campestris pv. vesicatoria* Race T3	Foolad (2007)	–	7	1	–	1

Table 1 Selected resistance loci inherited in Mendelian fashion harbouring predicted pathogen recognition genes, showing the locus name, the pathogen to which it gives resistance and the number of candidate genes identified subdivided by class. [a] CNL coiled coil/nucleotide-binding site/leucine-rich repeat, RLK receptor-like kinase, RLP receptor-like protein, TNL toll interleukin resistance/nucleotide-binding site/leucine-rich repeat, Unknown genes that encode novel domain associations or single domains.

4.3.3. Putative pathogen recognition gene functionality tests

A first prerequisite for testing the functionality of a gene is the identification of its transcript. In order to ascertain that the predicted genes derive from functional sequences, we categorized the expressed tomato predicted pathogen recognition sequences close to R loci. Table 2 reports the number of expressed predicted pathogen recognition genes co-localizing with R loci for each chromosome and the number of expressed genes falling in a cluster or array. On average, 80 % of the genes examined proved to have a transcript in the tomato genome ranging from 63 % (chromosome 10) to 100 % (chromosomes 3, 6 and 9). Of them, 197 genes are located in clusters or arrays that might have a resistance function. To verify that the predicted genes were actually present in tomato molecular analysis was carried out. Out of 37 gene sequences tested, 34 were shown to be also transcribed.

Chromosome	Predicted pathogen receptor genes (no.)	Expressed pathogen receptor genes (no.)	Identified cluster or array (no.)	Expressed gene in cluster or array (no.)[a]
1	7	6	1	2
2	54	54	8	29
3	55	55	11	28
4	31	25	8	20
5	38	32	6	19
6	16	16	4	8
7	25	22	7	21
9	4	4	1	4
10	46	28	11	15
11	50	43	13	35
12	40	39	8	25

Table 2 Results on predicted pathogen recognition genes co-localizing with R loci for each chromosome, the number of total and expressed pathogen recognition genes, as well as the number of expressed genes falling in the clusters or arrays, is shown. [a] The number reported is calculated from the number of total genes identified in cluster/arrays for each chromosome.

4.3.4. Identification of candidate genes in specific loci

In this paragraph we will present two interesting examples of how the integration of genomic and genetic knowledge can facilitate the identification of good candidates for resistance functions. The following cases described of a recently mapped but not yet cloned R gene, and another locus under high evolutionary pressure for which no R gene in tomato has been identified yet.

Two recent publications presented independently a set of four flanking markers for the R gene Ph-3 that confers resistance to certain *P. infestans* isolates in S. lycopersicum (Andolfo et al. 2013; Zhang et al. 2013). Alignment based anchoring of these marker sequences (Indel_3, CT220, TG591 and P55) to the reference chromosomes identifies a 600-kb region on the short arm of Chromosome 9 (Figure 2A). This genomic region includes sequences with high similarity to the tomato R genes Tm2 and Sw5, which confer resistance to Tomato mosaic virus (ToMV) and Tomato spotted wilt virus (TSWV), respectively. The Tm2 cluster in Heinz 1706 consists of four CC-NB-LRR genes that share over 90% pairwise identity and one unrelated TIR-NB-LRR gene. The Sw5 cluster is composed of three full length CC-NB-LRR and a partial CC-NB gene. Interestingly, the two independently identified marker pairs span a common region of only 30-kb, in which only one NB-LRR gene is located between TG591 and P55. The CNL Solyc09g092310 is the closest homologue in Heinz 1706 and is thus a potential candidate for Ph3 in the resistant tomato line. This CNL has an amino acid identity of 77.4% and 73% with Rpi-vnt1.1 and Tm2, respectively. Figure 2C shows the syntenic conservation of the R gene clusters around the Ph-3 candidate gene between tomato

and potato. A combined potato and tomato phylogenetic analysis of sequences found in this syntenic region did not result in a clear distinction of the sequences derived from both species, suggesting that these clusters were already present in the last common ancestor (Figure 2B). Five highly similar gene pairs with an identity between 82 and 89% (Figure 2C; blue arrows) were identified that might be most ancestral.

Chromosome 4 of Heinz 1706 harbours the largest NB-LRR gene cluster with 14 members (all located in CNL-11) (Figure 3A). All members of this cluster, jointly to Solyc04g009660 and Solyc04g009690 genes, share high sequence similarity to each other and the wild potato derived R genes R2, Rpi-blb3 and Rpi-abpt, that are located in a syntenic region of the potato chromosome 4 (Chunwongse et al. 1998; Li et al. 1998). Synteny is also shown by mapping the markers CT229 and TG339R that are linked to Rpi-blb3 (Chunwongse et al. 1998). A detailed phylogenetic analysis of proteins encoded by members of these clusters from tomato and potato show that all genes fall into a unique clade with mean identities of 80% and a bootstrap value of 83% (Figure 3B). Solyc04g009290 has high sequence identity to R2 (88%; Figure 3A). The phylogenetic tree further identifies nine duplication events in potato that must have occurred after the divergence of potato and tomato (Figure 3C). Microsyntenic analyses identified six NB-LRR genes with high sequence similarity between 78 and 85% in both species (blue arrows; Figure 3C). No functional R gene has yet been identified in tomato from this rapidly evolving cluster, but it can be speculated that some alleles of this locus might encode valuable resistance.

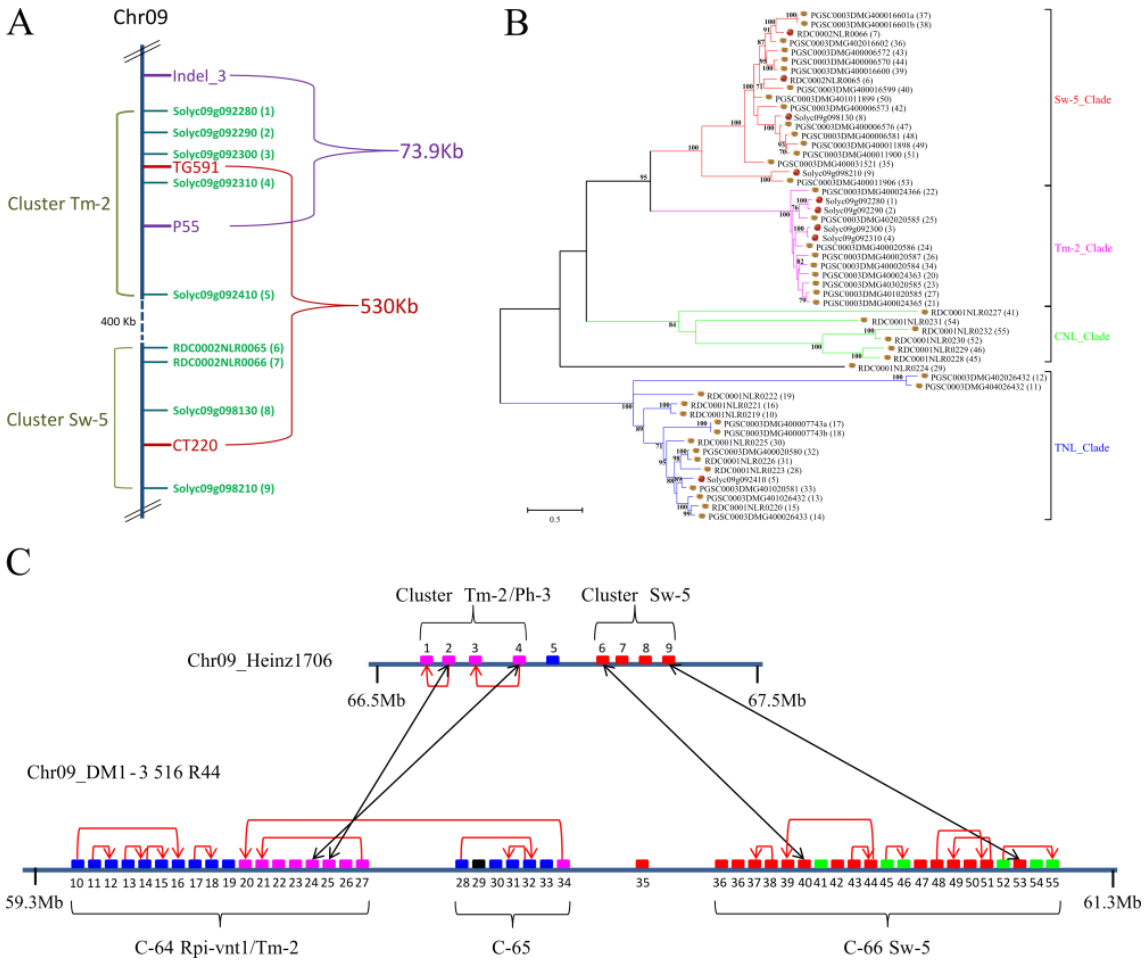

Figure 2. Comparison of the Tm-2 and Sw-5 clusters between Solanum lycopersicum Heinz 1706 and S. tuberosum clone DM and identification of the Ph-3 locus in the tomato genome. (A) Physical mapping position of NB-LRR gene clusters close to the physical Ph-3 locus, based on marker information derived from [15, 16]. (B) Phylogenetic analysis performed using the maximum likelihood method, based on the general time reversible model, for homologous sequences of the Tm-2 and Sw-5 clusters. Cartoon potatoes and tomatoes at the end of the branches indicate the origin of the sequence. Bootstrap values (100 replicates) are indicated above branches. (C) Schematic representation of hypothesised gene duplication events that occurred in the tomato and potato genomic region of Tm-2 and Sw-5 clusters. NB-LRR genes are depicted as boxes, and the colors relate to (B).

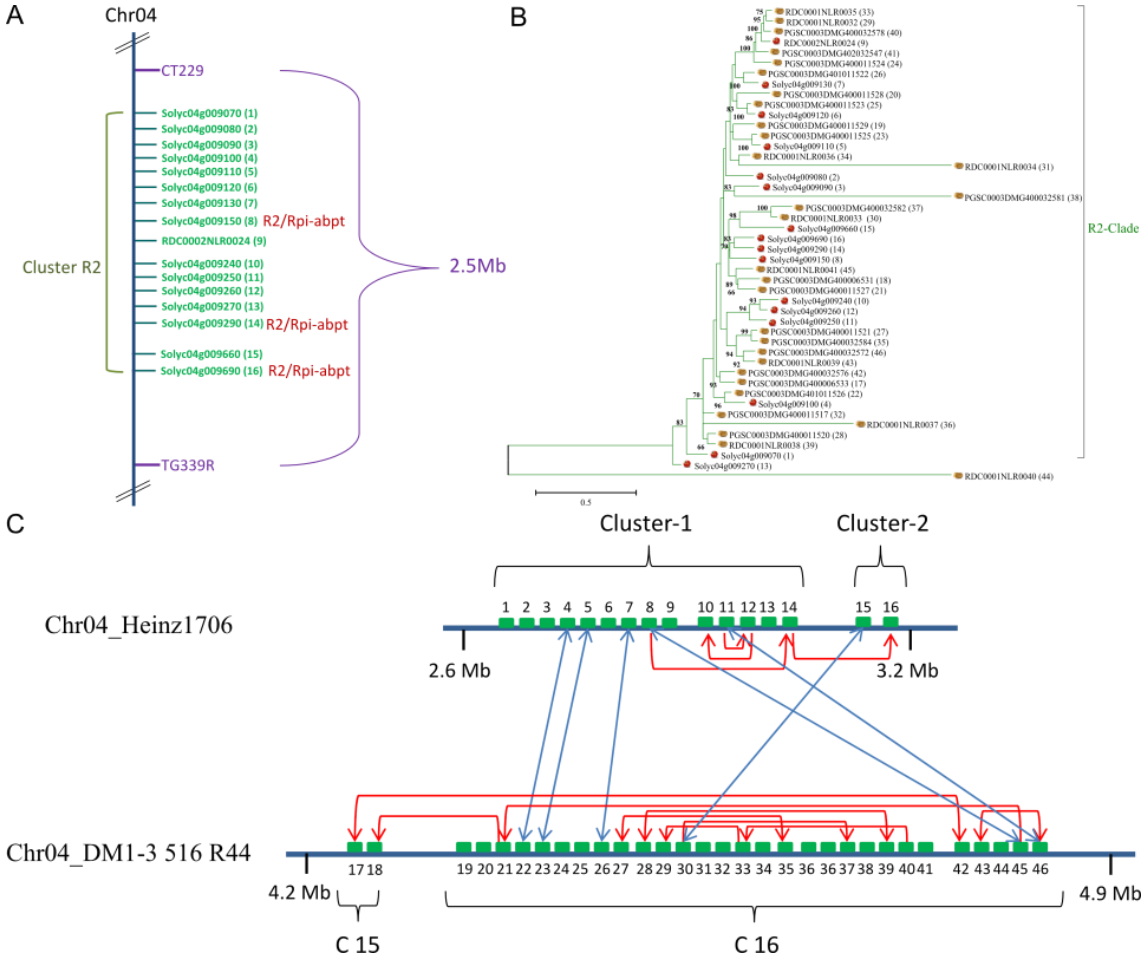

Figure 3. The evolutionary history of the largest NB-LRR gene cluster involving 16 NB-LRR genes on chromosome 4. (A) Physical mapping position of NB-LRR genes harbour in R-2 cluster. (B) The phylogenetic analysis was inferred using the maximum likelihood method based on the general time reversible model in MEGA5. Cartoon potatoes and tomatoes at the end of the branches indicate the origin of the sequence. Bootstrap values higher than 60 are indicated above branches. The tree is drawn to scale, with branch lengths measured in terms of the number of substitutions per site. (C) Schematization of the duplication events that occurred in these genomic regions. Arrows highlight the most probable gene duplication events. NB-LRR genes are depicted as boxes, and the colors relate to (B).

4.4. DISCSSION

4.4. Discussion

Tomato (Solanum lycopersicum L.) is subject to numerous pathogen attacks that can significantly reduce yields. Many tomato breeding projects have aimed to introduce resistance genes through classical and molecular genetic approaches (Foolad 2007). The immune response governed by resistance (R) genes has been investigated in depth in this species, contributing to elucidating important R gene molecular and genetic mechanisms in plants (Ercolano et al. 2012). Comprehensive knowledge of genomic R loci architecture in this species could help explain gene arrangement and diversification, as well as design a new genomic breeding strategy. The recent sequencing of the tomato genome (Tomato Genome Consortium2012) would appear very useful for improving the identification of disease resistance genes or genomic regions harboring them.

While some tomato disease resistance genes have been cloned using genetic map-based methods, many more have been mapped but not cloned to date (Foolad 2007; Ercolano et al. 2012). An integrated genomic approach could help to find a specific function for predicted genes. For this reason we built up a detailed R loci physical map, based on identifying predicted pathogen recognition genes located in the proximity of marker sequences associated with previously identified R loci (Andolfo et al. 2013). The visualization of information linked to a locus is a fundamental step in interpreting data and in suggesting correlations between genetic and genomic data, even if markers delimit regions that can include a variable number of genes (ranging from 1 to 50) which can belong to the same class of pathogen recognition genes or otherwise. Indeed, several markers co-localize with a large group of putative pathogen recognition genes, making identification of individual candidate genes difficult.

We investigated the pathogen recognition gene clades to reconstruct the evolution history of these loci. Importantly, phylogenetic profiling yielded a revised view of the organization of particular resistance gene subfamily. In particular the KIN-phylogenetic tree showed a first super-clade included proteins with a kinase domain similar to that of the protein Pto, while the second included proteins that possess a Ginkbilobin-2 (Gnk2) domain. The serine/threonine kinase Pto protein confers immunity to Pseudomonas syringae pv. tomato (Pedley and Martin 2003) and its overexpression has been shown to confer broad resistance (Tang et al. 1999). The Gnk2 is an antifungal protein found in the endosperm of Ginkgo seeds, which inhibits the growth of phytopathogenic fungi such as Fusarium oxysporum (Sawano et al. 2007; Miyakawa et al. 2009).

The molecular validation confirmed that there were neither inaccuracies of the predictor, nor prediction-distorted inaccuracies related to alignment defects. All the genes tested were found in the tomato genome. To date, all cloned tomato pathogen recognition genes with known resistance have been found to exist in gene clusters or arrays within the genome (Andolfo et al. 2013). The clustered arrangement of these genes may be a critical attribute allowing the generation of novel resistance specificities via recombination (Hulbert et al., 2001). These analyses also suggested mechanisms that led to lineage-specific diversification, providing insights into adaptive processes. The results presented in this paper can be useful for building up a framework of integration of genomics data, for breeding design, by using available marker data in conjunction with *R* gene annotation information.

5. A genome-wide transcriptional response to F. oxysporum and Tomato mosaic virus in tomato

5.1. Introduction

Tomato (Solanum lycopersicum) has served as an important model system for studying the genetics and molecular basis of resistance mechanisms in plants. The breadth of pathogen classes affecting tomato underscores the importance of tomato pathosystems as amenable models for studying the plant immune system (Ercolano et al. 2012). Plants build up-appropriate defence responses without draining energy resources to unsustainable levels through the cross-talk and fine-tuning of different defence pathways (Lodha and Basak 2012).The identification of host genes, involved in defence responses, is important both to understand plant resistance mechanisms against pathogens and constitute a starting point for building a biological model of a plant–pathogen interaction in order to direct future strategies of control.

Fusarium oxysporum f. sp. lycopersici (Fol) is one of the main diseases of tomato, which causes vascular wilt disease by colonizing the xylem vessels of roots and stems. Three physiological races (1, 2 and 3) have been identified. Resistant cultivars are the best wilt control strategy, since the pathogen remains in the soil for some decades and chemical control is ineffective (Reis et al. 2004). Tomato I2 is a CC-NB-ARC-LRR protein that confers resistance to race 2 of Fusarium oxysporum f. sp. lycopersici (Ori et al. 1997; Simons et al. 1998). Response mainly involves the callose deposition, the accumulation of phenolics and the formation of tyloses (outgrowths of xylem contact cells) and gels in the infected vessels (Beckman 2000). The interaction between tomato and Fusarium oxysporum f. sp.lycopersici has deeply investigated becoming a model system for disease resistance response (Takken and Rep 2010).

Tomato mosaic virus (ToMV) is a positive-sense ssRNA virus belonging to the Tobamovirus genus. ToMV infects tomato plants systemically, causing mosaic symptoms, which are characterized by intermingled light and dark green regions (He et al. 2012). The Tm2 gene of tomato and its allelic gene, Tm2(2), confer resistance to Tomato mosaic virus (ToMV) and encode a member of the coiled-coil/ nucleotide binding-ARC/leucine-rich repeat (LRR) protein class of plant resistance (R) genes (Kobayashi et al. 2011).

These two pathogens have different infection strategies because ToMV is a biothrophic foliar pathogen that causes tomato leaf mold, whereas Fol is a saprothrophic soilborne vascular pathogen that challenges several Solanaceae.

The aim of this study was to compare tomato global transcriptional profiles in response to host attack by ToMV and Fol in order to identify genomic differences and similarities in incompatible interactions between a foliar and a vascular pathogen. First we examined the global tomato transcriptional profiling during both incompatible interactions to compare transcriptional changes occurring in tomato plants. Then we highlighted GO categories enriched evidencing metabolic perturbation arisen. Finally, we explored genome arrangement of expressed genes along chromosome in order to connect genomic and transcriptional happening.

5.2. Materials and methods

5.2.1. Plant material and inoculation protocol

The tomato (S.lycopersicum) Rosso Delta variety, resistant to Fol race 1 (ex 2) and ToMV race 0 was used in our experiment. Tomato plants were inoculated with Fol race 1 (ex 2) strain ATCC 16605(Fol 1) (Plant Research International, NL). Plantlets were infected at the stage of expanded cotyledons: roots were cut and dipped in a suspension at a concentration of 1 x 106 conidia/ml, according to CPVO technical protocol TP/044/3 (Experiment I). The inoculation with Tomato Mosaic Virus was carried out with the strain GM6s of ToMV race 2a (Plant Research International, NL), according to the method reported in the CPVO technical protocol TP/044/3 (Experiment II). Plants with expanded cotyledons were inoculated with the sap obtained from infected desiccated tomato leaves. Viral transmission was ensured by mechanical tissues abrasion caused by diatomaceous powder.

5.2.2. Samples collection

Two 2 days after treatments, infected and non-infected tomato leave samples of two independently replicas of Experiment I and II were collected. Roots and leaves were removed from the plants, weighed and immediately frozen in liquid nitrogen and stored at -80°C. RNA was isolated from whole plants using Rneasy Plant Kit according to the manual instructions (Quiagen Valencia, USA). RNA samples concentration was determined using a Nanodrop photometer. Agilent 2100 Bioanalyzer was used for checking the quality of RNA.

5.2.3. Chip design and microarray hybridization

Transcriptomic analysis was performed using a 90K TomatArray 2.0 microarray synthesized using the CombiMatrix platform at the Plant Functional Genomics Center of the University of Verona. The chip (TomatoArray2.0) carries 25,789 nonredundant probes (23,282 unique probes and 2,507 probes with more than one target) randomly distributed in triplicate across the array, each comprising a 35-40-mer oligonucleotides. The source of sequence information included tentative consensus sequences (TCs) derived from the DFCI Tomato Gene Index Release 12.0 and expressed sequence tags. Eight bacterial oligonucleotide sequences provided by CombiMatrix, 8 probes designed on 8 Ambion spikes and 40 probes based on Bacillus anthracis, Haemophilus ducreyi and Alteromonas phage sequences were used as negative controls. Complete description of chip is available at the Gene Expression Omnibus under the platform accession GPL13934. Microarray analysis was used to investigate tomato gene expression profiles 2 days post the infection with Fol and ToMV comparing with the uninfected control profile. Total RNA (2 μg) was amplified

and labelled using the RNA ampULSe kit (Kreatech). After checking the quantity and quality of aRNA by spectrophotometry using NanoDrop™ 1000 (Thermo Scientific) and the quality subsequent labelling, 4 µg of labelled aRNA was hybridized to the array according to the manufacturer's recommendations. Cy5 labelled aRNA were hybridized with the microarray at 45°C for 16h; then arrays were washed with Hibr. Solution at 45°C for 5 minutes and 3xSS, 0,5 x SSPET, PBST wash and PBS wash at room temperature. Pre-hybridization, hybridization, washing and imaging stepswere carried out according to manufacturer instructions. Microarray were stripped for reuse (up to 3 reuse per chip) with CombiMatrix CustomArrayTM Stripping Kit according to the manufacturer instructions (product number 610049). After hybridization and washing, the microarray was scanned using a Perkin Elmer Scan Array 4000XL (software ScanArray Express Microarray Analysis System v4.0).

5.2.4. Data Analysis

Gene expression levels corresponding to 8 microarrays were processed using R software (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL http://www.R-project.org/.) and the limma package (Smyth, GK 2005). Raw data were investigated for quality assessment, preprocessed and normalized using a suitable subset of control probes available in the Combimatix array and a quantile normalization technique. Three technical replicates within each array and two biological replicates were employed to assess differential expression for each experiment (Experiment 1 and Experiment 2) to compare the different experimental conditions (non-inoculated vs inoculated). The significance of the differential expression was assessed taking into account the multiple testing setting and controlling the False Discovery Rate (FDR) at level FDR=0.01 and the emprical Bayes moderated t-statistics available in the limma package. The linear model fitting, used to assess differential expression of the two couples of contrasted experimental conditions (non inoculated vs inoculated), taking into account the presence of technical replicates as well as biological replicates.

5.2.5. Annotation of the gene chip probes

An in-house pipeline was developed to annotate tomato Tentative Consensus sequences (TCs) used to develop CombiMatrix CustomArrayTM probes. The queried tomato genes were identified by mapping TC sequences to the tomato CDS sequence using BLASTn (E-

value 1e-3). The latest version of the tomato gff3 annotation files was parsed to extract the cds sequences of genes probed. Blast2GO http://blast2go.bioinfo.cipf.es/ was used to provide automatic high-throughput annotation, gene ontology mapping and categorization of tomato proteins identify of TCs. An expectation value threshold of 1e-6 in BLASTp analysis was peformed. Blast2GO was used for the statistical analysis of GO term frequency differences. The enrichment analysis of the GO-terms was based on the Fisher's exact test and corrects for multiple testing. A cut-off FDR value of 0.05 was used. An interactive graph of the GO category enrichment REVIGO htt://revigo.org/ and a graphic representation Cytoscape htt://cytoscape.org// were also performed.

5.3. Results

5.3.1. Identification of differentially expressed genes induced by F. oxysporum f. sp. lycopersici (Fol) and ToMV inoculation

The transcriptional responses of tomato resistant seedlings, inoculated with Fol and ToMV, were evaluated interrogating 15.734 tomato genes. This analysis allowed us to compare changes occurring during both incompatible interaction reactions with a vascular fungus and with a virus. Differentially expressed tomato genes were identified at 2 DPI by linear models for microarray (Smyth 2004.) comparing inoculated and not inoculated plants (FDR-adjusted p-value <0.01). The time point was selected in order to identify genes involved in the initial stages of the defence against the pathogens.

In Fol incompatible interaction, transcriptome variation resulted in 3.753 differentially induced genes. In particular 2.392 genes (about 64%) were up-regulated (Figure 1A) indicating considerable genes activation during infection process. As for the vascular pathogens, 3.501 transcriptional changes were monitored in tomato upon inoculation with ToMV, of which about 2000 (52%) were overexpressed. When total number of differentially regulated genes between the two incompatible interactions were compared, roughly the half (2.205 genes) overlapped of which a small subset of 131 genes, had opposite expression direction during the two interactions (data not shown). As showed in figure 1B, a marked host gene induction was evident, since more than a half of the differentially expressed genes appeared to be induced in both interactions. Moreover, overlapped repressed genes in both interactions are e up to 50%. Taken together these observations suggested that most differences between two plant reactions were caused by genes activation. Transcriptional changes during the two tomato-pathogen interactions were further investigated in order to find the best strategy for building a biological model of a plant-pathogen system.

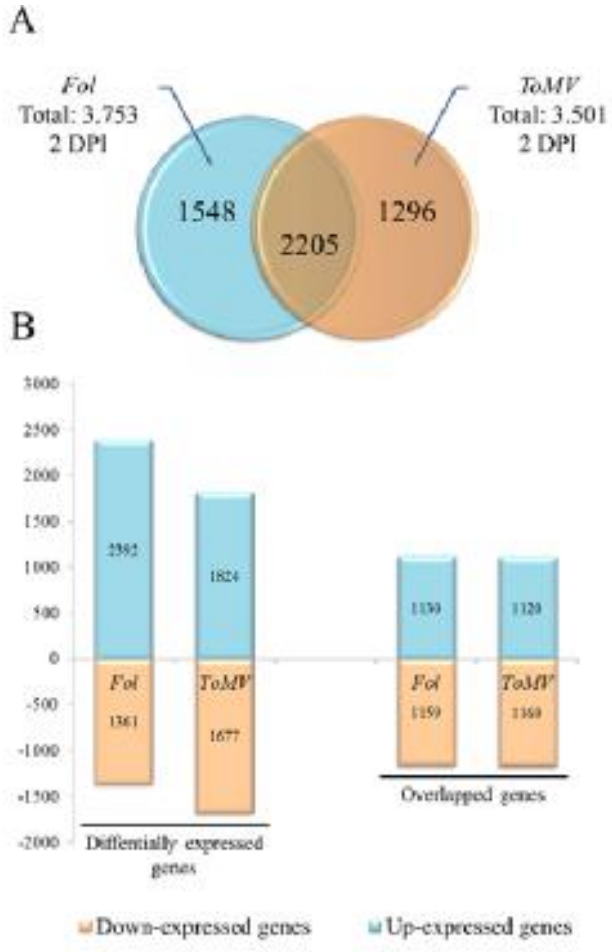

Figure 1. Comparison of tomato differentially expressed gene set during interaction with Fusarium oxysporum (Fol) and Tomato Mosaic Virus (ToMV) at 2 DPI post-inoculation. A, Venn diagram of differentially expressed genes in Fol and ToMV tomato incompatible interactions. B, Histogram of up regulated and down regulate genes in tomato-Fol interaction and tomato-ToMV interaction, respectively.

5.3.2. Gene set enrichment analysis

We performed a GO-term annotation analysis of all transcripts identified by gene chip probe matching. Through this analysis, we were able to assign functional annotations to 15378 chip transcripts. In order to facilitate the formulation of biological hypotheses, a categorization of gene expression data was performed. Indeed, the Gene Ontology effort was extremely useful for structuring data description.

In the tomato-ToMV interaction 226 enriched GO-terms were detected, of which 70% belonged to a biological process (P) category, while in the fungus-interaction less enriched categories (185) were identified of which 78% belonged to a biological process (P) category

(figure 2A). Comparing the total number of GO categories between the two incompatible interactions, roughly 60% overlapped.

Heatmap obtained through hierarchical analysis, evidenced four specific GO terms clusters. The clusters 1 and 4 included GO ontology terms that did not show significant differences in two performed experiments. The cluster 2 contained about 90% of the enriched GO-terms. The sub cluster 2b was composed of specific GO-terms of tomato-Fol interaction, whilst sub clusters 2a 2d as well as cluster 3 included tomato-ToMV interaction specific ontology terms (figure 2B).

To reduce redundancy of functional categories between the enriched GO-terms, we used the semantic similarity approach. The gene ontology interactive graph-based network produced (figure 3) encapsulated functional homology between genes of the two tomato-pathogen interactions. In the tomato-Fol interaction, the network was composed by 53 nodes and 310 edges, while for tomato-ToMV interaction, the network were more complex because it contained 58 nodes and 383 edges.

The network of tomato-Fol interaction showed that the biotic stimulus node (GO:0009607) correlated with the 5 internal nodes with gene ontology terms related to external stimuli response (GO:0009637, GO:0009746, GO:0009628, GO:0042742, GO:0042744). The GO:0009607 category includes a large number of genes, that play a role in signal transduction and regulation of gene expression in the defence response, representing approx. the 10% of differentially expressed genes in tomato-Fol interaction. The external stimuli sub network was also connected with hydrogen peroxide catabolism node (GO:0042744). It contained 28 genes, of which 70% was overexpressed. In particular, four haem peroxidases (Solyc01g006290.2.1; Solyc01g006300.2.1; Solyc06g050440.2.1; Solyc11g018800.1.1) are included which have a role in host defence by inhibiting the hyphal extension of invading pathogens.

Figure 3B showed the network obtained with significant GO terms changes in tomato-ToMV interaction. The biotic stimuli GO:0009607 node correlated with 4 GO-terms nodes (GO:0009637, GO:0006950, GO:0010038, GO:0009628), associated to 264 genes. Interestingly, for this interaction was not evidenced any external link. These data indicated that both incompatible interactions induced well documented stress-responsive genes as well as unknown genes that might play a role in multi-stress responses. Investigation of single GO categories could evidence GO terms related to each assessed specific response.

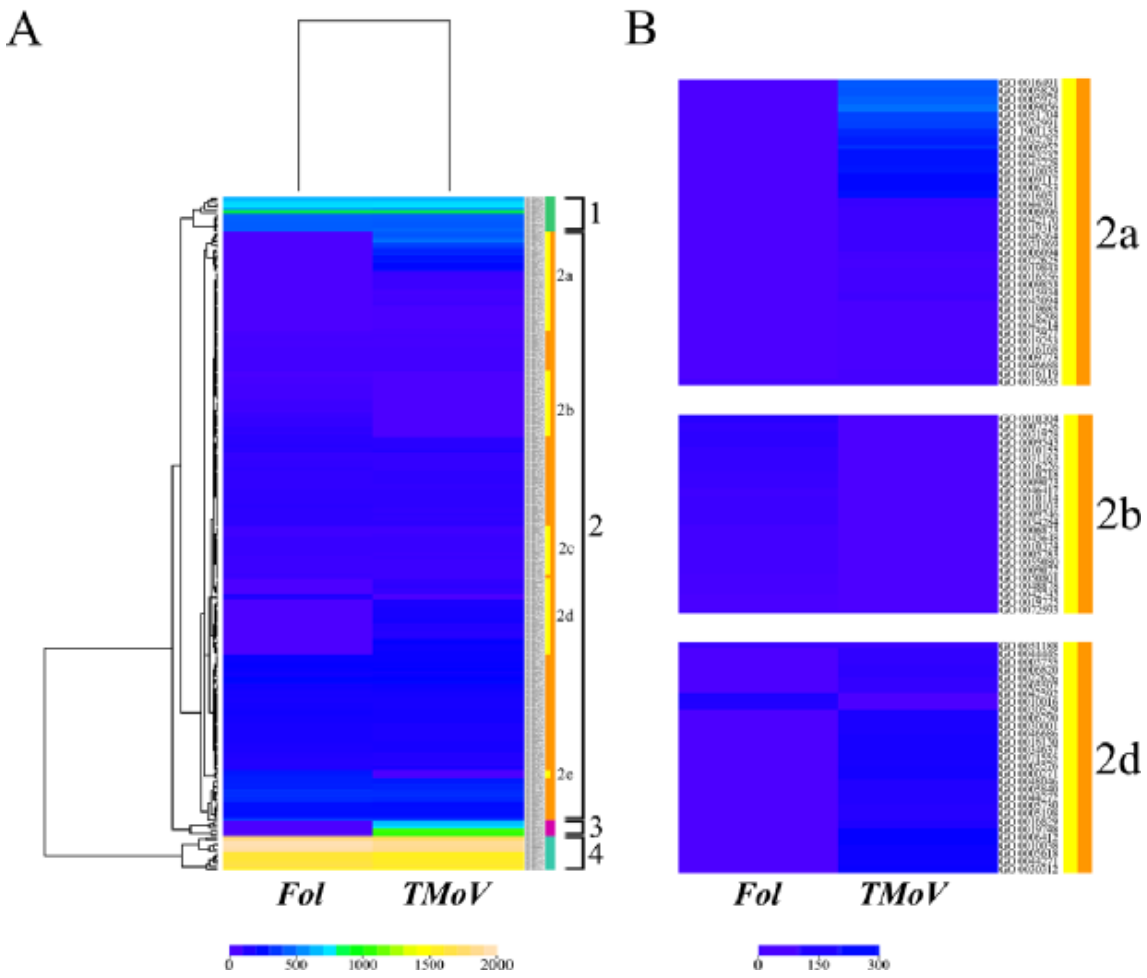

Figure 2. Heat map of Gene Ontology enriched terms in the two interactions. Gene Ontology terms enriched in a test group (differentials of the two interaction) when compared to a reference group (all gene of chip) using Fisher's Exact Test ,with Multiple Testing Correction of FDR (Benjamini and Hochberg, year), estimated at P value of <0.05.A, The rows of heat map represent GO-terms and the columns represent samples. Each cell is colorized based on the number of gene associated with that category go in that particular sample. B, Subclusters composed of specific GO-terms of the tomato-pathogen interactions.

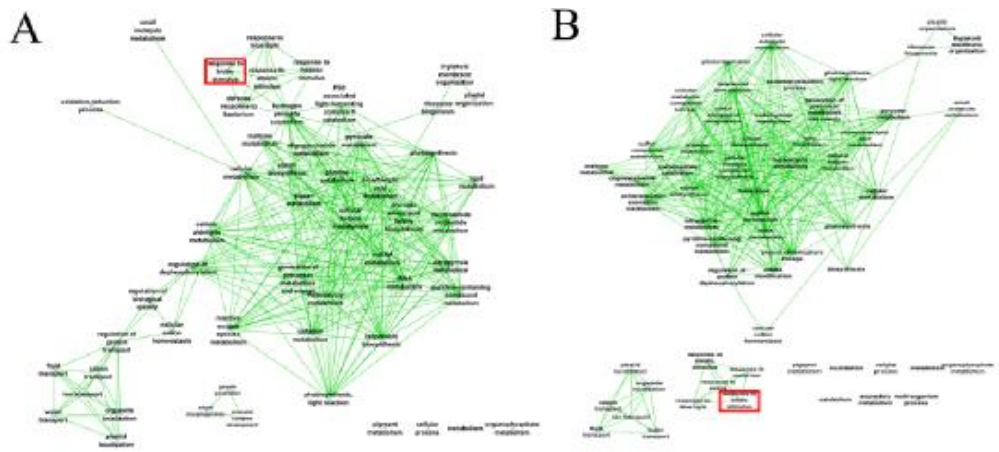

Figure 3. The interactive graph-based enriched gene-ontology (A) in fungus-tomato interaction and (B) in virus-tomato interaction.

5.3.3. Investigation of perturbed biological processes

To obtain an overview of processes involved in fungus and virus incompatible reaction a deeper GO (gene ontology) term analysis was performed. Significant differences were observed in specific regulated biological processes between the two incompatible interactions. Notably, 30 specific enriched GO-term categories for tomato-Fol interaction, and 71 for tomato-ToMV interaction were identified.

Tomato-Fol interaction revealed changes in cell structure, light perception, iron transport, aromatic amino acid synthesis and ROS response. In particular, 6 specific GO-terms associated to homeostatic process (GO:0006873 cellular ion homeostasis; GO:0055080 cation homeostasis; GO:0050801 ion homeostasis; GO:0019725 cellular homeostasis; GO:0048878 chemical homeostasis and GO:0042592 homeostatic process) that included 104 genes were found. This finding supports that homeostasis plays an important role in tomato defence to Fol. Interestingly, the master gene of inflammation, NF-kB (Solyc02g094530.1.1) was also up-regulated in tomato-Fol interaction. This gene is a key player in anti-apoptotic signaling and it is able to prevent the apoptotic signaling by inhibiting the map-kinases. Indeed, tomato map–kinas (Solyc12g019460.1.1; Solyc11g072630.1.1) were found down regulated heavily during this interaction. Table 1 showed 14 auxin binding genes and 2 abscisic acid receptor up-regulated in tomato-Fol interaction.

Photosynthetic and carbohydrate derivative processes, nitrogen biosynthesis, therpen and carotenoid metabolism, cadmium/ copper function were challenged in tomato-ToMV interaction. Three specific GO-terms associated to photosynthetic process (GO:0009853 photorespiration; GO:0019685 photosynthesis dark reaction and GO:0009773 photosynthetic electron transport in photosystem I) were identified and they were investigated further. Table 2 showed the enriched categories associated to photosynthetic process, containing 191 genes of which over 86% was down-regulated. Interestingly, photosynthetic gene transcription repression was inversely correlated with the pathogenesis-related genes induction (Table 3). Moreover, a close linkage between the plant carbohydrate-status with the out coming plant pathogen interaction was evident. We have found 17 carbon fixation enzymes differentially expressed during ToMV-tomato interaction. Furthermore, several genes involved directly in Calvin cycle were down expressed (Table 4) and the β-fructofuranosidase (Solyc10g085640.1.1), a key invertase, involved in sink source portioning, was overexpressed. We also found up-regulated 6 gibberellin modulated genes and 2 genes involved in SA (Solyc01g014320.2.1) and in JA (Solyc07g042170.2.1) synthesis (Table 5).

Gene ID	Probe ID	Putative function	GO-ID terms
Solyc01g091030.2.1	TC196630	Auxin responsive SAUR protein	GO:0009733; GO:0009862; GO:0003674
Solyc02g077560.2.1	TC192465	Auxin response factor	GO:0009725; GO:0009734; GO:0009850; GO:0009733
Solyc03g007310.2.1	TC197392	Abscisic acid receptor	GO:0010427
Solyc03g031970.2.1	TC207784	Auxin response factor	GO:0009734
Solyc03g120500.2.1	TC199757	Auxin responsive protein	GO:0009734
Solyc04g074980.2.1	TC207543	Auxin F-box protein	GO:0009734
Solyc05g047460.2.1	TC205766	Auxin response factor	GO:0009734
Solyc06g061180.1.1	TC193498	Abscisic acid receptor	GO:0010427; GO:0009738
Solyc07g016180.2.1	TC207832	Auxin response factor	GO:0009734
Solyc07g043610.2.1	TC207520	Auxin response factor	GO:0003677; GO:0009734
Solyc07g063850.2.1	TC216744	GH3 auxin-responsive promoter	GO:0010252; GO:0009734
Solyc09g014380.2.1	TC216697	Auxin transporter-like protein	GO:0009734; GO:0010328; GO:0060919; GO:0009926; GO:0010011
Solyc10g050710.1.1	TC202297	GH3 auxin-responsive promoter	GO:0010279; GO:0009733
Solyc10g083320.1.1	TC194758	Auxin responsive SAUR protein	GO:0009733
Solyc11g069190.1.1	TC208143	Auxin response factor	GO:0009734
Solyc12g005310.1.1	TC203007	GH3 auxin-responsive promoter	GO:0010279; GO:0010279; GO:0009734

Table 1. List of abscisic acid receptors, auxin response factors (ARFs) and auxin response genes up-regulated in the tomato-*Fol* interaction. For each entry it is reported: Gene, probe IDs, description of putative function and - GOID term correlated.

GO ID	Term	Category	n. Gene	% Downexpressed Genes
GO:0009521	photosystem	cellular component	25	100
GO:0009522	photosystem I	cellular component	16	100
GO:0009523	photosystem II	cellular component	18	100
GO:0009765	photosynthesis, light harvesting	biological process	17	100
GO:0009767	photosynthetic electron transport chain	biological process	24	100
GO:0009773	photosynthetic electron transport in photosystem I	biological process	15	86,7
GO:0010207	photosystem II assembly	biological process	50	94
GO:0015979	photosynthesis	biological process	125	91,2
GO:0019684	photosynthesis, light reaction	biological process	93	94,6
GO:0019685	photosynthesis, dark reaction	biological process	12	91,7
GO:0034357	photosynthetic membrane	cellular component	127	89,8

Table 2. List of the GO-terms involved in the photosynthetic process during the tomato-*ToMV* interaction. For each GO-term, the category, the number genes and the percentage of down regulation genes is reported.

Protein ID	Probe ID	PRs family	Enzyme	[a]Fold change
Solyc01g006290.2.1	TC197719	PR-9	Peroxidase	5.5
Solyc01g067860.2.1	TC195695	PR-9	Peroxidase	2.3
Solyc01g108320.2.1	TC209477	PR-9	Peroxidase	2.3
Solyc02g064970.2.1	TC201774	PR-9	Peroxidase	8.9
Solyc02g077300.1.1	TC205411	PR-9	Peroxidase	1.6
Solyc02g079510.2.1	TC196561	PR-9	Peroxidase	5.7
Solyc02g080530.2.1	TC209879	PR-9	Peroxidase	6.9
Solyc02g082960.2.1	TC200207	PR-3	Endochitinase	1.7
Solyc02g084780.2.1	TC201521	PR-9	Peroxidase	6.3
Solyc02g094180.2.1	TC211238	PR-9	Peroxidase	2.6
Solyc03g005200.2.1	TC203320	PR-14	Non-specific lipid-transfer protein	2.8
Solyc03g033710.2.1	TC203104	PR-9	Peroxidase	7.1
Solyc03g044100.2.1	TC199639	PR-9	Peroxidase	5.8
Solyc04g007750.2.1	TC199935	PR-10	Major latex-like protein	3.2
Solyc04g071890.2.1	TC193192	PR-9	Peroxidase	4.3
Solyc04g072000.2.1	TC212476	PR-3	Chitinase	1.2
Solyc04g080760.2.1	TC206568	PR-9	Peroxidase	2.1
Solyc05g046010.2.1	TC202014	PR-9	Peroxidase	2.4
Solyc05g050130.2.1	TC194980	PR-11	Acidic chitinase	2.8
Solyc06g050440.2.1	TC211424	PR-9	Peroxidase	6.9
Solyc06g072220.1.1	TC210670	PR-6	Kunitz trypsin inhibitor	8.5
Solyc06g076630.2.1	TC201658	PR-9	Peroxidase	7.9
Solyc07g005380.2.1	TC194633	PR-10	Norcoclaurine synthase	2.6
Solyc07g009260.2.1	TC210032	PR-12	Defensin-like protein	2.3
Solyc07g017880.2.1	TC209710	PR-9	Peroxidase	3.8
Solyc07g052510.2.1	TC208195	PR-9	Peroxidase	8.7
Solyc08g067510.1.1	TC213655	PR-14	Non-specific lipid-transfer protein	1.5
Solyc09g007520.2.1	TC200697	PR-9	Peroxidase	3.9
Solyc09g065430.2.1	TC214938	PR-14	Non-specific lipid-transfer protein	2.1
Solyc09g082280.2.1	TC209373	PR-14	Non-specific lipid-transfer protein	8.9
Solyc10g076220.1.1	TC209457	PR-9	Peroxidase	6.4
Solyc11g072760.1.1	TC215379	PR-11	Chitinase a	3.7
Solyc12g017660.1.1	TC198426	PR-9	Peroxidase	2.9

Table 3. List of up-regulated pathogenesis-related proteins involved in tomato-ToMV interaction. For each entry it is reported: gene and probe IDs, the PRs-family to which belongs,the description of its function and the fold change. [a]Fold change is calculated using the signal $\log_2$ ratio.

Gene ID	Function Description	Ezyme ID	Enzyme Type	Up/Down-expressed
Solyc08g013860.2.1	Malic enzyme	ec:1.1.1.39	dehydrogenase	Up-
Solyc12g044600.2.1	Malic enzyme	ec:1.1.1.40	dehydrogenase	Down-
Solyc03g071590.2.1	Malate dehydrogenase	ec:1.1.1.82	dehydrogenase	Up-
Solyc04g009030.2.1	Glyceraldehyde-3-phosphate dehydrogenase	ec:1.2.1.13	dehydrogenase	Down-
Solyc12g094640.1.1	Glyceraldehyde-3-phosphate dehydrogenase	ec:1.2.1.13	dehydrogenase	Down-
Solyc10g018300.1.1	Transketolase	ec:2.2.1.1	glycolaldehydetransferase	Down-
Solyc05g050970.2.1	Transketolase	ec:2.2.1.1	glycolaldehydetransferase	Up-
Solyc08g079750.2.1	1-aminocyclopropane-1-carboxylate synthase Aminotransferase	ec:2.6.1.1	transaminase	Down-
Solyc07g032740.2.1	Aspartate aminotransferase	ec:2.6.1.1	transaminase	Up-
Solyc07g055210.2.1	Aspartate aminotransferase	ec:2.6.1.1	transaminase	Up-
Solyc11g044840.1.1	LL-diaminopimelate aminotransferase	ec:2.6.1.1	transaminase	Up-
Solyc05g013380.2.1	Alanine aminotransferase	ec:2.6.1.2	transaminase	Down-
Solyc06g063090.2.1	Alanine aminotransferase	ec:2.6.1.2	transaminase	Up-
Solyc08g076220.2.1	Phosphoribulokinase	ec:2.7.1.19	phosphopentokinase	Down-
Solyc04g008740.2.1	Pyruvate kinase	ec:2.7.1.40	kinase	Up-
Solyc07g066600.2.1	Phosphoglycerate kinase	ec:2.7.2.3	kinase	Down-
Solyc07g066610.2.1	Phosphoglycerate kinase	ec:2.7.2.3	kinase	Down-
Solyc09g011810.2.1	Fructose-1 6-bisphosphatase	ec:3.1.3.11	hexose diphosphatase	Down-
Solyc10g086730.1.1	Fructose-1 6-bisphosphatase	ec:3.1.3.11	hexose diphosphatase	Down-
Solyc12g014250.1.1	Phosphoenolpyruvate carboxylase	ec:4.1.1.31	carboxylase	Down-
Solyc10g007290.2.1	Phosphoenolpyruvate carboxylase	ec:4.1.1.31	carboxylase	Up-
Solyc01g007330.2.1	Ribulose bisphosphate carboxylase	ec:4.1.1.39	carboxylase	Down-
Solyc03g034220.2.1	Ribulose bisphosphate carboxylase	ec:4.1.1.39	carboxylase	Down-
Solyc06g061280.2.1	Cinnamoyl-CoA reductase-like protein	ec:4.1.1.49	carboxykinase	Up-
Solyc01g110360.2.1	Fructose-bisphosphate aldolase	ec:4.1.2.13	aldolase	Down-
Solyc07g065900.2.1	Fructose-bisphosphate aldolase	ec:4.1.2.13	aldolase	Down-
Solyc09g009260.2.1	Fructose-bisphosphate aldolase	ec:4.1.2.13	aldolase	Up-
Solyc10g083570.1.1	Fructose-bisphosphate aldolase	ec:4.1.2.13	aldolase	Up-
Solyc06g005490.2.1	Triosephosphate	ec:5.3.1.1	isomerase	Down-

Table 4. List of genes involved in carbon fixation differentially expressed during the tomato-ToMV interaction. For each genes it is reported: the functional description, the enzyme ID, the enzyme class and direction of expression regulation.

Gene ID	Probe ID	Putative function	GO-ID terms
Solyc02g083880.2.1	TC203352	Gibberellin-regulated protein	GO:0005515
Solyc06g008870.2.1	TC202708	Gibberellin receptor	GO:0005515; GO:0008152
Solyc06g035530.2.1	TC202928	Gibberellin 20-oxidase-2	GO:0016491; GO:0045544; GO:0055114
Solyc10g005360.2.1	TC204547	Gibberellin 2-beta-dioxygenase	GO:0016491,GO:0045543
Solyc10g050880.1.1	TC198803	Gibberellin receptor	GO:0004091,GO:0008152
Solyc11g072310.1.1	TC211262	Gibberellin 20-oxidase-3	GO:0016491,GO:0045544,GO:0055114

Table 5. List of the gibberellin modulated genes, up-regulated in tomato-ToMV interaction. For each it is reported: genes and probe IDs, the description of putative function and the GO-ID terms correlated.

5.3.4. Relationship between genomics and transcriptional changes

In order to investigate chromosomal arrangement of differential expressed sequences identified by tomato chip, a genomic expression map was constructed (figure 4). Our results indicated that differentially genes expressed (about 33% of total analyzed genes) during the two interactions were arranged along the chromosomes with different level of expression.

The genome-wide distribution of differentially expressed genes, based both on the chromosome size and gene density calculated at was not uniformly distributed along the genome (chromosome size: fungus-interaction χ^2 =107; virus-interaction χ^2 =117. Gene density: χ^2 value included between 27 and 170).

Indeed, the genome-wide distribution of differentially expressed genes, based both on the chromosome size and gene density is clearly non-uniform both for under-expressed and over-expressed genes (Figure 4A and B NEW2). If we use the genome location of the coding genes spotted on the microarray as a reference we could highlight a significant difference in the distribution of over-expresses genes as well as under-expressed ones with respect to the reference distribution of the coding genes (Kolmogorov-Smirnov and Cramer von Mises tests in Table 6 and 7). Indeed, there is some weak evidence for either one sign of differentially expressed genes (p-value <0.10) on all chromosomes but not on the 2, 4 and 7. Stronger evidence of peculiar chromosomal displacement of over-expressed genes is on chromosomes 3, 5, 6 and 1 (p-value <0.05). Significantly differential displacement of under-expressed genes (p-value <0.05) has been detected mainly in chromosomes 5 and 9.

This confirmed the presence of specific genomic regions involved in pathogen response. The fungus and virus distribution showed slight differences. Indeed, the number of expressed genes identified for single chromosomes, was always greater in tomato-ToMV interaction than in tomato-Fol interaction, with exception of chromosome 8. Chromosomal areas (on chromosomes 1, 2, 8, 11 and 12) showed a higher density of genes up-regulated in the fungus-interaction. While on chromosomes 8 it was found a region with a higher density of genes down-regulated in the virus-interaction. In both interactions, the genomic regions rich in differentially expressed genes were located in the vicinity of telomeres, inter alia the richest regions of genes were long arm of chromosomes 10 and 12.

For each studied interaction, distribution of expressed specific genes is showed in figure 5. Most of them were located in same regions, but clear differences could be highlighted. The 151 common differentially expressed genes with opposite direction were mainly arranged on chromosome 7, 9 11 and 12. On chromosomes 9 and 11, in the vicinity of the genes Tm2 conferring resistance to Tomato mosaic virus and I2 conferring resistance to Fusarium oxysporum, we highlighted difference in gene expression between the two experiments. Four genes were specifically expressed in ToMV interaction in this area of chromosome 9 (Pentatricopeptide repeat, ATP dependent RNA helicase, Cytochrome P450 and Pectinesterase) and five of I2 region of chromosome 11 in Fol interaction (BEL1-like homeodomain protein 8, Calmodulin-like protein and MYB transcription factor and two protein of unknow function).

Figure 6 shows differential expressed pathogen recognition genes both in tomato-Fol and tomato-ToMV interactions. A pronounced activation of all pathogen recognition gene classes assessed was evidenced in tomato-Fol interaction since the 70% of the total differentially expressed genes were up regulated. Instead in tomato-ToMV interaction the number of up-regulated and down-regulated pathogen recognition genes was equivalent. The 45% of the total of differentially regulated pathogen recognition genes overlapped. Among them 10 genes had opposite expression between the two experiments.

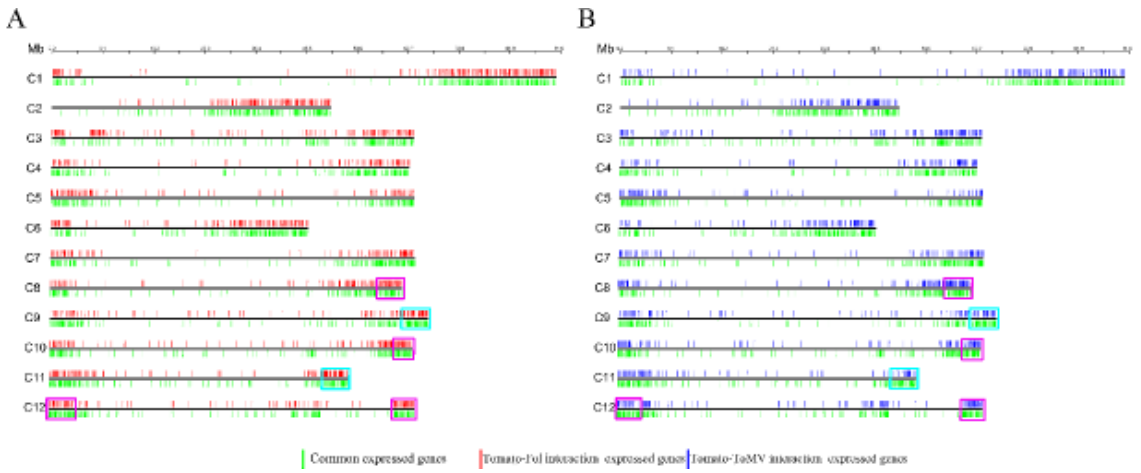

Figure 4. Physical genome locations of tomato (*Solanum lycopersicum*) differentially expressed genes. **A**, Physical map of the tomato-fungus interaction. **B**, Physical map of the tomato-virus interaction. Tomato chromosomes are represented as black horizontal bars, the approximate location of each gene is designated with vertical lines on each chromosome. The color used for each gene indicates specificity and overlap in differentially expressed genes in the two interactions. The cyan boxes on chromosomes 9 and 11 indicates the chromosomal regions in the vicinity of the Tm2 and I2 genes. The violet boxes on chromosomes 8, 10 and 12 indicates the richest regions of genes.

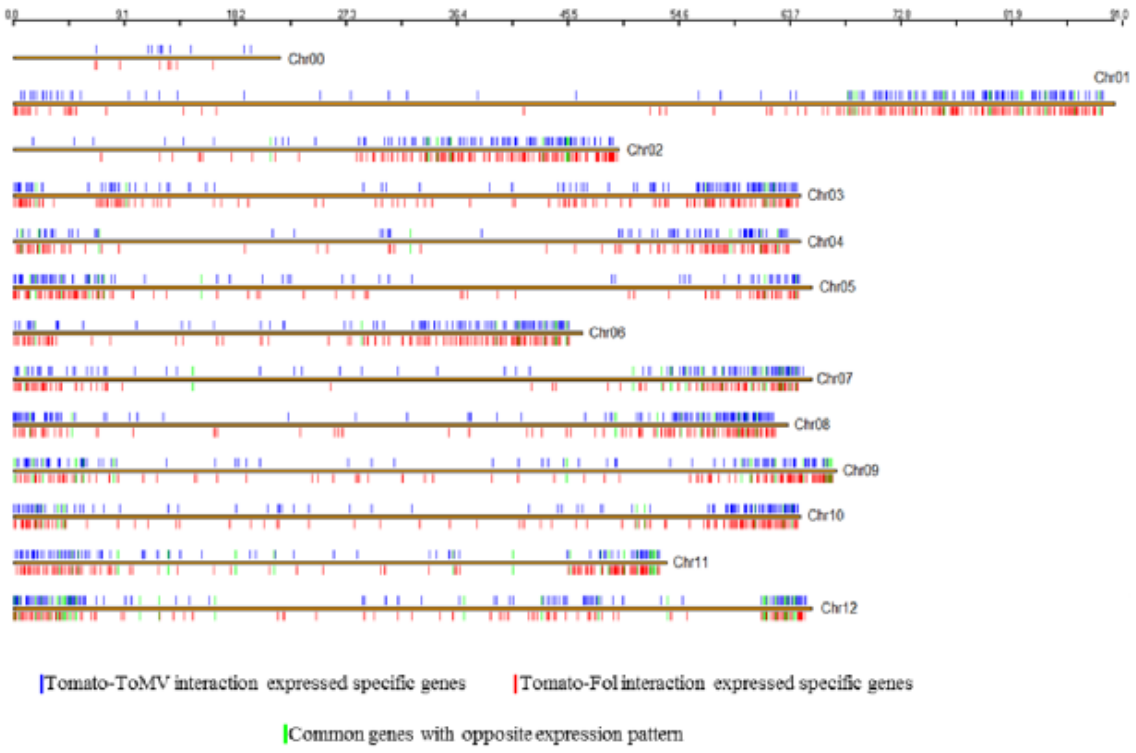

Figure 5. Physical map of specific differentially expressed genes in tomato-ToMV and tomato-Fol interactions. Tomato chromosomes are represented as brown horizontal bars, the approximate location of each gene is designated with vertical lines on each chromosome. The color used for each gene indicates specificity and overlap in differentially expressed genes in the two interactions.

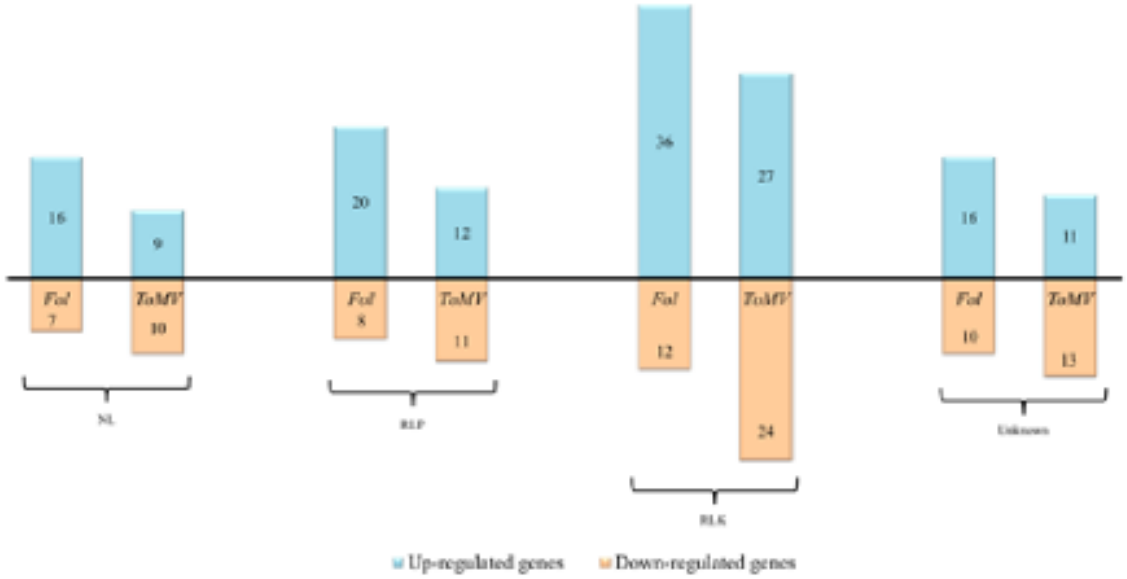

Figure 6. Comparison of differentially regulated tomato pathogen recognition genes sets during interaction with *Fusarium oxysporum (Fol)* and *Tomato Mosaic Virus (ToMV)*. Histogram displaying differentially regulated pathogen recognition genes for each structural class in two incompatible interactions at 2 DPI post-inoculation.

Chrom	Fol.ks	Fol.cvm	ToMV.ks	ToMV.cvm
01	0.753	0.617	0.290	*0.083*
02	0.597	0.660	0.711	0.509
03	**0.024**	*0.069*	0.163	0.275
04	0.885	0.762	0.634	0.678
05	*0.078*	**0.027**	0.202	*0.075*
06	0.322	0.260	**0.048**	**0.022**
07	0.812	0.874	0.940	0.894
08	0.715	0.925	0.764	0.816
09	0.293	0.356	0.632	0.561
10	0.851	0.592	0.123	*0.099*
11	**0.011**	**0.031**	0.816	0.610
12	0.460	0.315	0.510	0.376

Table 6. Results of Kolmogorov-Smirnov test and Cramer von Mises for analyzing the differences between the distribution of position of **over**-expressed genes and coding genes spotted on the microarray chip.

Chrom	Fol.ks	Fol.cvm	ToMV.ks	ToMV.cvm
01	0.650	0.578	0.828	0.869
02	0.531	0.403	0.341	0.281
03	0.501	0.655	*0.060*	0.130
04	0.731	0.881	0.978	0.934
05	**0.048**	*0.052*	0.128	0.122
06	0.799	0.724	0.982	0.943
07	0.806	0.733	0.288	0.144
08	0.135	*0.050*	*0.086*	0.105
09	*0.070*	**0.045**	0.534	0.554
10	0.782	0.898	0.993	0.955
11	0.363	0.510	0.216	0.367
12	0.161	*0.097*	0.752	0.648

Table 7. Results of Kolmogorov-Smirnov test and Cramer von Mises for analyzing the differences between the distribution of position of **under**-expressed genes and coding genes spotted on the microarray chip.

5.4. DISCUSSION

5.4. Discussion

Plant-expressed molecules during pathogen challenging can give insights into the underlying defence mechanisms. plants have evolved a defence system against microbial pathogens that involves the regulation of gene expression, cascade signaling activation, hormone balancing and synthesis of defensive metabolites (Mithöfer and Boland 2012). Several microarray experiments have been successfully carried out in molecular plant–microbe interaction area, elucidating mechanisms controlling plant disease resistance and the crosstalk among the signalling pathways involved (Lodha and Basak 2012). More of 3500 genes, out of 15734 assessed, were activated in tomato Fol and ToMV interactions, of which more of 60% overlapped. Pathogens from different kingdoms deploy independently evolved virulence proteins that interact with a limited set of highly connected plant cellular hubs (Mukhtar et al. 2011). Network analysis suggested that plant response to Fol and ToMV infection involved activation or repression of genes implicated in pathway shared by most response to environmental stimuli. However, GO enrichment analysis evidenced specific enriched categories in both interactions. Response to TMV seemed more multifaceted, since more of 70 specific categories were enriched versus the 30 detected in Fol interaction. Biotrophic plant pathogens are generally accepted to have a more intricate biological interaction with their host plant than saprophytic plant pathogenic fungi (Hammond - Kosack and Rudd 2008). Once an attack was perceived, plant metabolism must balance potentially competing demands for resources to support defence versus requirements for cellular maintenance, growth and reproduction (Herms and Mattson 1992; Zangerl and Berenbaum 1997, Zangerl and Berenbaum 2003; Berger et al. 2007).

In tomato-Fol interaction our investigation evidenced a number of overexpressed genes associated to maintenance of cellular structures and cellular homeostasis. maintenance of cellular structures and cellular homeostasis is a very important metabolic activity required by plants in order to survive fungus inflicted stresses. Fusarium oxysporum is a saprophytic fungus (Trusov et al. 2006) that kills host cells prior to the infection through the predicted deployment of toxins and enzymes, which induces cell death, therefore the expression of anti-apoptosis genes would confer resistance to Fol (Paul et al. 2011). The NF-kB signaling enhancement found in this work can inhibit apoptosis and combat the effects of oxidative stress switching down the anti-apoptotic signals mediated by map-kinase cascade. These events lead to hypothesize that resistant plant challenged by Fol induces a homeostatic response to prevent the attempt of fungus to kill the cells for getting its nourishment.

Fourteen auxin binding genes were also activated in this interaction. Auxin signaling is an important phytoregulator for resistance to the saprophytic/necrotrophic fungi (Llorente et al. 2008). The degradation of AUX/IAA proteins allows activation of auxin response factors (ARFs) and the expression of auxin-responsive genes (Hagen and Guilfoyle 2002). These protein families can have an enormous number of interactions, capable of fine-tuning specific responses within the auxin signaling pathway (Guilfoyle 2007). Moreover, the buildup of IAA nearby the sites of pathogen ingress constitutes one of the main host factors that determine plant resistance to Fusarium wilt (Dalila Paz-Lago et al. 2000). Essential regulators of Jasmonate-ZIM domain proteins (JAZs) in tomato-Fol interaction were activated, suggesting an attenuated JA signaling for reducing senescence process. Indeed, also a couple of abscisic acid receptors were up-regulated in Fol interaction. Interactions between multiple components of ABA and the JA-ethylene signalling pathways modulate defence and stress responsive gene expression in response to Fol, confirming that these hormones mediate resistance in necrothofic interaction (Badruzsaufari et al. 2004; Koornneef et al. 2008). SA pathway was switched down. Indeed silencing a tomato gene encoding SAmethyltransferase was found to enhance resistance to F. oxysporum f. sp. lycopersici biotic and abiotic stresses(Ament et al. 2010).

In tomato-ToMV interaction, we detected a reduced photosynthesis activity and elevated carbohydrate catabolism. It is possible that pathogen attack caused metabolic alterations that induced a hormone response for reprogramming cellular metabolism. The photosynthesis and assimilatory metabolism were switched off to initiate respiration and other processes required for the defence. This occurrence reflected the reallocation of plant metabolites from normal growth processes to defensive functions after the elicitation of induced plant responses by virus infection (Handford and Carr 2007). A close linkage between the plant carbohydrate-status with the out coming plant pathogen interaction was also evident. During interaction with the virus, host produced salicylic acid and increased the production of an invertase enzyme that is able to redirect the flux of carbohydrates acquisition. The infection of ToMV could have determined the alteration of the concentration gradient of sucrose in phloem. The plant overexpressed the β-fructofuranosidase (Solyc10g085640.1.1) to ensure the sucrose transport of source cell to the sink cell. Overexpression of invertase in tomato-virus interaction could be elicited by the expression of pathogenesis-related proteins and by salicylic acid in order to increase resistance against virus infection (Herbers et al. 2000). In order to ensure the success of defence, this mechanism appears to be crucial. The phenomenon of “high sugar resistance” was described long time ago (Horsfall and Dimond

1957) and the finding that various pathogenesis-related genes were sugar inducible (Herbers et al. 1996) supports this hypothesis. We also found 6 gibberellin, one SA gene (Solyc01g014320.2.1) and one JA (Solyc07g042170.2.1) modulating genes up-regulated, suggesting that GAs activated tomato immune responses to ToMV by modulating the levels of salicylic acid and/or jasmonic acid (de Torres - Zabala et al. 2009).

Plants rely on the innate immunity and on systemic signals emanating from metabolic alterations (Jones and Dangl 2006). ETI system seems to prompt response in the right direction thanks to metabolic clues and hormone signalling. Salicylic acid (SA) primarily triggers resistance against biotrophic and pathogens, whereas a combination of jasmonic acid (JA) and ethylene (ET) signalling activates resistance against necrotrophic pathogens (Glazebrook 2005; Robert-Seilaniantz et al. 2011). The metabolic changes associated with defence response in two incompatible tomato-pathogen interactions suggested that the response to the specific metabolic alteration (photosynthesis/ carbohydrate metabolism and homeostasis activity) in tomato was pathogen specific and contributed substantially to monogenetic gene-for-gene resistance. Our results confirm that resistance to pathogen depends on a sophisticated interplay among different biological pathways and that hormonal directionality is critical to the outcome of a response (carbohydrate biosynthesis and homeostasis activity). Several hormones involved in pathogen perception, activation of defence products restricting pathogen invasion, have been identified (Lòpez et al. 2008). How they trigger ETI defence is currently unclear. In the absence of a pathogen, NBS genes stand in an autoinhibition state, which is relieved upon pathogen perception (Lukasik and Takken 2009). Different forms of plant immunity share the same signalling mechanisms, but they use the same mechanisms in very different ways (Tsuda et al. 2009). The mechanisms that lead rapid metabolism switch and the connection among the overall defence pathways is still not clear. Signalling able to fine tuning the defence mechanism could be activated by genes in proximity of the main pathogen receptors.

Genome organization of functional gene networks to tolerate alterations can determine plasticity. In two specific interactions expressed genes chromosome distribution showed wide overlapping regions, except that for region holding I2 gene and Tm2 gene. Indeed, genome regions can be enriched in genes with specific function for fine tuning gene expression in a compensatory way. In Fol interaction a calmdulin gene, a myb factor and a BEl like were specifically expressed. Calmodulin (CaM) plays an important role in sensing and transducing changes in cellular Ca2+ concentration in response to several biotic and

abiotic stresses. In ToMV interaction Pentatricopeptide repeat, ATP dependent RNA helicase could be involved in the process of silencing and virus replication Cytochrome P450 and Pectinesterase.

A high level of activation NBS genes and of others gene classes potentially involved in pathogen recognition was also found suggesting a host-coordinated reaction of guard machinery to monitor integrity of cellular proteins. NBS system minimizes the cost to the plant for defence, as multiple NBS-LRR proteins can be maintained at a low level in the absence of a pathogen and rapidly induced under pathogen attack through a miRNA regulation (Li et al. 2011). Pathogen-encoded suppressors of RNA silencing mechanisms might result in the induction of multiple NBS-LRR defence proteins (Shivaprasad et al. 2012). Investigation on pathogen recognition genes differentially regulated could lead the identification of specific modulation patterns. Although still fragmented, our depiction provides a global view of the I2 an Tm2 mediated resistance process, considering gene expression network as a starting point to construct a genomic model for the R mediated response into two investigated photosystems.

7. Conclusions and perspectives

In this work we get insight in the complexity of plant innate immune system through omic-approaches. For realizing this objective, it was necessary to acquire a deep knowledge of genomic organization of pathogen recognition gene family. The functional and structural genomic investigations were necessary to frame the plant innate immunity in a dynamic system, where pathogen recognition genes occupy a leading role. This genomic overview was the fundamental point from which the exploration of the transcriptomic profile started in a plant-pathogen interaction study.

Specifically, the four overall addressed aims were:

- A *Solanum* genome-wide R loci spatial arrangement investigation. A complete catalogue of *Solanum lycopersicum* and *Solanum tuberosum* nucleotide-binding site (NBS) NBS, receptor-like protein (RLP) and receptor-like kinase (RLK) gene repertories was generated. NBS genes showed to occur frequently in clusters of related gene copies that included RLP or RLK genes. This scenario is compatible with the existence of selective pressures optimizing coordinated transcription. A number of duplication events associated with lineage-specific evolution were also discovered. These findings suggest that different evolutionary mechanisms shaped pathogen recognition gene cluster architecture to expand and to modulate the defense repertoire.
- A refined NB-LLR *Solanum* annotation . A RenSeq bait library to reannotate the full NB-LRR gene complement in *Solanum lycopersicum* Heinz 1706 and S. pimpinellifolium LA1589 was produced. It allowed us to identify novel sequences that were not picked up by the automated gene prediction software. Phylogenetic analyses showed a high conservation of all NB-LRR classes between Heinz 1706, LA1589 and the potato clone DM, suggesting that all sub-families were already present in the last common ancestor. Use of RenSeq on cDNA from uninfected and late blight-infected tomato leaves allows the avoidance of sequence analysis of non-expressed paralogues. cDNA RenSeq enables for the first time next-gen sequencing approaches targeted to this very low-expressed gene family without the need for normalization.
- An integrated genomic approach for identifying new resistance (R) gene candidates. In this thesis, we show that information on the tomato genome can be used predictively to link resistance function with specific sequences. An R gene functional map was created

by co-localization of candidate pathogen recognition genes and anchoring molecular markers associated with resistance phenotypes. In-depth characterization of the identified pathogen recognition genes was performed. Such methodology can help to better direct positional cloning, reducing the amount of effort required to identify a functional gene. The resulting candidate loci selected are available for exploiting their specific function.

- A tomato transcriptional response to *F. oxysporum* and *Tomato Mosaic Virus* analysis. In order to identify a set of genes of interest in tomato plants infected with *F. oxysporum f. sp. lycopersici* (Fol) and *Tomato Mosaic Virus* (ToMV) a transcriptional analysis was performed. A large overlap was found in differentially expressed genes throughout the two incompatible interactions. Response to ToMV seems more multifaceted, since more than 70 GO tspecific categories were enriched versus the 30 detected in Fol interaction. Genomic mapping of transcripts suggested that specific genomic regions are involved in pathogen resistance response. Coordinated R gene machinery could have an important role in prompt the response, since the 60% of pathogen receptor genes were differentially expressed during both interactions.

The results produced in this work pose new questions, showing the unknown hugeness of the plant biology. However, the accumulated knowledge have allowed us to better understand the system of plant-pathogen interaction and to start to draw a new plant immune system model.

8. References

Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. Current Opinion in Plant Biology 8: 135–141.

Al-Shahrour F, Minguez P, Marques-Bonet T, Gazave E, Navarro A, Dopazo J. 2010. Selection upon genome architecture: conservation of functional neighborhoods with changing genes. PLoS Computational Biology 6: e1000953.

Alvarez, M.E., F. Nota and D.A. Cambiagno. 2010. Epigenetic control of plant immunity. Molecular Plant Pathology. 11: 563-76.

Alvarez, M.E., R.I. Pennell, P.J. Meijer, A. Ishikawa, R.A. Dixon and C. Lamb. 1998. Reactive oxygen intermediates mediate a systemic signal network in the establishment of plant immunity. Cell. 92: 773-84.

Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB. 2008. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant Medicago truncatula. Plant Physiology 146: 5–21.

Andolfo G, Sanseverino W, Aversano R, Frusciante L, Ercolano MR. 2013. Genome-wide identification and analysis of candidate genes for disease resistance in tomato. Mol breeding, 9928: 7.

Andolfo G, Sanseverino W, Rombauts S, Van der Peer J, Bradeen JM, Carputo D, Frusciante L, Ercolano MR. 2013. Overview of tomato (Solanum lycopersicum) candidate pathogen recognition genes reveals important Solanum R locus dynamics. New Phytol 197:223–237

Antony, G., J. Zhou, S. Huang, T. Li, B. Liu, F. 2010. White and B. Yang, Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os- 11N3. The Plant cell. 22: 3864-76.

Badruzsaufari, E., Schenk, P.M., Manners, J.M., Desmond, O.J., Ehlert, C., Maclean, D.J., Ebert, P.R., Kazan, K. and Anderson, J.P. 2004. Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis. Plant Cell. 16: 3460-3479.

Bakker, E.G., C. Toomajian, M. Kreitman and J. Bergelson. 2006. A genome-wide survey of R gene polymorphisms in Arabidopsis. The Plant cell. 18: 1803-18.

Bari, R. and J.D. Jones. 2009. Role of plant hormones in plant defence responses. Plant Molecular Biology. 69: 473-88.

Beckman, C.H. 2000. Phenolic-storing cells: keys to programmed cell death and periderm formation in wilt disease resistance and in general defence responses in plants? Physiol. Mol. Plant Pathol. 57: 101–110.

Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt RL, Giraudat J, Leung JL, Staskawicz BJ. 1994. RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant disease resistance genes. Science. 265, 1856-1860.

Bergelson J, Kreitman M, Stahl EA, Tian D. 2001. Evolutionary dynamics of plant R-genes. Science 292: 2281–2285. Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16: 1667–1678.

Berger, S., Sinha, A.K., and Roitsch, T. 2007. Plant physiology meets phytopathology: plant primary metabolism and plant-pathogen interactions. J Exp Bot. 58: 4019-26.

Caicedo AL, Schaal BA. 2004. Heterogeneous evolutionary processes affect R-gene diversity in natural populations of Solanum pimpinellifolium. Proc Natl Acad Sci USA 101:17444–17449

Chen Q, Han Z, Jiang H, Tian D, Yang S. 2010. Strong positive selection drives rapid diversification of R-genes in Arabidopsis relatives. Journal of Molecular Evolution 70: 137–148.

Chen, L.Q., B.H. Hou, S. Lalonde, H. Takanaga, M.L. Hartung, X.Q. Qu, W.J. Guo, J.G. Kim, W. Underwood, B. Chaudhuri, et al.. 2010. Sugar transporters for intercellular exchange and nutrition of pathogens. Nature. 468: 527-32.

Chisholm ST, Coaker G, Day B, Staskawicz BJ. .2006. Host-microbe interactions: shaping the evolution of the plant immune response. Cell. 124:803–814.

Choulakian, V. R. A. Lockhart, and M. A. Stephens (1994). Cramer-von Mises statistics for discrete distributions. The Canadian Journal of Statistics, 22: 125-137.

Chunwongse J, Chunwongse C, Black L, Hanson P. 1998. Mapping of Ph-3 gene for late blight from L. pimpinellifolium L3708. Report of the Tomato Genetics Cooperative, 48:13–14.

Cock, J.M., V. Vanoosthuyse and T. Gaude. 2002. Receptor kinase signalling in plants and animals: distinct molecular systems with mechanistic similarities. Current Opinion in Cell Biology. 14: 230-236.

Collier SM, Hamel LP, Moffet P: Cell Death Mediated by the N-Terminal Domains of a Unique and Highly Conserved Class of NB-LRR Protein. Molecular Plant-Microbe Interactions 2011, 24, 918-931.

Conover William J. (1972), A Kolmogorov Goodness-of-Fit Test for Discontinuous Distributions.Journal of American Statistical Association, Vol. 67, No. 339, 591–596.

CPVO-TP/44/3 technical protocol. Community plant variety office. Protocol for distinctness, uniformity and stability tests. Lycopersicon lycopersicum L. Karsten ex. Farw. TOMATO. 21/03/2007. website www.cpvo.europa.eu

Cui, H., Y. Wang, L. Xue, J. Chu, C. Yan, J. Fu, M. Chen, R.W. 2010. Innes and J.M. Zhou, Pseudomonas syringae effector protein AvrB perturbs Arabidopsis hormone signaling by activating MAP kinase 4. Cell Host & Microbe. 7: 164-75.

Dalila Paz-Lago, A., Borges, Jr., A., Gutiérrez, A. Borges, G., Cabrera, M. A. Ramírez A., and Falcón, A.. .2000. Tomato Fusarium oxysporum interactions:ii-chitosan and msb induced resistance against fol in young tomato plants. Cultivos Tropicales 21: 17-20.

de Torres Zabala, M., Bennett, M.H., Truman, W.H., and Grant, M.R. 2009 Antagonism between salicylic and abscisic acid reflects early host-pathogen conflict and moulds plant defence responses. Plant J. 59: 375-386.

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD. 2002. A comparative genetic linkage map of eggplant (Solanum melongena) and its implications for genome evolution in the Solanaceae. Genetics 161:1697–1711

Dong, X. 2004. NPR1, all things considered. Curr. Opin. Plant Biol. 7, 547–552.

Durrant, W.E. & Dong, X. 2004. Systemic acquired resistance. Annu. Rev. Phytopathol. 42, 185–209.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–1797.

Ercolano MR, Sanseverino W, Carli P, Ferriello F, Frusciante L. 2012. Genetic and genomic approaches for R-gene mediated disease resistance in tomato: retrospects and prospects. Plant Cell Reports 31: 973–985.

Ercolano, M.R., Sanseverino, W., Carli, P., Ferriello, F., and Frusciante, L. 2012. Genetic and genomic approaches for R-gene mediated disease resistance in tomato: retrospects and prospects. Plant Cell Rep. 31: 973-85.

Erickson FL, Holzberg S, Calderon-Urrea A, Handley V, Axtell M, Corr C, Baker B. 1999. The helicase domain of the TMV replicase proteins induces the N-mediated defence response in tobacco. Plant Journal 18: 67–75.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791.

Feng, F., F. Yang, W. Rong, X. Wu, J. Zhang, S. Chen, C. He and J.M. Zhou. 2012. A Xanthomonas uridine 5'-monophosphate transferase inhibits plant immune kinases. Nature. 485: 114-8.

Fenyk, S., S. Campillo Ade, E. Pohl, P.J. Hussey and M.J. Cann. 2012. A nucleotide phosphatase activity in the nucleotide binding domain of an orphan resistance protein from rice. The Journal of Biological Chemistry, 287: 4023-32.

Flor H.H.. 1971. "Current status of the gene-for-gene concept". Annu Rev Phytopathol 9: 275–296.

Foolad MR. 2007. Genome mapping and molecular breeding of tomato. International Journal of Plant Genomics. 64358:52.

Fritz-Laylin LK, Krishnamurthy N, Tor M, Sjolander KV, Jones JD. 2005. Phylogenomic analysis of the receptor-like proteins of rice and Arabidopsis. Plant Physiology 138: 611–623.

Gary Loake and Murray Grant. 2007. Salicylic acid in plant defence—the players and protagonists. Current Opinion in Plant Biology. 10, 466–472.

Gebhardt C, Valkonen JP. 2001. Organization of genes controlling disease resistance in the potato genome. Annual Review of Phytopathology 39: 79–102.

Glazebrook, J. 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annual Review of Phytopathology. 43: 205–227.

Gohre, V., T. Spallek, H. Haweker, S. Mersmann, T. Mentzel, T. Boller, M. de Torres, J.W. Mansfield and S. Robatzek. 2008. Plant pattern-recognition receptor FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB. Current Biology. 18: 1824-32.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution 11: 725–736.

Gomez-Gomez, L., Boller T.. 2000. "FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis". Molecular Cell 5: 1003–1011.

Grube RC, Radwanski ER, Jahn M. 2000. Comparative genetics of disease resistance within the solanaceae. Genetics 155: 873–887.

Gu Z, Cavalcanti A, Chen F-C, Bouman P, Li WH. 2002. Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. Molecular Biology and Evolution 19: 256–262.

Guilfoyle, T.J., and Hagen, G. 2007. Auxin response factors Curr Opin Plant Biol. 10: 453-60.

Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D. 2011. Genomewide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in Arabidopsis. Plant Physiology 157: 757–769.

Hagen, G., and Guilfoyle, T. 2002. Auxin-responsive gene expression: genes, promoters and regulatory factors. Plant Mol Biol. 49: 373-85.

Hammond-Kosack, K.E and Rudd, J.J. 2008. Plant resistance signalling hijacked by a necrotrophic fungal pathogen Plant Signal Behav. 3: 993–995.

Handford, M.G., and Carr J.P. 2007. A defect in carbohydrate metabolism ameliorates symptom severity in virus-infected Arabidopsis thaliana. J Gen Virol. 88: 337-41.

Hanson PM, Bernacchi D, Green S, Tanksley SD, Muniyappa V, Padmaja AS, Chen H, Kuo G, Fang D, Chen J. 2000. Mapping a wild tomato introgression associated with Tomato yellow leaf curl virus resistance in a cultivated tomato line. J Am Soc Hort Sci 125:15–20

He, M., He, C.Q., and Ding, N.Z. 2012. Natural recombination between tobacco and tomato mosaic viruses.Virus Res. 163: 374-379.

Heidrich, K., S. Blanvillain-Baufume and J.E. Parker. 2012. Molecular and spatial constraints on NB-LRR receptor signaling. Current opinion in plant biology. 15: 385-91.

Herbers, K., Meuwly, P., Métraux, J.P., and Sonnewald, U. 1996 Salicylic acid-independent induction of pathogenesis-related protein transcripts by sugars is dependent on leaf developmental stage. FEBS Lett. 397: 239-44.

Herbers, K., Takahata, Y., Melzer, M., Mock, H.P., Hajirezaei, M., and Sonnewald, U. 2000 Regulation of carbohydrate partitioning during the interaction of potato virus Y with tobacco. Mol Plant Pathol. 1: 51-59.

Herms, D.A, and Mattson, W.J. 1992. The dilemma of plants: to grow or defend. Quarterly Review of Biology. 67: 283-335.

Holub EB. 2001. The arms race is ancient history in Arabidopsis, the wildflower. Nature Reviews Genetics 2: 516–527.

Horsfall, J. G. and Diamond, A. E., 1957. The diseased plant in plant pathology - An advanced treaties. Academics press New York. 5: 1-17.

Huang S, van der Vossen EA, Kuang H, Vleeshouwers VG, Zhang N, Borm TJ, van Eck HJ, Baker B, Jacobsen E, Visser RG. 2005. Comparative genomics enabled the isolation of the R3a late blight resistance gene in potato. Plant Journal 42: 251–261.

Hulbert SH, Webb CA, Smith SM, Sun Q. 2001. Resistance gene complexes: evolution and utilization. Annual Review of Phytopathology 39: 285–312.

Iwasaki, A. and R. Medzhitov. 2010. Regulation of Adaptive Immunity by the Innate Immune System. Science. 327: 291-295.

Jia, Y., S.A. McAdams, G.T. Bryan, H.P. Hershey and B. Valent. 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. The EMBO Journal. 19: 4004-14.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97–100.

Jones, J.D.G. and J.L. Dangl. 2006. The plant immune system. Nature. 444: 323-329.

Jupe F, Pritchard L, Etherington GJ, Mackenzie K, Cock PJ, Wright F, Kumar Sharma S, Bolser D, Bryan GJ, Jones JD, Hein I. 2012. Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics, 13:75.

Jupe F, Witek K, Verweij W, Sliwka J, Leighton P, Etherington GJ, Maclean D, Cock PJ, Leggett RM, Bryan GJ, Milne L, Hein I, Jones JDG. 2013: Resistance gene enrichment sequencing (RenSeq) enables re-annotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. Plant J. 76, 530-540.

Kaloshian I, Yaghoobi J, Liharska T, Hontelez J, Hanson D, Hogan P, Jesse T, Wijbrandi J, Simons G, Vos P, Zabel P, Williamson VM. 1998. Genetic and physical localization of the root-knot nematode resistance locus Mi in tomato. Mol Gen Genet 257:376–385

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30: 3059–3066.

Kay, S., S. Hahn, E. Marois, G. Hause and U. Bonas. 2007. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science. 318: 648-51.

Kobayashi, M., Yamamoto-Katou, A., Katou, S., Hirai, K., Meshi, T., Ohashi, Y., and Mitsuhara, I. 2011. Identification of an amino acid residue required for differential recognition of a viral movement protein by the Tomato mosaic virus resistance gene Tm-2(2). J Plant Physiol. 168: 1142-5

Koornneef, A., Verhage, A., Leon-Reyes, A., Snetselaar, R., Van Loon, L., and Pieterse, C.M. 2008. Towards a reporter system to identify regulators of cross-talk between salicylate and jasmonate signaling pathways in Arabidopsis. Plant Signal Behav. 3: 543-546.

Kumar, H., T. Kawai and S. Akira. 2011. Pathogen recognition by the innate immune system. International Reviews of Immunology. 30: 16-34.

Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, Tung J, Sun H, Kumar P, Baker B. 2012. MicroRNA regulation of plant innate immune receptors. Proc Nat Acad Sci. 109.

Li J, Ding J, Zhang W, Zhang Y, Tang P, Chen JQ, Tian D, Yang S. 2010. Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. Molecular Genetics and Genomics 283: 427–438.

Li X., Van Eck HJ, Rouppe van der Voort JNAM, Huigen DJ, Stam P, Jacobsen E. 1998. Autotetraploids and genetic mapping using common AFLP markers: The R2 allele conferring resistance to Phytophthora infestans mapped on potato chromosome 4: Theor. Appl Genet, 96:1121-1128.

Li, J., Zhang, W., Wu, H., Guo, T., Liu, X., Wan, X., Jin, J., Hanh, T.T., Thoa, N.T., Chen, M., Liu, S., Chen, L., Liu, X., Wang, J., Zhai, H., and Wan, J. 2011. Fine mapping of stable

QTLs related to eating quality in rice (Oryza sativa L.) by CSSLs harboring small target chromosomal segments. Breed Sci. 61: 338-46.

Liu J, Liu X, Dai L, Wang G. 2007. Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. Journal of Genetics and Genomics 34: 765–776.

Liu, J., J.M. Elmore and G. Coaker. 2009. Investigating the functions of the RIN4 protein complex during plant innate immune responses. Plant Signaling & Behavior. 4: 1107-10.

Llorente, F., Muskett, P., Sánchez-Vallet, A., López, G., Ramos, B., Sánchez-Rodríguez, C., Jordá, L., Parker, J., and Molina, A. 2008. Repression of the auxin response pathway increases Arabidopsis susceptibility to necrotrophic fungi. Mol Plant. 1: 496-509.

Lodha, T.D., and Basak, J. 2012. Plant-pathogen interactions: what microarray tells about it ?. Mol Biotechnol. 50: 87-97. 2.

López, M.A., Bannenberg, G., and Castresana, C. 2008. Controlling hormone signaling is a plant and pathogen challenge for growth and survival. Curr Opin Plant Biol. 11: 420-427

Lukasik, E., and Takken, F.L. 2009. .Standing strong, resistance proteins instigators of plant defence. Curr Opin Plant Biol. 12: 427-436.

Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. Science 252: 1162–1164.

Mackey, D., B.F. Holt, 3rd, A. Wiig and J.L. Dangl. 2002. RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. Cell,. 108: 743-54.

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262: 1432–1436.

Martin GB, Williams JG, Tanksley SD. 1991. Rapid identification of markers linked to a Pseudomonas resistance gene in tomato by using random primers and near-isogenic lines. Proc Natl Acad Sci USA 88:2336–2340

Mazourek M, Cirulli ET, Collier SM, Landry LG, Kang BC, Quirin EA, Bradeen JM, Moffett P, Jahn MM. 2009. The fractionated orthology of Bs2 and Rx/Gpa2 supports shared synteny of disease resistance in the Solanaceae. Genetics 182:1351–1364

McDowell JM, Simon SA. 2008. Molecular diversity at the plant-pathogen interface. Dev Comp Immunol 32:736–744 McHale L, Tan X, Koehl P, Michelmore RW(2006) Plant NBSLRR proteins: adaptable guards. Genome Biol 7:212

McHale L, Tan X, Koehl P, Michelmore RW. 2006. Plant NBS-LRR proteins: adaptable guards. Genome Biology 7: 212.

Meyers BC, Kaushik S, Nandety RS. 2005. Evolving disease resistance genes. Curr Opin Plant Biol 8: 129–134.

Meyers, B.C., A. Kozik, A. Griego, H. Kuang and R.W. Michelmore. 2003. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. The Plant cell. 15: 809-34.

Michelmore RW, Meyers BC. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research 8: 1113–1130.

Mishina, T.E. and J. Zeier. 2007. Pathogen-associated molecular pattern recognition rather than development of tissue necrosis contributes to bacterial induction of systemic acquired resistance in Arabidopsis. The Plant Journal. 50: 500-13.

Mithöfer, A., and Boland, W. 2012. Plant defense against herbivores: chemical aspects. Annu Rev Plant Biol. 63: 431-50.

Miyakawa T, Miyazono K, Sawano Y, Hatano K, Tanokura M. 2009. Crystal structure of ginkbilobin-2 with homology to the extracellular domain of plant cysteine-rich receptorlike kinases. Proteins 77:247–251

Monaghan, J. and C. Zipfel. 2012. Plant pattern recognition receptor complexes at the plasma membrane. Current Opinion in Plant Biology. 15: 349-57.

Mukhtar, M.S., Carvunis, A.R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M.T., Pevzner, S.J., Donovan, S.E., Ghamsari, L., Santhanam, B., Romero, V., Poulin, M.M., Gebreab, F., Gutierrez, B.J., Tam, S., Monachello, D., Boxem, M., Harbort, C.J., McDonald, N., Gai, L., Chen, H., He, Y.. 2011. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science. 33: 596-601

Mur, L.A., Kenton, R. Atzorn, O. Miersch and C. Wasternack. 2006. The outcomes of concentration-specific interactions between salicylate and jasmonate signaling include synergy, antagonism, and oxidative stress leading to cell death. Plant Physiology. 140: 249-62.

Murphy WJ, Pevzner PA, O'Brien SJ. 2004. Mammalian phylogenomics comes of age. Trends in Genetics 20: 631–639.

Nei M, Kumar S. 2000 Molecular evolution and phylogenetics. New York, NY, USA: Oxford University Press.

Nimchuk Z, Marois E, Kjemtrup S, Leister RT, Katagiri F, Dangl JL. 2000. Eukaryotic fatty acylation drives plasma membrane targeting and enhances function of several type III effector proteins from Pseudomonas syringae. Cell 101: 353–363.

Nobuta K, Ashfield T, Kim S, Innes RW. 2005. Diversification of non-TIR class NB-LRR genes in relation to whole-genome duplication events in Arabidopsis. Molecular Plant-Microbe Interactions 18: 103–109.

Okayama H and Berg P. 1982, High-efficiency cloning of full-length cDNA. Molecular and cellular biology. 2:161-170.

Ori, N., Eshed, Y., Paran, I., Presting, G., Aviv, D., Tanksley, S., Zamir, D., and Fluhr, R. 1997. The I2C family from the wilt disease resistance locus I2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. Plant Cell. 9: 521-32.

Pan Q, Liu YS, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamir D, Fluhr R. 2000. Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and Arabidopsis. Genetics 155: 309–322.

Pan Q, Liu YS, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamir D, Fluhr R. 2000. Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and arabidopsis. Genetics 155:309–322

Pan Q, Wendel J, Fluhr R.. 2000. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol. 50:203–213.

Park T, Gros J, Sikkema A, Vleeshouwers VGAA, Muskens M, Allefs S, Jacobsen E, Visser RGF, van der Vossen EAG. 2005. The Late Blight Resistance Locus Rpi-blb3 from Solanum bulbocastanum Belongs to a Major Late Blight R Gene Cluster on Chromosome 4 of Potato: Mol. Plant-Microbe Interact. 18: 722–729.

Paul, J.Y., Becker, D.K., Dickman, M.B., Harding, R.M., Khanna, H.K., and Dale, J.L. 2011. Apoptosis-related genes confer resistance to Fusarium wilt in transgenic 'Lady Finger' bananas. Plant Biotechnol J. 9: 1141-1148.

Pedley KF, Martin GB. 2003. Molecular basis of Pto-mediated resistance to bacterial speck disease in tomato. Annu Rev Phytol Pathol 41:215–243

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. Nucleic Acids Research 33: W116–W120.

Quirin EA, Mann H, Meyer RS, Traini A, Chiusano ML, Litt A, Bradeen JM. 2012. Evolutionary meta-analysis of solanaceous resistance gene and Solanum resistance gene analog sequences and a practical framework for cross-species comparisons. Molecular Plant-Microbe Interactions 25: 603–612.

R Core Team 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reis, A., Giordano, L.B., Lopes, C.A., and Boiteux, L.S. 2004. Novel sources of multiple resistance to three races of Fusarium oxysporum f. sp. lycopersici in Lycopersicon germplasm. Crop Breeding and Applied Biotechnology. 4: 495-502.

Riely B, Martin G. 2001. Ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene Pto. Proc Natl Acad Sci USA 98:2059–2064

Robatzek S, Bittel P, Chinchilla D, Kochner P, Felix G, Shiu SH, Boller T. 2007. Molecular identification and characterization of the tomato flagellin receptor LeFLS2, an orthologue of Arabidopsis FLS2 exhibiting characteristically different perception specificities. Plant Molecular Biology 64: 539–547.

Robert-Seilaniantz, A., MacLean, D., Jikumaru, Y., Hill, L., Yamaguchi, S., Kamiya, Y., Jones, J. D. G. 2011. The microRNA miR393 re-directs secondary metabolite biosynthesis away from camalexin and towards glucosinolates. Plant J. 67: 218–231.

Salmeron JM, Oldroyd GE, Rommens CM, Scofield SR, Kim HS, Lavelle DT, Dahlbeck D, Staskawicz BJ. 1996. Tomato Prf is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the Pto kinase gene cluster. Cell 86: 123–133.

Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, Frusciante L, Ercolano MR. 2010. PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Research 38: D814–D821.

Sawano Y, Miyakawa T, Yamazaki H, Tanokura M, Hatano K. 2007. Purification, characterization, and molecular gene cloning of an antifungal protein from Ginkgo biloba seeds. Biol Chem 388:273–280

Schenk, P.M., K. Kazan, I. Wilson, J.P. Anderson, T. Richmond, S.C. Somerville and J.M. Manners. 2000. Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proceedings of the National Academy of Sciences of the United States of America. 97: 11655-60.

Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences, USA 95: 5857–5864.

Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. Trends in Genetics 20: 461–464.

Shah J. 2003. The salicylic acid loop in plant defense. Curr Opin Plant Biol , 6:365-371.

Shen KA, Chin DB, Arroyo-Garcia R, Ochoa OE, Lavelle DO, Wroblewski T, Meyers BC, Michelmore RW. 2002. Dm3 is one member of a large constitutively expressed family of nucleotide binding site-leucine-rich repeat encoding genes. Molecular Plant-Microbe Interactions. 15:251-261.

Shiu, S.-H. and A.B. Bleecker. 2001. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proceedings of the National Academy of Sciences of the United States of America. 98: 10763-10768.

Shivaprasad, P.V., Chen, H.M., Patel, K., Bond, D.M., Santos, B.A., and Baulcombe, D.C. 2012 A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell. 24: 859-874.

Sidhu D, Gill KS. 2005. Distribution of genes and recombination in wheat and other eukaryotes. Plant Cell, Tissue and Organ Culture 79: 257–270.

Simons, G., Groenendijk, J., Wijbrandi, J., Reijans, M., Groenen, J., Diergaarde, P., Van der Lee, T., Bleeker, M., Onstenk, J., De Both, M., Haring, M., Mes, J., Cornelissen, B., Zabeau, M. and Vos, P. 1998. Dissection of the Fusarium I2 gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell. 10:1055–1068.

Singer GA, Lloyd AT, Huminiecki LB, Wolfe KH. 2005. Clusters of coexpressed genes in mammalian genomes are conserved by natural selection. Molecular Biology and Evolution 22: 767–775.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 3 Article 3.

Smyth, G.K. 2005 Limma: linear models for microarray data. In: 'Bioinformatics and Computational Biology Solutions using R and Bioconductor'. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds), Springer, New York, pages 397-420.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. American Journal of Botany 96: 336–348.

Song, W.Y. et al.. 1995. "A receptor kinase-like protein encoded by the rice disease resistance gene, XA21". Science 270: 1804–1806.

Spoel, S.H. and Dong, X. 2012. How do plants achieve immunity? Defence without specialized immune cells. Nature Reviews Immunology. 12: 89-100.

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. 1999. Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. Nature 400: 667–671.

Stevens MR, Lamb EM, Rhoads DD. 1995. Mapping the Sw-5 locus for tomato spotted wilt virus resistance in tomatoes using RAPD and RFLP analyses. Theor Appl Genet 90: 451–456

Stoye J, Wittler R. 2009. A unified approach for reconstructing ancient gene clusters. IEEE/ACM Transactions on Computational Biology and Bioinformatics 6: 387–400.

Takken FLW, Tameling WIL. 2009. To nibble at plant resistance proteins. Science 324:744–746.

Takken, F., and Rep, M. 2010. The arms race between tomato and Fusarium oxysporum. Mol Plant Pathol. 11: 309-14.

Tamura K, Kumar S. 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Molecular Biology and Evolution 19: 1727–1736.

Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences, USA 101: 11030–11035.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28: 2731–2739.

Tang PZY, Sun X, Tian S, Yang S, Ding J. 2010. Disease resistance signature of the leucine-rich repeat receptor-like kinase genes in fourplant species. Plant Science 179: 399–406.

Tang X, Xie M, Kim YJ, Zhou J, Klessig DF, Martin GB. 1999. Overexpression of Pto activates defense responses and confers broad resistance. Plant Cell 11:15–29

The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485: 635–641.

Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K, Jones JD. 1997. Characterization of the tomato Cf-4 gene for resistance to Cladosporium fulvum identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. Plant Cell 9: 2209–2224.

Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 22, 4673–4680.

Ton, J., Van Pelt, J.A., Van Loon, L.C. & Pieterse, C.M.J. 2002. Differential effectiveness of salicylate-dependent and jasmonate/ethylene-dependent induced resistance in Arabidopsis. Mol. Plant Microbe Interact. 15, 27–34.

Trusov, Y., Rookes, J.E., Chakravorty, D., Armour, D., Schenk, P.M., and Botella, J.R. 2006. Heterotrimeric G proteins facilitate Arabidopsis resistance to necrotrophic pathogens and are involved in jasmonate signaling. Plant Physiol. 140: 210-220.

Tsuda, K., Sato, M., Glazebrook, J., Cohen, J.D. & Katagiri, F. 2008. Interplay between MAMPtriggered and SA-mediated defense responses. Plant J. 53, 763–775.

Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J., and Katagiri, F. 2009. Network Properties of Robust Immunity in Plants. PLoS Genet 5: e1000772.

van der Hoorn, R.A. and S. Kamoun. 2008. From Guard to Decoy: a new model for perception of plant pathogen effectors. The Plant cell. 20: 2009-17.

Van Oosten, V.R. et al. 2008. Differential effectiveness of microbially induced resistance against herbivorous insects in Arabidopsis. Mol. Plant Microbe Interact. 21, 919–930.

Vandenhaute, J., Roth, F.P., Hill, D.E., Ecker, J.R., Vidal, M., Beynon, J., Braun, P., and Dangl, J.L. 2011. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science. 333: 596-601.

Vlot, A.C., Klessig, D.F. & Park, S.-W. 2008. Systemic acquired resistance: the elusive signal(s). Curr. Opin. Plant Biol. 11, 436–442.

Wang G, Ellendorff U, Kemp B, Mansfield JW, Forsyth A, Mitchell K, Bastas K, Liu CM, Woods-Tor A, Zipfel C et al. 2008a. A genome-wide functional investigation into the roles of receptor-like proteins in Arabidopsis. Plant Physiology 147: 503–517.

Wang Y, Diehl A, Wu FN, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD. 2008b. Sequencing and comparative analysis of a conserved syntenic segment in the solanaceae.Genetics 180: 391–408.

Ward, J.L., S. Forcat, M. Beckmann, M. Bennett, S.J. Miller, J.M. Baker, N.D. Hawkins, C.P. Vermeer, C. Lu, W. Lin, et al.. 2010. The metabolic transition during disease following infection of Arabidopsis thaliana by Pseudomonas syringae pv. tomato. The Plant Journal. 63: 443-457.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular Biology and Evolution 18: 691–699.

Williams, S.J., Sornaraj, E. deCourcy-Ireland, R.I. Menz, B. Kobe, J.G. Ellis, P.N. Dodds and P.A. Anderson. 2011. An autoactive mutant of the M flax rust resistance protein has a preference for binding ATP, whereas wild-type M protein binds ADP. Molecular Plant-Microbe Interactions. 24: 897-906.

Xiang, T., Zhong, N., Zou, Y., Wu, Y., Zhang, J., Xing, W., Li, Y., Tang, X., Zhu, L., Chai, J., and Zhou, J.-M.. 2008. Pseudomonas syringae effector AvrPto blocks innate immunity by targeting receptor kinases. Curr. Biol. 18: 74–80.

Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M, Turner JG. 2001. Broad-spectrum mildew resistance in Arabidopsis thaliana mediated by RPW8. Science 291:118–120

Xiaoxun N, Jingjing Y, Silong S, Wencai Y. 2013. Identification and analysis of resistance-like genes in the tomato genome. Jurnal of phytopathology, 117/199.

Xing, W., et al. 2007. The structural basis for activation of plant immunity by bacterial effector protein AvrPto. Nature 449: 243–247.

Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J et al. 2011. Genome sequence and analysis of the tuber crop potato. Nature 475: 189–195.

Yang S, Zhang X, Yue JX, Tian D, Chen JQ. 2008. Recent duplications dominate NBS-encoding gene expansion in two woody species. Molecular Genetics and Genomics 280: 187–198.

Yi G, Sze SH, Thon MR. 2007. Identifying clusters of functionally related genes in genomes. Bioinformatics 23: 1053–1060.

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W,HuS, ZengCet al. 2005. The Genomes of Oryza sativa: a history of duplications. PLoS Biology 3: e38.

Zangerl, A.R., and Berenbaum, M.R. 1997. Cost of chemically defending seeds: furanocoumarins and Pastinaca sativa. Am Nat. 150 (4): 491-504.

Zangerl, A.R., and Berenbaum, M.R. 2003. Phenotype matching in wild parsnip and parsnip webworms: causes and consequences. Evolution. 57: 806-15.

Zhang C, Liu L, Zheng Z, Sun Y, Zhou L, Yang Y, Cheng F, Zhang Z, Wang X, Huang S, Xie B, Du Y, Bai Y, Li J. 2013. Fine mapping of the Ph‐3 gene conferring resistance to late blight (Phytophthora infestans) in tomato. Theor Appl Genet, 126:2643–2653.

Zhang M, Wu YH, Lee MK, Liu YH, Rong Y, Santos TS, Wu C, Xie F, Nelson RL, Zhang HB. 2010. Numbers of genes in the NBS and RLK families vary by more than four-fold within a plant species and are regulated by multiple factors. Nucleic Acids Research 38: 6513–6525.

Zhang, J., W. Li, T. Xiang, Z. Liu, K. Laluk, X. Ding, Y. Zou, M. Gao, X. Zhang, S. Chen, et al.. 2010. Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a Pseudomonas syringae effector. Cell Host & Microbe. 7: 290-301.

Zhang, Z., Y. Wu, M. Gao, J. Zhang, Q. Kong, Y. Liu, H. Ba, J. Zhou and Y. Zhang. 2012. Disruption of PAMP-Induced MAP Kinase Cascade by a Pseudomonas syringae Effector Activates Plant 41 Immunity Mediated by the NB-LRR Protein SUMM2. Cell Host & Microbe. 11: 253-263.

Zhou, H., J. Lin, A. Johnson, Robyn L. Morgan, W. Zhong and W. Ma. 2011. Pseudomonas syringae Type III Effector HopZ1 Targets a Host Enzyme to Suppress Isoflavone Biosynthesis and Promote Infection in Soybean. Cell Host & Microbe. 9: 177-186.

Zipfel, C. and J.P. Rathjen. 2008. Plant immunity: AvrPto targets the frontline. Current Biology. 18: R218-20.