A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Molecular genetic analysis of *Tobamovirus* resistance in *Capsicum* spp.

고추 토바모바이러스 저항성의 분자생물학적 분석

February, 2014

Hee-Bum Yang

MAJOR IN HORTICULTURAL SCIENCE

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Molecular genetic analysis of *Tobamovirus* resistance in *Capsicum*

UNDER THE DIRECTION OF DR. BYOUNG-CHEORL KANG
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF
SEOUL NATIONAL UNIVERSITY

BY
HEEBUM YANG

MAJOR IN HORTICULTURAL SCIENCE
DEPARTMENT OF PLANT SCIENCE

February, 2014

APPROVED AS A QUALIFIED DISSERTATION OF HEEBUM YANG
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS

CHAIRMAN _____
Doil Choi, Ph.D.

VICE-CHAIRMAN _____
Byoung-Cheorl Kang, Ph.D.

MEMBER _____
Chang-Sik Oh, Ph.D.

MEMBER _____
Tae-Jin Yang, Ph.D.

MEMBER _____
Hwa Jin Cho, Ph.D.

Molecular genetic analysis of *Tobamovirus* resistance in *Capsicum* spp.

Hee-Bum Yang

Department of Plant Science, Seoul National University

ABSTRACT

Tobamovirus is one of the most destructive viruses in *Capsicum.* Because it is hard to control *Tobamovirus* by chemical agents or sanitation, development of resistant cultivars would be the best option to manage *Tobamovirus*. Molecular genetic analysis provides deep understanding of *Tobamovirus* resistance, which leads usefully application to *Tobamovirus* resistance breeding.

In the first chapter, comparative genetic relationship between *I2* of tomato, *R3* of potato, and *L* of pepper were investigated and bacterial artificial chromosome (BAC) clones corresponding to *I2C-1* and *R3/7* were isolated. Selected BAC clones were sorted by PCR screening, sequencing, and genetic mapping. The BAC clone 082F03 of one sorted group was anchored near TG36. Additional BAC clones were isolated using 082F03 sequence and 224kb contig was constructed. Three markers developed based on contig sequence were closely linked to the L^4 gene (about 1.2cM). The linkage analysis of that marker co-segregated with the L^3 gene was performed in L^4-segregating populations. The result suggests the possibility that L^3 and L^4 may be different genes that are closely linked to each other instead of different alleles in same locus.

In the second chapter, the L^4 gene candidate isolated by homology based PCR using the L^3 gene was validated by linkage analysis of L4segF&R marker developed based on LRR sequence of L^4 gene candidate. L4segF&R was closely linked to the L^4 gene (0.3cM), however it did not co-segregate perfectly in L^4-segregating population. Linkage analysis of L4segF&R of breeding materials provided by three seeds companies was performed to confirm the mapping result and several recombinants were also found. A number of genetic architecture models were postulated based on recombinants data. They demonstrated that *Tobamovirus* resistance is not fully explained by L^4 gene candidate alone. Finally allele specific HRM marker, L4RP-3F/L4RP-3R, was developed based on LRR sequence of the L^4 gene candidate sequence.

In the third chapter, *de novo* assembly of *L*-homologous transcripts using transcriptome sequence of *C. chacoense* 'PI260429' containing L^4 was performed. Two indexes including the number of contigs and maximum contig length, were designed to validate *de novo* assembly products. Transcripts assembled by Velvet with single k-mer = 59 showed the highest score in both indices, and these were chosen for next analysis. A total of 96 *L*-homologous contigs were isolated and digital expression analysis was performed between four *Tobamovirus* resistant accessions and four susceptible accessions. Eight contigs were expressed only in resistance accessions. These are derived from four transcripts, the L^4 candidate, PIX-4, CA00g34020_CNL_G4_C3 and *L*_homologue 97_98. This indicates the other three transcripts except the L^4 candidate which is known to be responsible for *Tobamovirus* resistance may also play an important role for *Tobamovirus* resistance.

The results of this research are expected to provide molecular genetic understanding of *Tobamovirus* resistance in *Capsicum* spp. It is expected that this study will consequently contribute to developing molecular breeding system for

Tobamovirus resistance.

Keywords: *Capsicum*, *Tobamovirus*, disease resistance, molecular marker, *de novo* assembly

Student Number: 2007-23120

CONTENTS

CHAPTER III

De novo Assembly and Digital Expression Profiling of the *L* Homologs in *Capsicum* spp.

LIST OF TABLES

CHAPTER I

CHAPTER II

CHAPTER III

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

LIST OF ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
BAC	Bacterial Artificial Chromosome
BSA	Bulked Segregant Analysis
CAPS	Cleaved Amplified Polymorphic Sequence
CC-NBS-LRR	Coiled Coil-Nucleotide Binding Site-Leucine Rich Repeat
CMV	*Cucumber Mosaic Virus*
DH	Doubled Haploid
HRM	High Resolution Melting
HR	Hypersensitive Response
MAS	Marker Assisted Selection
MAB	Marker Assisted Backcrossing
NILs	Near Isogenic Lines
NGS	Next Generation Sequencing
ORFs	Open Reading Frames
PaMMV	*Paprika Mild Mottle Virus*
PMMoV	*Pepper Mild Mottle Virus*
QTL	Quantitative Trait Loci

RAPD	Random Amplified Polymorphic DNA
RLK	Receptor Like Kinase
RLP	Receptor Like Protein
RGA	Resistance Gene Analog
RFLP	Restriction Fragment Length Polymorphism
RT-PCR	Reverse Transcriptase PCR
SNP	Single Nucleotide Polymorphism
TMV	*Tobacco Mosaic Virus*
TIR	Toll/Interleukin-1 Receptor
ToMV	*Tomato Mosaic Virus*
TSWV	*Tomato Spotted Wilt Virus*
TYLCV	*Tomato Yellow Leaf Curl Virus*
LTRs	Long Terminal Repeats

GENERAL INTRODUCTION

Tobamovirus is one of the most devastating virus genus. A total of 31 virus species were reported so far, and among them , *Tobacco mosaic virus* (TMV), *Tomato mosaic virus* (ToMV), *Paprika mild mottle virus* (PaMMV) and *Pepper mild mottle virus* (PMMoV) were major *Tobamovirus* species found in *Capsicum* (Boukema 1980; Mizumoto et al. 2012; Tomita et al. 2011). Peppers infected by *Tobamovirus* show symptoms such as mild chlorosis, stunting, stem necrosis and light green mottle in both vegetative and reproductive tissues. *Tobamovirus* causes serious damage on yield and quality worldwide. Virus is barely controlled by chemical reagents, and sanitation methods entailing hard labor is not enough to prevent *Tobamovirus* infection. On the other hands, breeding disease resistant cultivars would be the best option to control *Tobamovirus*.

Tobamovirus resistance is known to be inherited by a single dominant resistance locus *L*. A total of four alleles, L^1, L^2, L^3, and L^4, were found in *C. annuum, C. frutescens, C. chinense,* and *C. chacoense* by Boukema (1980), and additionally temperature dependent alleles, L^{1a}, L^{1c}, and L^{2b} were also reported (Tomita et al. 2011). These alleles were classified by interaction with

Tobamovirus pathotypes P_0, P_1, $P_{1.2}$ and $P_{1.2.3}$ (Berzal-Herranz et al. 1995; Boukema 1980; de la Cruz et al. 1997; Tomita et al. 2011). The L^4 locus conferred resistance against all kinds of *Tobamovirus* pathotypes until the appearance of new L^4-breaking pathotype $P_{1.2.3.4}$ (Genda et al. 2007). Fortunately, this pathotype found only in Japan and the L^4 gene is still a valuable resistance gene for commercial breeding of sweet and hot pepper.

At first, the *L* locus was mapped on the end of long arm of chromosome 11 (Lefebvre et al.1995). Since then, a number of *L*-linked markers have been developed using random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and resistance gene analog (RGA) mapping strategies (Kim et al. 2008; Matsunaga et al. 2003; Tomita et al.2008). High resolution linkage map of the *L* locus was constructed and then the L^3 gene was isolated (Tomita et al. 2011). Other *L* allele candidates were isolated by homology based PCR strategy. *In vivo* co-infiltration analysis indicates that these candidates were interacted with corresponding *Tobamovirus* pathotypes in *Nicotiana benthamiana* system (Tomita et al. 2011). However genetic analysis of *L* candidates was not confirmed yet.

Comparative analysis of resistance genes in three Solanaceae crops, *Lycopersicum* (tomato), *Solanum* (potato) and *Capsicum* (pepper), revealed that resistance genes are not distributed randomly but clustered in the syntenic region

(Grube et al. 2000). Three resistance genes, *I2* (resistant to *Fusarium oxysporum* f. sp lycopersici), *R3a* (resistant to *Phytophthora infestans*), and *L* (resistant to *Tobamovirus*), were located in the corresponding genetic region of *Lycopersicum*, *Solanum* and *Capsicum*, respectively. *R3a* was isolated using a comparative genomic strategy (Huang et al.2004, 2005). *R3a*, *I2* and L^3 belong to the coiled coil-nucleotide binding site-leucine rich repeat (CC-NBS-LRR) class, which is one of the major resistance gene structures in plant (Huang et al. 2005; Ori et al. 1997; Tomita et al. 2011).

NBS-LRR includes highly repetitive homologs in a cluster (Meyers et al. 1999; Michelmore and Meyers 1998). A number of homologs were reported in the flanking region of *R3a*, *I2* and L^3 (Huang et al. 2005; Ori et al. 1997; Tomita et al. 2011; Yang et al. 2012). Function of these resistance gene analogs (RGAs) is not figured out yet. Transcriptome sequencing using NGS technologies and bioinformatics programs provide tools to isolate RGAs and analyze digital expression level of RGAs in a single assay. However research on RGAs using NGS techniques have limitations because current commercially available NGS platforms generate short read sequences less than 400 bp.

In this study, molecular markers linked to the *L* locus were developed by investigating comparative relationship among *I2* of *Lycopersicum*, *R3/7* of *Solanum*, and *L* of *Capsicum*. Then genetic architectures of the *L* locus were

elucidated by linkage analysis of the L^4 candidate. The result demonstrated that though the L^4 candidate is a L^4 functional homolog, another genetic factor is required to explain *Tobamovirus* resistance. To find another genetic factor responsible for *Tobamovirus*, *de novo* assembly and digital expression of *L*-homologous transcripts in PI260429 was performed in this study.

LITERATURE REVIEW

1. *Tobamovirus*

Tobamovirus is a rod-shaped type IV RNA virus found worldwide. *Tobacco mosaic virus* (TMV), a type species of *Tobamovirus*, consists of RNA genome and about 2,140 of coat proteins and constitute long wound cylindrical structure. This genus belongs to Virgaviridae family and thirty one species and several tentative species have been reported to date (Kim et al. 2012, Rhie et al. 2007). *Tobamovirus* spp. infect plants species of Apocynaceae, Brassicaceae, Cactaceae, Cucurbitaceae, Fabaceae, Gesneriaceae, Malvaceae, Orchidaceae, Passifloraceae, and Solanaceae family (Rhie et al. 2007). Phylogenetic analysis based on sequence of 126-KDa replicase and 53-KDa polymerase showed that each subgroup of *Tobamovirus* largely correlated with *Tobamovirus* groups classified by their own host family. This supports hypothesis that *Tobamovirus* have coevolved and codiverged from a common ancestor with their host plants (Gibbs et al. 1999; Rhie et al. 2007, Stobbe et al. 2012).

Tobamovirus is one of most devastating virus group in *Capsicum* species including hot pepper, sweet pepper, and ornamental pepper grown in both of field and plastic- or glasshouse. Four *Tobamovirus* species, *Tobamovirus* species,

Tobacco mosaic virus (TMV), *Tomato mosaic virus* (ToMV), *Paprika mild mottle virus* (PaMMV) and *Pepper mild mottle virus* (PMMoV), have been reported to be able to infect *Capsicum* plants (Boukema 1980; Mizumoto et al. 2012; Tomita et al. 2011). Mild chlorosis, stunting, stem necrosis and light green mottle symptoms developed in *Capsicum* plants infected by *Tobamovirus*. This cause severe yield and quality loss in pepper production.

1) Open reading frames (ORFs) in *Tobamovirus* genome

Genus *Tobamovirus* contains single strand positive sense RNA genome. Sequence analysis of total RNA of TMV, which is a type virus of *Tobamovirus* genus, was finished by cDNA synthesis using oligonucleotide primers and shotgun sequencing strategy (Goelet et al. 1982). Total genome size of TMV is 6,395 nucleotides (nt), and it contains five ORFs encoding proteins for replication, movement and assembly activities (Bagley 2001; Dawson 1992; Levy et al. 2013; Lewandowski and Dawson 2000; Tenllando et al. 1995).

126-kDa protein is translated from first ORF located on 69 nt apart from the 5' terminus of TMV RNA genome. 183-kDa protein is translated when read through occurs by suppression of stop codon of 126-kDa with 10% of probability, approximately. 183-kDa replication associated proteins contains methyl-transferase (MT) domain, helicase (HEL)-like and polymerase (POL) domains, while 126-kDa contains only MT and HEL domains. Mutation analysis

expressing only 183-kDa or 126-kDa unveiled functions of these proteins (Lewandowski and Dawson 2000). 183-kDa protein is capable to carry out all of RNA synthesis unilaterally. It can detect the promoters of both of positive- and negative- sense RNAs to produce subgenomic RNAs (sgRNAs). On the other hand, 126-kDa protein is not able to synthesize sg RNAs because of absent of POL domain. However 126-kDa protein supports 183-kDa protein to enhance replication efficiency. 126-kDa protein binds viral RNA and leads to 183-KDa protein to form a heteroduplex structure for the replication with high efficiency.

TMV replication proteins initiate synthesis of RNA genomes including 54k, 30k, and 17.5k sgRNAs (Dawson 2001). 54k subgenomic RNA (sgRNA) is produced by readthrough. However, translation product is not detected *in vivo*. The function and feature of the 54k sgRNA are not unveiled yet. Movement proteins (MP) translated from the 30k sgRNA are required for virus spreading from infected cell into adjacent uninfected cell (cell to cell movement). MP is involved in intra-cellular trafficking of TMV by modification of permeablility of the plasmodesmata (PD) using its ablility of binding RNA genome and altering gate structure of PD (Levy et al. 2013). Recently novel function of MP of *Turmip vein clearing virus* (TVCV), a species of *Tobamovirus* subgroup 3, was reported. MP targets F-actin associated with chromatin and is imported to the host nucleus. It is assumed that MP can modify host cellular physiology for its replication and

movement (Levy et al. 2013).

Lastly, 17.5kb sgRNA encodes coat protein (CP). Primary function of this protein is protection of viral RNA genome by construction of external capsid structure. CP subunits have intrinsic properties to auto-assemble into various aggregated forms, monomers, oligomers and two-layered 20S disks, by carboxylate pair in aqueous solution (Butler 1999; Kegel and van der Schoot, 2006). At the assembly stage of viruses, structure of 20S disks are twisted into "lock-washer" shaped 20S helices. Viral RNA genome is intercalated in the inner radii surface of "lock-washer" 20S aggregate and other "lock-washer" shaped 20S aggregates are stacked bidirectionaly to constitute right-handed extended helix structure finally. CPs are required for both cell to cell and long-distance movements. It is possible to replicate viral genome and to move toward neighboring cells without CPs. However replication efficiency is low. Therefore CPs are essential for stable long-distance movements (Dawson 1992).

2) Infection and management of *Tobamovirus*

Tobamovirus are very stable virus particles and able to retain viability and infectivity for a long time in infected plant tissue debris (Dawson 2001; Naqvi 2004). Therefore, remained infected plant tissues of Solanaceae crops, for example, *Capsicum*, *Solanum*, *Nicotiana* and *Lycopersicum*, and also weeds belonged to Solanacease family, for example *Solanum carolinense* in south

eastern USA, can be the origin of initial inoculation. Any kind of insect vectors which transmit *Tobamovirus* have not been reported so far. However *Tobamovirus* are able to remain on seed coat and be transmit to uninfected young leaf embryo (Naqvi 2004). All these cases are primary infection of *Tobamovirus*. To avoid infection from infected seed coat, virus-free seeds or virus-free seedling should be purchased or seeds should be treated by heat, acid or chemical reagents such as trisodium phosphate. *Tobamovirus* can survive in infected plant debris, however, if debris are degraded completely, it lose infectivity. Therefore, plowing field soil can be very helpful to control *Tobamovirus* (Dawson 2001). Crop rotation is another good option for the management of *Tobamovirus* by preventing introduction source of infection such as infected plant debris. However rotational crops should be other family crop except Solanaceae.

Infected individuals can be new inoculum source of secondary infection and any kinds of contacts with infected plants, usually mediated by worker's handling, can transmit *Tobamovirus* (Naqvi 2004). Infected plants should be eliminated for preventing secondary infection. Tools, equipment, and hands are contaminated by handling such as pruning, stalking and tying due to the contact of the nutrient solution of infected plants. Hence, the sanitation with strong soap, 70% alcohol, trisodium phosphate solution and bleach should be carried out for management of *Tobamovirus*. Milk is known to be very useful material for

preventing infection of *Tobamovirus*. Cleaning of simple personal equipment, for example pruning shears, and hands after handling for each individual is very effective way to control *Tobamovirus* (Dawson 2001).

2. Disease resistance (R) gene

Biotic diseases occurred by fungi, bacteria, viruses, and nematodes are one of main factors to decline the yield and quality of crops along with environmental stress such as drought and flooding. Breeders and geneticist discovered that some disease resistance traits were inherited dominantly or semi-dominantly. A number of resistance genes were isolated from various crops, and considerable biochemical and genetics progresses of R genes, structure, function, recognition of pathogen, evolution, and resistance responses, were accumulated. These knowledge permits understanding of R genes and enable to isolate other unknown resistance genes and to assist disease resistance breeding by application to MAS and producing of transformants resistant to diseases.

1) Structure and function of R proteins

Over fifty R genes were isolated from various crops, and nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4 (NB-ARC), leucinse rich repeat (LRR), coiled coil (CC) and Toll/interleukin-1 receptor (TIR) domains were found by amino acid sequence comparison of R proteins (Martin et al. 2003;

Meyers et al. 2005; Ooijen et al. 2007). NB-ARC is responsible to regulate resistance response signaling by catalyzing the hydrolysis of ATP to ADP. NB-ARC domain consists of subdomains, NB, ARC1, and ACR2 and P-loop, RNBS-A, Walker B, RNBS-B, RNBS-C, RNBS-D and MHD motives. LRR, most common domain in various R proteins, is involved in recognition of pathogen, regulation of R gene-activation, and signal transduction (Padmanabhan et al. 2009). LRR domain comprise 2-48 repeats, and each repeat consists of 24-28 residues with core consensus LxxLxxLxLxxC/Nxx. TIR and CC domains are supposed to be required for receptor signaling and intramolecular interaction, respectivley.

Isolated R genes falled into four classes, TIL-NBS-LRR (TNL), CC-NBS-LRR (CNL), receptor like protein (RLP) and receptor like kinase (RLK), by architecture of domains, mainly except for some R genes such as *Xa21* and *Asc-1* (Ooijen et al. 2007; Padmanabhan et al. 2009). TNL contains TIL, NB-ARC, and LRR domain, and CNL contains CC, NB-ARC, and LRR domains. R genes belong to TNL or CNL are located in cytoplasm. RLP contains extracellular LRR (eLRR) domain only, and RLK contains kinase and eLRR. R genes classified to RLP and RLK classes traverse plasma membrane.

2) Three models for gene-for-gene theory

Classical model explaining resistance gene, “Gene-for-gene” theory was

proposed by research on host *Linum usitatissimum* (flax) and its pathogen *Melampsora lini* (flax rust disease) (Flor 1971; Hammond-Kosack and Jones 1997). The "gene-for-gene" theory suggests concept that disease resistance is induced only in the case when resistance gene (R gene) of host and avirulence gene (Avr gene) of pathogen are incompatible. R genes encode proteins which can recognize Avr proteins, then activate resistance response. Pathogens have evolved to overcome R genes by amino acid residues substitution of Avr proteins, whereas R genes have evolved to recognize mutated Avr proteins (co-evolution). Co-evolution was supported by diversifying selection occurred in R genes and Avr genes of flax (*Lilium usitatissimum*)-flax rust (*Melampsora lini*) system (Dodds and Thrall 2009). Evolution tracking by phylogenetic analysis of R genes (*Pik*) and Avr genes (*AvrPik*) showed the older Avr gene tended to be recognized by a larger group of newer R alleles, which is strong evidence of co-evolution (Kanzaki et al 2012). Avr genes encodes virulence factors which are essential for pathogenicity, for example, coat protein, movement protein, and replicase of virus, effectors secreted from type III secretion system of bacteria, and effector of fungus (Hammond-Kosack and Jones 1997; Ooigen et al. 2007). It is reinforced by fitness penalty, which is loss or decline of pathogenicity often occur in R gene resistance-breaking races of pathogen.

As molecular biological research progressed, "guard hypothesis" was

suggested because direct interaction between R protein and Avr protein (ligand-receptor model) found only in few cases, such as Pita/AvrPita, L/AvrL567, PopP2/RRS-1, and p50 helicase/N (Deslandes et al. 2003; Dodds et al. 2006; Jia et al. 2000; Ueda et al. 2006). Avr proteins, for example effectors, interact with host target proteins (guardee). “Guard hypothesis” implies R proteins is activated by monitoring the state of target protein indirectly (Jones and Dangl 2006; Ooijen et al. 2007). Rin4 is the guardee of AvrB or AvrRpm1/RPM1 and AvrRpt2/RPS2, and PBS1 is the guardee of RPS5/AvrPphB in *Arabidoposis*. “Decoy hypothesis” was postulate based on the research progresses of plant-microbe interaction and evolution of R genes and Avr genes (Keith and Mitchell-Olds 2013; van der Hoorn and Kamoun 2008). “Guardee” a target protein responsible for resistance or susceptibility which is modified by effector protein. “Decoy” is a faking “target” protein, which is specialized for the perception of effector but is not related to pathogenicity of host. “Decoy hypothesis” is supported by R protein/decoy/Avr protein cases in *Lycopersicum*, Pto/Prf/AvrPto and AvrPtoB and Cf-2/Rcr3/Avr2 (Kim et al. 2005; Mackey et al. 2002; Mucyn et al. 2003; Rooney et al. 2005; Shao et al. 2003).

3. Resistance to *Tobamovirus* in *Capsicum* spp.

In spite of strict managements such as crop rotation and sanitation for preventing *Tobamovirus* infection, there is still significant chance that plants can be infected. Furthermore, any kinds of chemical agent to control *Tobamovirus* and infected plants are hardly recovered from *Tobamovirus*. However disease resistance give a perfect protection to plants. Therefore utilization of disease resistance is very useful to control *Tobamovirus* in pepper production.

1) Resistance locus *L*

L is a single dominant gene conferring *Tobamovirus* resistance in *Capsicum* spp.. *Capsicum* plants containing the *L* gene showed hypersensitive response (HR) to induce cell death in not only infected cell but also neighboring cells, forming necrotic local lesion, finally. Four *L* alleles, L^1, L^2, L^3 and L^4 alleles are found in *C. annuum*, *C. frutescense*, *C. chinense*, and *C. chacoense*, respectively. These alleles are classified by responses against *Tobamovirus* stains, which are designated to race P_0, P_1, $P_{1.2}$ and $P_{1.2.3}$. (Boukema 1980; Boukema 1984). L^1 is resistant to P_0 pathotype only and is susceptible to other kinds of pathotypes. L^2 is resistant to P_0 and P_1 pathotype, L^3 is resistant to P_0, P_1 and $P_{1.2}$ pathotypes, and L^4 is resistant to P_0, P_1, $P_{1.2}$ and $P_{1.2.3}$ pathotypes. Breeders used L^4 for commercial pepper breeding because resistance spectrum of the L^4 gene

covered all kinds of pathotypes. However, new pathotype $P_{1.2.3.4}$ overcoming resistance of the L^4 allele has reported recently (Genda et al. 2007). Additionally temperature sensitive alleles L^{1a} was reported. The L^{1a} allele confers resistance against P_0 pathotype like L^1. However the resistance is overcome at and above 26°C (Sawada et al. 2004).

2) Elicitor responsible for activating resistance

HR is the resistance response of the *L* locus triggered by infection of *Tobamovirus* inducing local necrotic lesion on plant tissues. The viral determinant of HR had been revealed by construction of chimeric viral genome (Berzal-herranz et al. 1995; de la Cruz et al. 1997; Gilardi et al. 1998; Gilardi et al. 2004; Matsumoto et al. 2008). Coat protein (CP) translated from 17.5kb sgRNA of *Tobamovirus* encoding capsid subunit is a elicitor activating HR. Single amino-acid substitution and comparison of amino acid sequence between HR-inducing *Tobamovirus* and *L*-breaking *Tobamovirus* provided means to figure out key amino acid residues of coat protein activating HR (Berzal-herranz et al. 1995; Genda et al. 2007; Hamada et al. 2002; Mizumoto et al. 2012). However, key amino acid residues suggested by research groups are not consistent. The single amino acid substitution of position 138 of PMMV-I ($P_{1.2.3}$ pathotype) CP, two amino acid substation of position 46 and 85 of PMMoV-J ($P_{1.2.3.4}$ pathotype) CP, and two amino acid substation of position 7 and 81 of

TPW1 (one of PMMoV strain, $P_{1.2.3}$ pathotype) CP have been shown to be responsible for overcoming resistance conferred by the *L* locus.

3) Molecular genetic researches of the *L* locus

Construction of linkage groups in *Capsicum* spp. provides tools to locate genes controlling important horticultural traits and comparative genetic researches enable to discover syntenic relation of resistance gene in three crops of Solanaceae, *Capsicum*, *Solanum* and *Lycopersicum* (Grube et al. 2000; Lefebvre et al. 1995; Lefebvre et al. 2002). The *Tobamovirus* resistance locus *L* was mapped at telomeric region of long arm of chromosome 11 in intraspecific linkage map with Random amplified polymorphic DNA (RAPD) and Restriction fragment length polymorphism (RFLP) markers (Lefebvre et al. 1995). Linkage groups of *Capsicum* spp. were constructed using three interspecific populations, and five genes, *C* (pungency contents), *L* (*Tobamovirus* resistance), *pvr2* (Potyvirus resistance), *Pvr4* (Potyvirus resistance), and *up* (erect fruit habit), were mapped (Lefebvre et al. 2002). *Tobamovirus* resistance was segregated in two doubled haploid (DH) populations, 'H3' x 'Vania' and 'Perennial' x 'Yolo wonder', and the *L* locus was located in 4.0cM from TG36 marker. Comparative genetic analysis of resistance genes of *Capsicum*, *Lycopersicum* and *Solunum* demonstrated that resistance genes are not distributed randomly but found in syntenious genetic region (Grube et al. 2000). The *L* locus belongs to one

resistance gene cluster located in chromosome 11 with two of QTLs, *cmv4* (CMV resistance) and *phyt3* (Phytophthora resistance) in *Capsicum* genome. Corresponding resistance clusters are observed in both of *Lycopersicum* and *Solanum* genomes. Three *Phytophthora* resistance genes (*R3*, *R6* and *R7*), one Phytophthora QTL (*phyto7*), and one nematode QTL (*Gro1.3*) are located in syntenious *Solanum* cluster, and *Fusarium* resistance gene *I2*, *Stemphylium* resistance gene *Sm*, and *Tomato yellow leaf curl virus* (TYLCV) resistance gene *Ty-2* are located in syntenic *Lycopersicum* cluster.

Molecular markers closely linked to the *L* locus were developed for application to marker assisted selection (MAS) and isolation of the *L* gene (Kim et al. 2008; Matsunaga et al. 2003; Sugita et al. 2004; Tomita et al. 2008; Yang et al. 2009; Yang et al. 2012). A total of 516 arbitary primers were screened to find markers linked to the L^4 gene, and one primer WA31 amplified 1,500 bp DNA fragment from resistant individuals specifically. SCAR marker WA31-1500S was redesigned to amplified 1,500 bp fragment, and this marker located in the distance of 1.5cM from the L^4 gene (Matsunaga et al. 2003). Two codominant RAPD markers, $E18_{272}$ and $E18_{286}$, were used to develop markers linked to the L^3 allele using DH population and backcross population (Sugita et al. 2004). These markers converted to three SCAR markers, $PMFR11_{269}$, $PMFR11_{283}$ and $PMFR21_{200}$, and were linked to the L^3 gene in a distance of 4.0 cM. Amplified

fragment length polymorphism (AFLP) – bulked segregant analysis (BSA) strategy using near isogenic lines (NILs) populations was performed to develop markers linked to the *L*^{4} locus (Kim et al. 2008). A total of 19 primer combinations amplified specific band in resistant bulks. Linkage analysis of 19 pirmer pairs in $BC_{10}F_2$ NIL populations T102 was carried out, and finally three primer sets, L4-a, L4-b and L4-c, were found to be linked to the *L4* gene in a distance of 2.5 cM. L4-b was converted into SCAR marker L4SC340. High resolution linkage map of the *L*^{3} gene was constructed through BSA-AFLP, resistance gene analogs (RGA), and bacterial artificial chromosome (BAC) library screening strategies using two intergenic F_2 populations (Tomita et al. 2008). A total of eight markers were located within a 1.5 cM from the *L*^{3} gene and two markers A339 and 189D23M were perfectly co-segregated in both F_2 populations, NK (n=2,016) and YB (n=3,391). Two BAC contigs separated about 30kb from each others were constructed and the *L*^{3} gene was mapped on these contigs.

4) Cloning of the *L*^{3} gene

The *L*^{3} gene was encompassed in about 400-kb region covered by two contigs (Tomita et al. 2008). Sequence of BAC contigs were analyzed and a total of nine resistance gene homologs of *I2* and *R3a* genes were found in this region (**Tomita et al. 2011)**. Seven homologs were pseudogene, and *PIH-X* homolog

was positioned out of expected genetic region of the L^3 gene by linkage analysis. The other homolog, the L^3 candidate, was isolated and co-infiltrated with CP of P_0, P_1, $P_{1.2}$, $P_{1.2.3}$ and $P_{1.2.3.4}$ pathotype of *Tobamovirus* in *Nicotiana benthamiana* system and interacted HR occurred with CP of P_0, P_1 and $P_{1.2}$ pathotype. The L^3 gene encodes 4,856 bp transcript and belongs to coiled-coil nucleotide binding leucine rich repeat (CNL) class like *I2* and *R3a*. Other *L* allele candidates, L^1, L^{1a}, L^{1c}, L^2, L^{2b} and L^4, were isolated using homology based PCR method using sequence of the L^3 gene. Co-expression analysis demonstrated that these candidates can induce HR by interaction with CP of *Tobamovirus*.

REFERENCES

Bagley CA (2001) Controlling *Tobacco mosaic virus* in tobacco through resistance. Thesis. Virginia Polytechnic Institute and State University

Berzal-herranz A, de la Cruz A, Tenllado F, Diaz-ruiz JR, López L, Sanz AI, Vaquero C, Serra MT, Garcia-luque I (1995) The Capsicum L3 gene-mediated resistance against the tobamoviruses is elicited by the coat protein. Virology 209:498-505

Boukema IW (1980) Allelism of genes controlling resistance to TMV in *Capsicum* L. Euphytica 29:433-439

Boukema IW (1984) Resistance to TMV in *Capsicum chacoense* Hunz. is governed by allele of the *L*-locus. Pepper News Letters 3:47-48

Butler PJG (1999) Self-assembly of assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed. Philosophical Transactions of the Royal Society B: Biological Sciences 354:537-550

Dawson WO (1992) *Tobamovirus*-plant interactions. Virology 186:359-367

de la Cruz A, López L, Tenllado F, Díaz-Ruíz JR, Sanz AI, Vaquero C, Serra MT, and García Luque I (1997) The coat protein is required for the elicitation of the *Capsicum* L^2 gene-mediated resistance against the tobamoviruses. Molecular Plant-Microbe Interactions 10:107-113

Deslandes L, Olivier J, Peeters N, Feng DX, Khounlotham M (2003) Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. Proceedings of the National Academy of Sciences 100:8024–29

Dodds P and Thrall P (2009) Recognition events and host–pathogen co-evolution in gene-for-gene resistance to flax rust. Functional Plant Biology 36:395-408

Dodds PN, Lawrence GJ, Catanzariti AM, Teh T, Wang CIA (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proceedings of the National Academy of Sciences 103:8888–93

Flor HH (1971) Current status of the gene-for-gene concept. Annual Review of Phytopathology 9:275-296

Gibbs A (1999) Evolution and origins of tobamoviruses. Philosophical Transactions of the Royal Society B: Biological Sciences 354:593-602

Gilardi P, García-Luque I, Serra MT (1998) *Pepper mild mottle virus* coat protein alone can elicit the capsicum spp. L^3 gene-mediated resistance. Molecular Plant-Microbe Interactions 11:1253-1257

Gilardi P, García-Luque I, Serra MT (2004) The coat protein of tobamovirus acts as elicitor of both L^2 and L^4 gene-mediated resistance in *Capsicum*. Journal of General Virology 85:2077-2085

Genda Y, Kanda A, Hamada H, Sato K, Ohnishi J, Tsuda S (2007) Two amino acid substitutions in the coat protein of *Pepper mild mottle virus* are responsible

for overcoming the L^4 gene-mediated resistance in *Capsicum* spp. Phytopathology 97:787-793

Goelet P, Lomonossoff GP, Butler PJG, Akam ME, Gait MJ, Karn J (1982) Nucleotide sequence of tobacco mosaic virus RNA. Proceedings of the National Academy of Sciences of the United States of America 79:5818-5822

Grube RC, Randwanski ER, Jahn M (2000) Comparative genetics of disease resistance within the Solanaceae. Genetics 155:873-887

Hamada H, Takeuchi S, Kiba A, Tsuda S, Hikichi Y, Okuno T (2002) Amino acid changes in *Pepper mild mottle virus* coat protein that affect L^3 gene-mediated resistance in pepper. Journal of General Plant Pathology 68:155-162

Hammond-Kosack KE and Jones JDG (1997) Plant disease resistance gene. Annual Review of Plant Physiology and Plant Molecular Biology 48:575-607

Huang S, Vleeshouwers VG, Werij JS, Hutten RC, van Eck HJ, Visser RG, Jacobsen E (2004) The *R3* resistance to *Phytophthora infestans* in potato is conferred by two closely linked R genes with distinct specificities. Molecular Plant-Microbe Interactions 17:428–435

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. The Plant Journal 42:251–261

Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance.

EMBO Journal 19:4004–14

Jones JDG and Dangl JL (2006) The plant immune system. Nature 444:323-329

Kanzaki H, Yoshida K, Saitoh H, Fujisaki K, Hirabuchi A, Alaux L, Fournier E, Tharreau D, Terauchi R (2012) Arms race co-evolution ofMagnaporthe oryzae AVR-Pikand ricePikgenes driven by their physical interactions. The Plant Journal 72:894-907

Kegel WK and van der Schoot P (2006) Physical regulation of the self-assembly of *Tobacco mosaic virus* coat protein. Biophysical Journal 91:1501-1512

Keith R and Mitchell-Olds (2013) Genetic variation for resistance to herbivores and plant pathogens: hypotheses, mechanisms and evolutionary implications. Plant Pathology DOI: 10.1111/ppa.12134

Kim HJ, Han JH, Yoo JH, Cho HJ, Kim BD (2008) Development of a sequence characteristic amplified region marker linked to the L^4 locus conferring broad spectrum resistance to tobamoviruses in pepper plants. Molecules and Cells 25:205-210

Kim HS, Desveaux D, Singer AU, Patel P, Sondek J, Dangl JL (2005) The *Pseudomonas syringae* effector AvrRpt2 cleaves its C-terminally acylated target, RIN4, from Arabidopsis membranes to block RPM1 activation. Proceedings of the National Academy of Sciences 102:6496–501

Kim NR, Hong JS, Song YS, Chung BN, Park JW, Ryu KH (2012) The complete genome sequence of a member of a new species of *Tobamovirus* (*Rattail cactus necrosis-associated virus*) isolated from *Aporcactus flagelliformis*. Archives of

Virology 157:185-187

Lefebvre V, Palloix A, Caranta C, Pochard E (1995) Construction of an intraspecific integrated linkage map of pepper using molecular markers and doubled-haploid progenies. Genome 38:112-121

Lefebvre V, Pflieger S, Thabuis A, Caranta C, Blattes A, Chauvet JC, Daubèze AM, Palloix A (2002) Towards the saturation of the pepper linkage map by alignment of three intraspecific maps including known-function genes. Genome 45:839-854

Levy A, Zheng JY, Lazarowitz SG (2013) The *Tobamovirus Turnip vein clearing virus* 30-kilodalton movement protein localizes to novel nuclear filaments to enhance virus infection. Journal of Virology 87:6428-6440

Lewandowski DJ and Dawson WO (2000) Functions of the 126- and 183-kDa proteins of *Tobacco mosaic virus.* Virology 271:90-98

Mackey D, Holt BF, Wiig A, Dangl JL (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, Bogdanove AJ, Sessa G (2003) Understanding the function of plant disease resistance gene. Annual Review of Plant Biology 54:23-61

Matsumoto K, Sawada H, Matsumoto K, Hamada H, Yoshimoto E, Ito T, Takeuchi S, Tsuda S, Suzuki K, Kobayashi K, Kiba A, Okuno T, Hikichi Y (2008) The coat protein gene of tobamovirus P0 pathotype is a determinant for activation of temperature-insensitive L^{1a}-gene-mediated resistance in *Capsicum* plants.

Achives of Virology 153:645-650

Matsunaga H, Saito T, Hirai M, Nunome T, Yoshida T (2003) DNA markers linked to *Pepper mild mottle virus* (PMMoV) resistance locus (L^4) in *Capsicum.* Journal of the Japanese Society for Horticultural Science 72:218-220

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. The Plant Cell 15:809–834

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

Mizumoto H, Nakamura I, Shimomoto Y, Sawada H, Tomita R, Sekine KT, Kiba A, Nishiguchi M, Kobayashi K, Hikichi Y (2012) Amino acids in *Tobamovirus* coat protein controlling pepper L^{1a} gene-mediated resistance. Molecular Plant Pathology 13:915-922

Mucyn TS, Clemente A, Andriotis VME, Balmuth AL, Oldroyd GED (2006) The tomato NBARC-LRR protein Prf interacts with Pto kinase *in vivo* to regulate specific plant immunity. Plant Cell 18:2792–806

Naqvi SAMH (2004) Disease of fruits and vegetables volume II. Dordrecht: Kluwer Academic Publishers

Ori N, Eshed Y, Paran I, Presting G, Aviv D, Tanksley S, Zamir D, 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. The Plant Cell 9:521–532

Padmanabhan M, Cournoyer P, Dinesh-Kumar SP (2009) The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. Cellular Microbiology 11:191-198

Rhie MJ, Min BE, Hong JS, Song YS, Ryu KH (2007) Complete genome sequence supports *Bell pepper mottle virus* as a species of the genus *Tobamovirus*. Archives of Virology 152:1401-1407

Rooney HC, Van't Klooster JW, van der Hoorn RA, Joosten MH, Jones JD, de Wit PJ (2005) Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. Science 308:1783–86

Sawada H, Takeuchi S, Hamada H, Kiba A, Matsumoto M, Hikichi Y (2004) A new Tobamovirus-resistance gene, L^{1a}, of sweet pepper (*Capsicum annuum* L.) Journal of the Japanese Society for Horticultural Science 73:552-557

Shao F, Golstein C, Ade J, Stoutemyer M, Dixon JE, Innes RW (2003) Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. Science 301:1230–33

Sugita T, Ymaguchi K, Sugimura Y, Nagata R, Yuji K, Kinoshita T, Todoroki A (2004) Development of SCAR markers linked to L^3 gene in *Capsicum*. Breeding Science 54:111-115

Stobbe AH, Melcher U, Palmer MW, Roossinck MJ, Shen G (2012) Co-divergence and host-switching in the evolution of tobamoviruses. Journal of General Virology 93:408-418

Tellando F, Garcia-luque I, Serra MT, Diaz-ruiz JR (1995) *Nicotiana benthamiana* plants transformed with the 54-kDa region of the *Pepper mild mottle tobamovirus* replicase gene exhibit two type of resistance responses against viral infection. Virology 211:170-183

Tomita R, Sekine KT, Mizumoto H, Sakamoto M, Murai J, Kiba A, Hikichi Y, Suzuki K, Kobayashi K (2011) Genetic basis for the hierarchical interaction between *Tobamovirus* spp. and *L* resistance gene alleles from different pepper species. Molecular Plant-Microbe Interactions 24:108-117

Ueda H, Yamaguchi Y, Sano H (2006) Direct interaction between the tobacco mosaic virus helicase domain and the ATP-bound resistance protein, N factor during the hypersensitive response in tobacco plants. Plant Molecular Biology 61:31–45

van der Hoorn RAL and Kamoun S (2008) From to guard to decoy : a new model for perception of plant pathogen effectors. The Plant Cell 20:2009-2017

van Ooijen G, van der Burg HA, Cornelisson BJC, Takken FLW (2007) Structure and function of resistance protein in Solanaceous plants. Annual Review of Phytopathology 45:43-72

CHAPTER I

Development of SNP markers linked to the *L* locus in *Capsicum* spp. by a comparative genetic analysis

ABSTRACT

In pepper, the TMV resistance locus *L* is syntenic to the tomato *I2* and the potato *R3* loci on chromosome 11. In this report, I identified pepper bacterial artificial chromosome (BAC) clones corresponding to the *I2* and *R3* loci and developed *L*-linked markers using the BAC sequence information. A BAC library was screened using the tomato *I2C-1* gene as a probe. The resulting clones were sorted further by PCR screening, sequencing, and genetic mapping. A linkage analysis revealed that BAC clone 082F03 could be anchored to the target region near TG36 on chromosome 11. Using the 082F03 sequence, more BAC clones were identified and a BAC contig spanning 224 kb was constructed. Gene

prediction analysis showed that there were at least three *I2*/*R3* R gene analogs (RGAs) in the BAC contig. Three DNA markers closely linked (about 1.2 cM) to the L^4 gene were developed by using the BAC contig sequence. The single nucleotide polymorphism marker 087H3T7 developed in this study was subjected to linkage analysis in L^4- and L^3-segregating populations together with previously developed markers. The 189D23M marker, which is known to co-segregate with L^3, was located on the opposite side of 087H3T7, about 0.7 cM away from L^4. This supports the idea that L^3 and L^4 may be different genes closely linked within the region instead of different alleles at the same locus. Finally, use of flanking markers in molecular breeding program for introgression of L^4 to elite germplasm against most aggressive tobamoviruses pathotype $P_{1.2.3}$ is discussed.

INTRODUCTION

Comparative genetic analysis of disease resistance in crop plants suggests that disease resistance genes (R genes) do not distribute randomly within a genome (Grube et al. 2000; Wisser et al. 2005, 2006). Clustering of R genes conferring resistance to several unrelated pathogen types often occurs and their syntenic positions across the genomes of tomato, potato, and pepper indicate that this clustering is conserved within the Solanaceae (Grube et al. 2000).

An earlier comparative genetic map had revealed the syntenic relationship between the *I2* gene cluster in tomato and *R3* gene complex in potato (Grube et al. 2000). In tomato, the *I2C-1* gene has been cloned and demonstrated to confer partial resistance to race 2 of the soil-borne fungus *Fusarium oxysporum* f. sp lycopersici. The gene belongs to a multigene *I2* complex (*I2C*) family and is located within the SL8Dcluster on chromosome 11 in tomato. At least four other *I2C* homologs have been mapped and found to be distributed on three chromosomes in the tomato genome, of which the SL8E cluster has been mapped closely with SL8D. Genomic dissection of the SL8D and SL8E regions consequently led to the revelation of a microsyntenic relationship between the SL8Ecluster in tomato and *R3a* cluster in potato (Huang et al. 2005; Ori et al. 1997). Based on colinearity, the *R3a* gene, which confers resistance to

Phytophthora infestans in potato, has been isolated and cloned via a comparative genomics approach with the resistance gene analog of *I2* (Huang et al. 2005). *R7*, which also confers late blight resistance and maps to the same locus, has been demonstrated to be allelic to the *R3* gene complex (Huang et al.2004, 2005). Sequence analysis also demonstrated that *R3a* is an *I2* gene analog (GA) and shares 88% DNA sequence similarity (Huang et al.2005). Furthermore, both *I2C-1* and *R3a* are coiled-coil nucleotidebinding site leucine-rich repeat (CC–NBS–LRR) type R genes (due to the presence of a predicted nucleotide binding site and leucine-rich repeats with a coiled coil region at the N-terminus) (Ballvora et al.2002; Huang et al.2005; Kuang et al. 2005; Song et al. 2003; van der Vossen et al.2003, 2005), which is the most abundant type of R gene in plant genomes among the five classes of R genes responsible for disease resistance in plants (Dangl and Jones2001; Meyers et al.1999, 2003; Michelmore and Meyers 1998). The orthologous relationship, supported by syntenic positions and sequence similarity, indicate that the two genes belong to an ancient R gene locus prior to the divergence of tomato and potato.

In pepper, the syntenic position of the *I2C* and *R3a* loci contains an R-gene cluster having several quantitative trait loci (QTL) and also a dominant resistance gene, *L* (Grube et al. 2000). At the *L* locus, four sources of resistance genes have been found from cultivars and wild species of pepper including

Capsicum annuum (*L*^{1}), *Capsicum frutescens* (*L*^{2}), *Capsicum chinense* (*L*^{3}), and *Capsicum chacoense* (*L*^{4}) (Berzal-Herranz et al. 1995; Boukema 1980;dela Cruz et al.1997). Allelism of the four genes controlling resistance has been demonstrated in an early genetic study of TMV resistance (Boukema1980). *L* has been demonstrated to confer resistance to several *Tobamovirus* infecting pepper, including *Tobacco mosaic virus* (TMV), *Tomato mosaic virus* (ToMV), *Paprika mild mottle virus* (PaMMV) and *Pepper mild mottle virus* (PMMoV) (Csillery et al.1983; Pernezny et al.2003; Wetter et al. 1984). These *Tobamovirus* have been classified into different pathotypes—P_0, P_1, $P_{1.2}$, and $P_{1.2.3}$—based on their pathogenicity and interaction with the counterpart resistance *L* genes (Alonso et al. 1991; Takeuchi et al. 2005). Several PMMoV isolates classified as P1.2.3 have been shown to be the most aggressive, all *L* gene-mediated resistance breaks down where systemic infection occurs with the exception of *L*^{4} (Tsuda et al. 1998; Velasco et al.2002; Matsunaga et al.2003; Kim et al. 2008). *L*^{4} confers resistance against most pathotypes of *Tobamovirus* by inducing hypersensitive response (HR). The coat protein of the *Tobamovirus* has been demonstrated to elicit *L* gene-mediated resistance by inducing a distinctive HR (Berzal-Herranz et al.1995; de la Cruz et al.1997).

L has been mapped at the lower arm of pepper chromosome 11 (Lefebvre et al.1995). Since then, several *L*-linked markers have been developed using

various DNA markers and mapping populations. Random amplified polymorphic DNA (RAPD) markers, which are located at 4.0 cM from the L^3 locus, were developed by Sugita et al. (2004). Recently, more tightly linked markers to the L^3 locus (<0.1 cM away) were developed using bulked segregant analysis-amplified fragment length polymorphism (BSA-AFLP) and resistance gene analog (RGA) mapping (Tomita et al.2008). Several L^4-linked markers were also developed. Matsunaga et al. (2003) reported L^4-linked RAPD markers. One of the RAPD markers (WA31-1500), located 1.5 cM from the L^4 locus, was converted to a SCAR marker. Another mapping study showed three AFLP markers linked to L^4 and the closest marker (L4SC340) was converted into a SCAR marker (Kim et al. 2008). When this marker was tested in two populations, the marker was mapped 0.9 and 1.8 cM away from L^4. Availability of these markers provides a great opportunity for fine mapping of the L^4 gene and for elucidating the allelic relationship between L^3 and L^4.

In this study, we have performed a comparative genomic study between the *R3*/*R7* locus in potato, the *I2* locus in tomato and the *L* locus in pepper to investigate the possible orthologous relationships among the R-gene clusters on chromosome 11. The goal of this study was to identify molecular markers, particularly single nucleotide polymorphism (SNP) markers, closely linked to the *L* gene through comparative genomics of the syntenic regions among the three

Solanaceous crops. We have also defined available markers and constructed a localized map flanking the *L* locus. Furthermore, we elucidated the allelic relationships between L^3 and L^4 using *L*-linked markers.

MATERIALS AND METHODS

Plant and virus materials

Capsicum annuum 'Early California Wonder (ECW)' and 'NuMex RNaky (RN)' were purchased from Asgrow Seed Co. (San Juan Bautista, CA, USA); 'Special' (SP) and 'Cupra' (CP) were obtained from Enza Zaden (Enkhuizen, The Netherlands); MyoungSung (MS) was purchased from Seminis Korea Inc. Hungnong Seeds. (Seoul, Korea). *C. chinense* PI159234 (234) and PI159236 (236) and *C. chacoense* PI260429 (429) were obtained from the USDA Southern Regional Plant Introduction Station (Experiment, GA, USA). *C. chacoense* PI260429 was used as a TMV resistant control, whereas ECW was the TMV susceptible control. CP had the L^3 allele, and SP and MS had the L^4 allele. For virus resistance and genotype analysis, F_2-segregating populations were developed from the commercial F_1 hybrid cultivars CP, SP, and MS. An F_2 population derived from a cross between RN and 234 was used for mapping DNA markers.

Virus strains and resistance screening

Two different strains of TMV P_0 and $P_{1.2.3}$ were used for virus resistance screening: TMV P_0 was provided by K. H. Paek (Korea University, Seoul, Korea),

and TMV $P_{1.2.3}$ was provided by K. H. Ryu (Seoul Women's University, Seoul, Korea). The virus strains were maintained and multiplied on *N. tabaccum* var. 'Samsun'. For virus inoculation, infected tobacco leaves were ground in chilled 50 mM potassium phosphate buffer (pH 7.5). Mechanical inoculation was carried out by applying virus inoculum onto the two oldest leaves of plants at the 5- to 6-leaf stage. Plants were monitored daily and observed for the presence of virus symptoms and hypersensitive response. Resistance to TMV was determined by the presence of hypersensitive response. In order to confirm HR response, virus accumulation was monitored in inoculated plants by the DAS-ELISA method.

BAC clone analysis and sequencing

The pepper BAC library was provided by J. Giovannoni (Cornell University, NY, USA). The BAC library was constructed using nuclear DNA from *C. frutescens* BG2816, which is known to contain the L^2 allele. The library is publically available at Arizona Genomics Institute (http://www2.genome.arizona.edu). The average insert size of this BAC library was estimated to be 125–130 kb. A total of 220,000 BAC clones were screened using the 30 sequence of the tomato *I2C-1* gene as a probe. Positive BAC clones were subjected to group using a BAC fingerprinting method. BAC fingerprinting was outsourced to Amplicon Express (Pullman, WA: http://www.genomex.com). After obtaining

BAC fingerprints, we assembled BAC contigs with FPC program with tolerance setting of 3 and a cutoff of 1 x e^{-10}. The 89 *I2C* containing BAC clones also screened by PCR using a set of three primers: R7-1, R7-2, and LRR (Table 1). These primers were provided by B. Baker (U.C.Berkeley, CA, USA) and were designed to specifically amplify *R3a* homolog sequences. Amplified PCR fragments from 22 BAC clones using LRR primers were cloned and sequenced [Biotechnology Resource Center (BRC), Cornell University, NY, USA]. Sequences were analyzed and grouped using DNAstar and MegAlign program (DNASTAR, Inc., USA) to identify candidate BAC clones. The nine candidate BAC clones were sent to the Institute for Genomic Research (TIGR, MD, USA) for draft sequencing. After anchoring draft sequenced BAC clones (BAC clone 060I2 and 082F3) near the *L* locus, the remaining BAC clones were rescreened by PCR with primers (pepBAC060I2-H3) derived from BAC 060I2. BAC end sequences identified by pepBAC060I2-H3 were determined [National Instrumentation Center for Environmental Management (NICEM), Seoul National University, Korea], and primers were redesigned for the next round of screening. This screening for IC2-positive BAC clones was repeated until no further positive clones were obtained. After obtaining a BAC contig, three BAC clones—043M10, 158K24, and 268G7—were selected and sequences were determined by a shotgun sequencing method (NICEM, Seoul National University,

Korea).

BAC sequence annotation

The FGENESH program (http://linux1.softberry.com/berry.phtml) was used to predict genes from the contig sequence based on tomato organism information. To search for reported proteins with similar predicted amino acid sequences, the BLASTP program (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) was used. BLASTP was also used to calculate similarity and identity. To identify coiled-coil domains, the MARCOIL program (http://www.isrec.isbs-ib.ch/webmarcoil/webmarcoilC1.html) was used with default settings. Sequences having total coiled-coil probability over 90% were determined to code for a coiled-coil domain. Multiple sequence alignment was performed using ClustalW2 program (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Global pairwise analysis was performed using Matcher program (http://bioweb.paste-ur.fr/seqanal/interfaces/matcher.html) with default setting. Score of match was set 1, and gap open penalty and gap extend penalty were set 16 and 6 respectively. Neighbor-joining, minimum evolution, maximum parsimony, and UPGMA analysis were conducted using MEGA program version 4 (Tamura et al. 2007). Phylogenetic test was estimated using 1,000 bootstrap replicates and the other options follow default value.

Development of markers linked to the *L* locus based on BAC sequences

1) Simple sequence repeat marker

Simple sequence repeat markers were designed from the SSR Discovery program (PGG Bioinformatics, http://hornbill.cspp.latrobe.edu.au/ssrdiscovery.-html) based on the draft sequences of the nine selected BAC clones. In order to obtain markers with a higher chance of polymorphism, only primers flanking putative 59 tri-nucleotide repeats or 79 bi-nucleotide repeats were used. Three to six SSR markers were designed for each selected BAC clone and a total of 40 SSR markers were developed and tested. SSR marker analysis was conducted according to Yi et al. (2006).

2) Single nucleotide polymorphism marker

Single nucleotide polymorphism markers linked to the *L* loci were developed from the end sequences of BAC clones. To find SNPs, sequences of PCR products amplified using BAC-derived primers were compared between *C. chacoense* (L^4) and *C. annuum* (L^0), resistant and susceptible plants from *L*-segregating populations respectively. SNPs were converted to cleaved amplified polymorphic sequence (CAPS) or high resolution melting (HRM) markers. The CAPS finder program (www.sgn.cornell.edu) was used to identify restriction enzymes (RE) cutting potential SNPs. SspI (087H3T7 CAPS marker) was used

according to protocols provided by the manufacturers. Cut DNA products were electrophoresed on 1.5% agarose gels. For HRM analysis, PCR was performed using a Rotor-Gene TM 6000 (Corbett, Australia). PCR was carried out in a 20ll volume containing 60 mM KCl, 10 mM Tris–Cl, 2.5 mM MgCl2, 0.25 mM each dNTP, 5 pmol each primer, 1 unit Taq polymerase, 1.25 μM Syto9, and 50 ng gDNA stock. After the PCR, HRM was performed each 0.1^{o}C from 65 to 90^{o}C.

Linkage analysis of molecular markers

Linkage analysis of markers developed in this study and previously was performed using AC99, SP, CP, and MS F_2 populations. Linkage analysis was performed using CarthaGene software (de Givry et al.2005) with a LOD score threshold of 4.0 and a maximum distance of 30 cM.

Table 1. Primer information lists in this research

Primer Name	Primer sequence (5’ → 3’)	Experimental purpose
R7-1F	GAGAAATGGAGATTGGCTTAG	BAC library screening
R7-1R	AGATTTGAAGATGCTATTTGG	BAC library screening
R7-2F	CCTGTGGATTGCTAATGGTCT	BAC library screening
R7-2R	CTCAAGAGAGTTAAATGGCTTT	BAC library screening
LRR-F1	TGCATGGAATAACAGAGGTGA	BAC library screening
LRR-R1	TAGAGAGGGAGGAGGGCAGT	BAC library screening
pepbac082F3-5F	TCTTACCTGTTGACTGCTGA	BAC anchoring
pepbac082F3-5R	ATACTTAGGGCTTACCCGTC	BAC anchoring
pepbac101J3-4F	AAACAGGAGAGGCATAGTGA	BAC anchoring
pepbac101J3-4R	CTTCTCCCTTGTTGTTCTTG	BAC anchoring
pepbac181M9-1F	CTTCTCCCTTGTTGTTCTTG	BAC anchoring
pepbac181M9-1R	AAACAGGAGAGGCATAGTGA	BAC anchoring
pepbac337L21-1F	AGGATGTAAGGGATTTAGGC	BAC anchoring
pepbac337L21-1R	GTTCAAATCAATTTCATGGC	BAC anchoring
pepbac060I2-E1F	GGAAAACGATTTGTGGTACG	Construction of a contig
pepbac060I2-E1R	GACCCTCCAAATCTCCACTT	Construction of a contig
pepbac060I2-E2F	TGAACAACTGTCTGCGAAAA	Construction of a contig
pepbac060I2-E2R	TGTGGTCTTCCGTACCTGAT	Construction of a contig
pepbac060I2-E3F	CCTCTATTGTTCCGATGTGG	Construction of a contig
pepbac060I2-E3R	CACGGGATACTTGTCACCTC	Construction of a contig
pepbac060I2-H1F	CTGAAAGAGGTTGGTTTTTG	Construction of a contig

Primer Name	Primer sequence (5’ → 3’)	Experimental purpose
pepbac060I2-H1R	GGAACTCTTTCTTAGCAGCA	Construction of a contig
pepbac060I2-H2F	CTTTGCAAATTCAAGCTTCT	Construction of a contig
pepbac060I2-H2R	CGTGCTCTAAATCAAGCTCT	Construction of a contig
Pepbac060I2-H3F	ATGCTCTGAGAGAGGGGAGT	Construction of a contig
pepbac060I2-H3R	AGACACCCATCCAACTTTCA	Construction of a contig
060I2END-1F	AAGGGACTGTTTCCACGGCT	Construction of a contig
060I2END-1R	TTCTAGCCAAGGGTGGCCTT	Construction of a contig
060I2END-2F	GCACATCAGCAGGTTTAGTACG	Construction of a contig
060I2END-2R	CCAACTGTCAAACCTCGGTT	Construction of a contig
087H3SP6F	CAGAATTTGGGTGGATCAAAGAG	Construction of a contig
087H3SP6R	TCACCTCAAGTGTGATCTGCC	Construction of a contig
087H3T7F	CCTTTGCCTGCATTATTCTTG	Construction of a contig
087H3T7R	GCCCAAATTTATTCCCAAATGC	Construction of a contig
207E13SP6F	GGCCATGATTTCAATGACTATAA	Construction of a contig
207E13SP6R	TTCGAGGTCTGCTATTCTAG	Construction of a contig
207E13T7F	GCACTAAGGCTATTTTGAGAC	Construction of a contig
207E13T7R	GACACAAGAGAGGCATCCAC	Construction of a contig
158K24SP6F	CGAGGAGGTAGTGGTAAGGT	Construction of a contig
158K24SP6R	CTGGTTAAGATTCTCTGAGGTTG	Construction of a contig
158K24T7F	GAATTACAACAACAAGTGCAACT	Construction of a contig
158K24T7R	CGTTGGCTTGAACGTAGTCAG	Construction of a contig
290J13SP6F	CACCCAATAGCTTAACAAGGG	Construction of a contig

Primer Name	Primer sequence (5’ → 3’)	Experimental purpose
290J13SP6R	CTTCAGCACCTCACTTCGCT	Construction of a contig
290J13T7F	ACAAGTGCAACTAATCTCCATC	Construction of a contig
290J13T7R	GCTTGAACGTAGTCAGAGTAAC	Construction of a contig

RESULTS

Identification of BAC clones cross-hybridized with tomato *I2C-1* gene and potato *R3*/*R7*

To identify pepper *I2* homologs, a pepper BAC library containing 221,184 clones was screened with a probe developed from the 30 sequence of the *I2C-1* gene. A total of 89 positive BAC clones were obtained. In the *R3*/*R7* region in potato, 17 copies of RFLP markers GP285 and TG105 are tightly linked to the *I2* homolog. Therefore, I attempted to hybridize the 89 BAC clones with the two RFLP markers to identify BAC clones containing sequences related to the *R3*/*R7* region. However, none of the 89 clones were cross-hybridized with the two RFLP markers. This demonstrates that potato and pepper have different genomic structures in this region.

To identify clones simultaneously homologous to potato *R3*, the 89 *I2*-positive BAC clones were screened using primer sets developed from potato genomic sequences. The three primer sets—R7-1, R7-2, and LRR—designed from sequence near *R3a* on chromosome XI in potato, amplified 1.5, 0.8–1.3, and 1.2 kb bands from 37, 52, and 22 BAC clones in pepper respectively (Table 1). Twelve BAC clones containing LRR priming sequences also contained both R7-1 and R7-2 priming sequences, while the other ten clones contained either R7-1

or R7-2 priming sequences. In order to pool the BAC clones into groups, the resulting 22 PCR fragments amplified with LRR primers were cloned and sequenced. Sequence similarity analysis revealed that the 22 clones could be grouped into nine different sequence groups. Based on these results, the BAC clones from which both the tomato *I2C-1* gene and the potato *R3a* gene could be amplified were chosen as candidate clones.

Mapping of the candidate BAC clones

The *I2* gene family in tomato is distributed at least at five genomic locations across three chromosomes. It is hard to estimate how many copies there might be for the *L* gene or gene family in pepper. Nevertheless, *L* was previously mapped 5.2 cM away from RFLP marker TG36 on chromosome 11 (Paran et al.2004), this region of the pepper genome is syntenic to the *R3*/*R7* region in potato and the *I2* region in tomato. I therefore focused our study on the candidate BAC clones mapped in the target region. To locate the candidate BAC clones on the pepper genetic map, SSR markers were developed using the nine draft BAC clone sequences. I successfully anchored four BAC sequences using six SSR markers on our reference mapping population AC99 (Livingstone et al. 1999) (Table 1). One SSR marker, pepBAC082F3-5 (developed from BAC clone 082F3), was cosegregated with TG36 on chromosome 11 (syntenic with

I2/R3loci) (Fig.1). Another marker representing BAC clone 337L21 was mapped on the same chromosome but near TG105. Other markers mapped outside of the target syntenic region and therefore were discarded for the following analysis.

Based on the BAC fingerprinting results, BAC clone 060I2 was closely related to BAC clone 082F3 (Fig.2). Therefore, I also mapped BAC clone 060I2. Since no polymorphic SSR markers had been developed for this BAC, six markers were developed from the BAC end sequences of BAC clone 060I2. One of which, pepBAC060I2-H3, polymorphic marker was cosegregated with pepBAC082F3-5 and mapped near TG36 on chromosome 11 (Fig.3a).

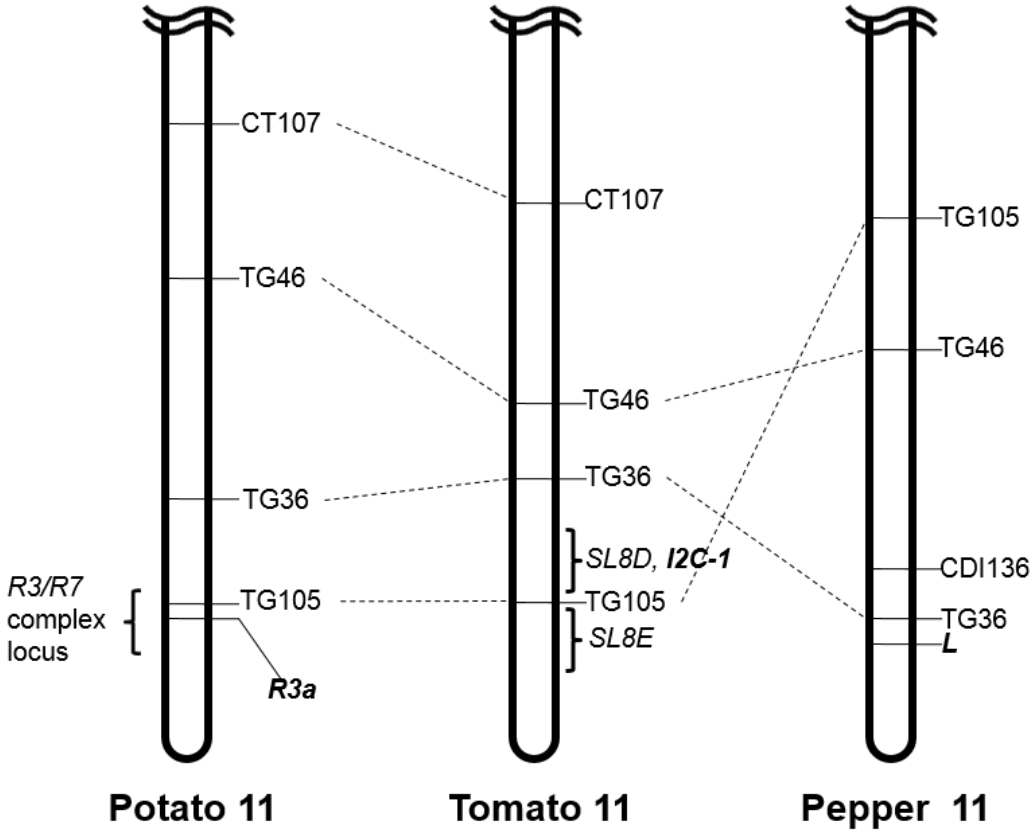


Figure 1. Comparative map showing positions of resistance genes on the lower arms of chromosome 11 of potato, tomato, and pepper. Marker order close to syntenic gene group including *L*, *I2*, and *R3* was conserved well except TG105 marker

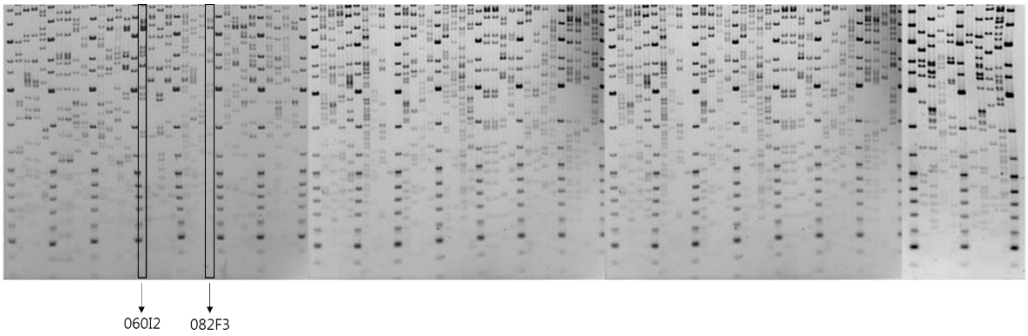


Figure 2. Fingerprinting result of 89 BAC clones. Band patterns of 060I2 and 082F3 are similar to each other.

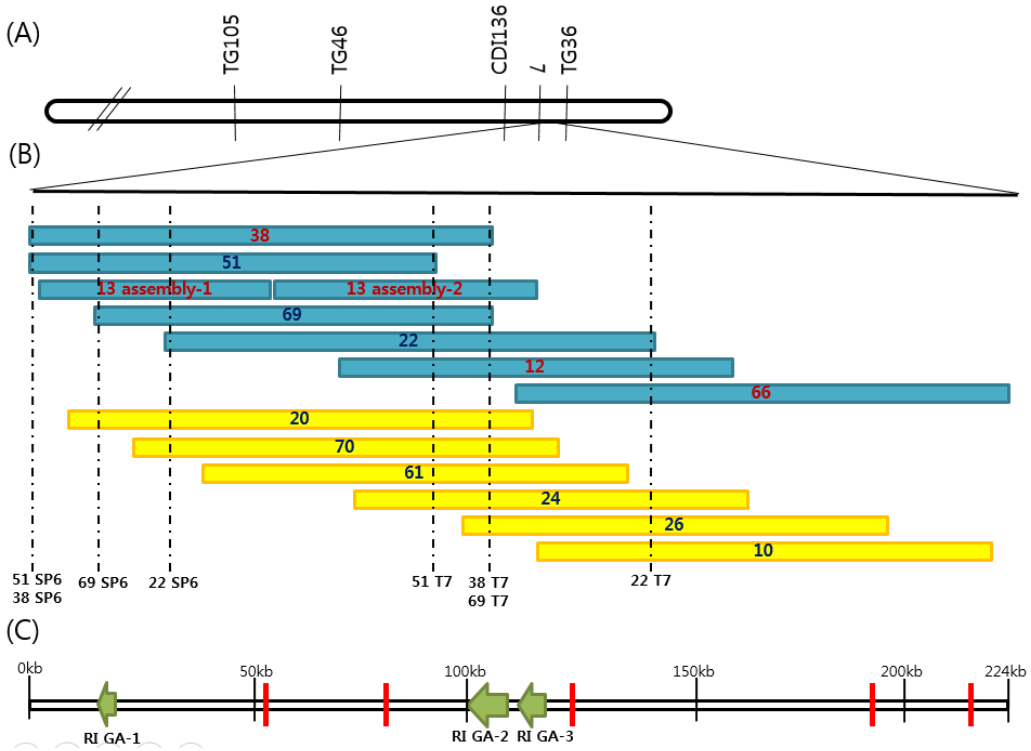


Figure 3. Partial contig of 13 BAC clones. **a, b** Partial contig was constructed using six primers (under dotted lines). Dark grey rectanglesquare located based on sequence information and light grey rectangles are located based on PCR information of six primers. About 224 kb DNA sequences were obtained by assembly of four BAC clones (158K24, 060I2, 043M10, and 268G7). **c** Gene prediction conducted by FGENESH program. Three *R3a*, *I2* family gene analogs (RIGA, arrows) and five transposable elements with long terminal repeats (LTRs, thick bars) were predicted

Construction of a partial contig

The BAC contig assembled with the FPC program was re-evaluated by PCR using primers derived from BAC clone 060I2. However, there were some discrepancies between PCR analysis and the BAC fingerprinting: some BAC clones predicted to have overlapping sequences by BAC fingerprints turned out to be false positive in PCR analysis. This may be due to repetitive nature of R gene homologs. Therefore, I attempted to construct a new contig by PCR using BAC end primers. Six primers were designed from four BAC clones (087H3, 158K24, 207E13, and 290J13), which had strong positive signals by PCR using primer pepBAC060I2-E4, were selected to rescreen 89 BAC clones. I was able to obtain an additional 12 BAC clones. A new contig was constructed by presence/absence of amplicons giving overlap order (Fig.3b). Three BAC clones (043M10, 158K24, and 268G7) in the contig were selected and fully sequenced.

Development of SNP markers linked to the *L* alleles

I developed *L*-linked markers by making use of the BAC contig information (Fig.3b). It is known that the L^4 gene originated from *C. chacoense* and had been introgressed into the *C. annuum* genome during the breeding process. Since the sequence information in the vicinity of the L^2 allele is available, it may be possible to develop markers linked to L^4 by simply comparing syntenic

sequences of resistant plants containing L^4 and susceptible plants containing L^0.

To develop SNP markers, a total of 16 primer pairs were designed using BAC sequences from the contig (Table 1). PCR using nine primer pairs (pepBAC 060I2-H1, pepBAC060I2-H3, pepBAC060I2-E1, pepBAC060I2-E2, pepBAC060I2-E3, 060I2END-1, K24D2SP6, K24D2T7, 207E13T7) resulted in multiple bands for both *C. annuum* ECW and *C. chacoense* 429. The 087H3T7 and 207E13SP6 primers amplified single bands for both genotypes. These single bands were sequenced and polymorphisms were surveyed. Nine and five SNPs were discovered for 087H3T7 (420 bp) and 207E13SP6 (422 bp) sequences respectively. For HRM analysis, primers were redesigned to amplify smaller fragments containing one or two SNPs. When HRM analyses were performed, I was able to distinguish SNPs between ECW and 429 for only 087H3T7 (Fig.4a, d). The pepBAC060I2-H2 and 290J13SP6 primers amplified very faint single bands. The 087H3SP6, 290J13T7, and 060I2END primers amplified a 700 bp single band for ECW, while no PCR products were obtained for 429. A dominant SCAR marker was developed for 060I2END based on the presence or absence of the 700 bp fragment (Fig.4c).

To test if the markers, which were polymorphic between ECW and 429 could be used for an L^4-segregating population, polymorphism and segregation of 087H3T7HRM and 060I2END markers were investigated using an SP F_2

population (Table 2). These markers showed polymorphism and segregated as expected (Fig.4a, c). However, the melting curves for 087H3T7HRM between resistant plants and 429 were slightly different, indicating sequence differences.

After confirming that the BAC contig sequences could be utilized for *L*-linked marker development, I directly compared sequences of resistant and susceptible F_2 plants and developed markers. The 158K24HRM marker was developed using the left end sequences of the BAC contig from resistant and susceptible plants in the L^4-segregating population (Fig.4b, Table 2). To test cosegregation of putative *L*-linked markers with the resistance phenotype, three *L*-segregation populations were evaluated. In the L^3 allele segregation population, there were 189 resistant and 54 susceptible plants, which fitted to an expected 3:1 model (Pvalue=0.3173). The phenotype segregation ratio was observed in two L^4 allele segregation populations—the resistance/susceptible ratios were 537:109 and 504:341 in SP and MS populations respectively. The segregation ratios of both populations were significantly deviated from the expected model of 3:1 segregation ratio (P values < 0.0001). When I tested the 087H3T7HRM marker in the L^3 population, the genotype ratio of L^3/L^3:L^3/L^0:L^0/L^0 was 57:128:58. Among the individuals, three plants were recombinant. The genotype ratio of L^4/L^4:L^4/L^0:L^0/L^0 was 222:315:107 in SP populations and the genotype ratio of L^4/L^4:L^4/L^1:L^1/L^1 was 159:299:341 in MS populations. Among them, 5 and 11

plants were recombinants in SP and MS populations, respectively.

Table 2. L^4-linked SNP markers developed based on BAC sequences

Marker	Primer name	Primer sequence (5' → 3')	Size	Type
087H3T7HRM	087H3T7150F	CATGATTACATTTTATGTTGC	150bp	Codominant
	087H3T7150R	AAAAGGAAGGTTCTCATTGTT		
060I2END	060I2END-2F	GCACATCAGCAGGTTTAGTACG	751bp	Dominant
	060I2END-2R	CCAACTGTCAAACCTCGG TT		
158K24HRM	158K24HRMF	CAGATTAAGTGTTCAAAATGAGTGATG	125bp	Codominant
	158K24HRMR	TGATTCCATGAAAATAAATTGTAAAGA		

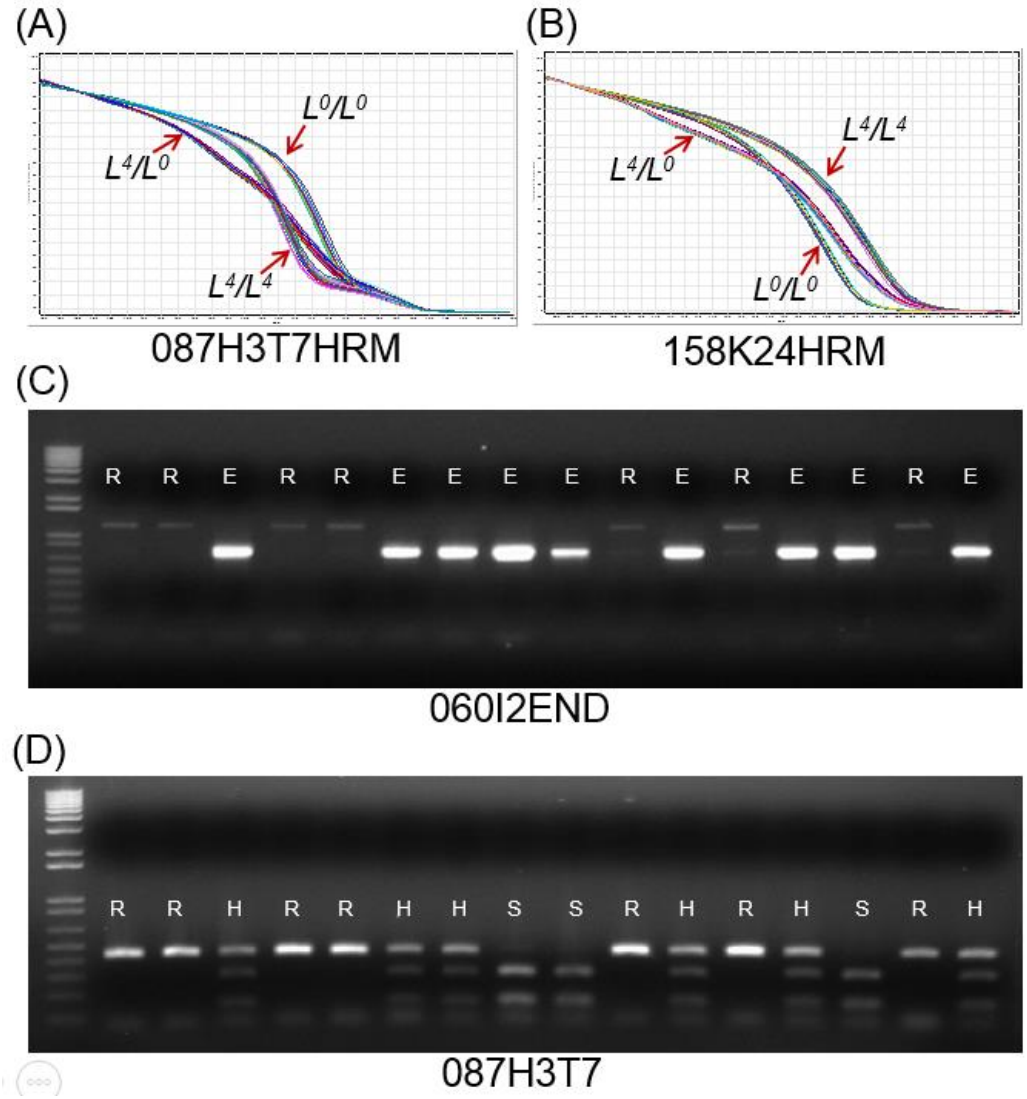


Figure 4. Analysis of molecular markers developed in this study. **a**, **b** HRM marker analysis using two markers 087H3T7HRM and 158K24HRM. There were three distinct melting curves that distinguish three genotypes. **c**, **d** Electrophoresis gels of 060I2END-2 and 087H3T7 markers that were stained by EtBr. 'R' means L^4/L^4 genotype, 'H' means L^4/L^0 genotype, 'S' means L^0/L^0 genotype, and 'E' means both L^0/L^0 and L^4/L^0 genotype

Linkage analysis and allelic relationship of *L*-linked markers

In order to have an overview of the recently published *L*-linked markers from different studies, I have mapped all the reported markers (Kim et al. 2008; Tomita et al. 2008) together with our newly developed SNP marker, 087H3T7 on a single population. To determine the relative positions of the L^4-linked markers, linkage analysis was performed using MS F2population. L4SC340 and L2kstd (Kim et al.2008) are dominant SCAR markers, which were reported to locate about 0.9–1.8 cM away from L^4. When the two markers were tested in the MS F_2 population together with 087H3T7, 11, 14, and 16 recombinants were detected for 087H3T7, L4SC340, and L2kstd respectively.

In order to determine the allelic relationship of L^3 and L^4 alleles, seven L^3-linked markers (A339, YB2A25, IH1-04, 189D23M, 197AD5R, 253A1R, and 213E3R) reported by Tomita et al. (2008) were also tested in the MS F_2 population using HRM analysis. One marker, 189D23M, demonstrated to be completely linked to L^3 in tested 2,016 individuals of an interspecific F_2 population and one recombination out of 3,391 individuals of an intraspecific F_2 population respectively (Tomita et al. 2008). Two primers sets were reported for the 189D23M marker and both were polymorphic in our mapping population. 189D23M showed six recombinants in 858 individuals of MS F_2 population. Occurrence of recombination of 189D23M was significantly higher in our L^4-

segregating population than that reported in L^3 indicates that L^3 and L^4 might not be allelic. This was further confirmed with SP F_2 population, in which 189D23M showed five recombinants out of 631 individuals.

Taken together, I was able to construct a linkage map around the *L* locus using four tested markers: L4SC340, L2kstd, 087H3T7, and 189D23M. Our results showed that the nearest markers were 087H3T7 and 189D23M located on opposite sides of the *L* locus at a distance of 1.2 and 0.8 cM, respectively (Fig.5, Table 3).

Table 3. Number of recombinants in SP, CP, and MS *L*-segregating populations

Marker	Recombinants / Total number of F_2 individuals		
	SP(L^4/L^0)	CP(L^3/L^0)	MS(L^4/L^1)
087H3T7 (CAPS) 087H3T7HRM (HRM)	5/631	4/243	11/858
L4SC340	-	-	14/858
L2kstd	-	-	16/858
189D23M	-	2/243	6/858

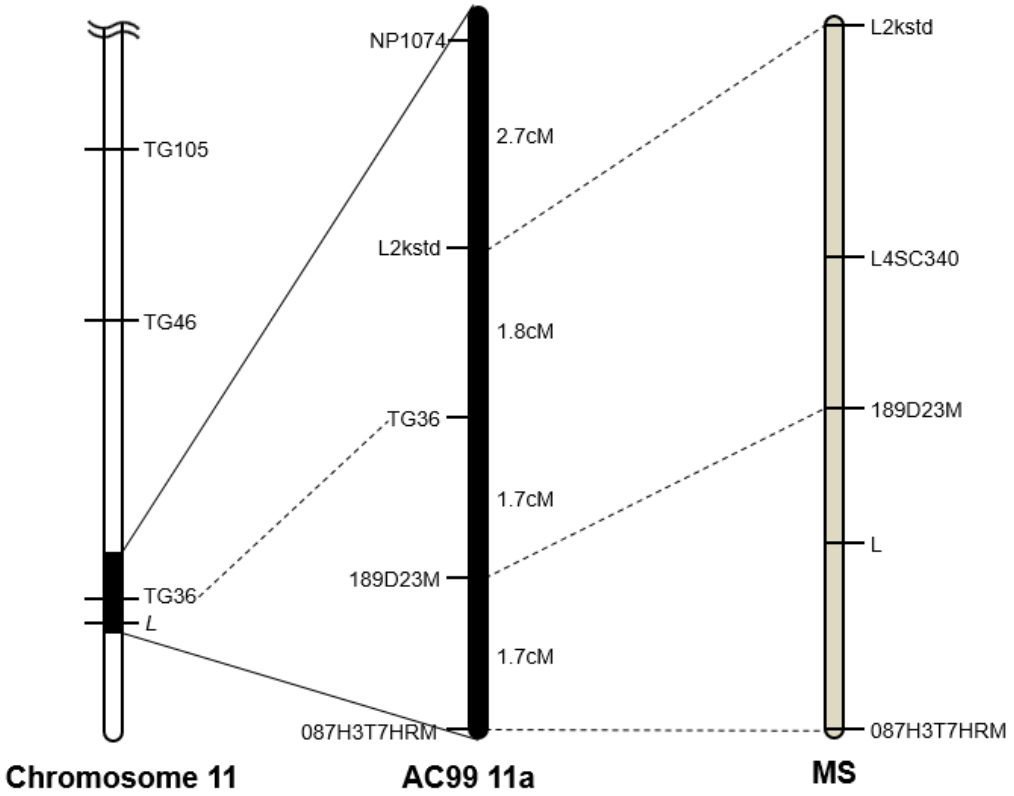


Figure 5. Genetic maps of four molecular markers (087H3T7, L4SC340, 189D23M, and L2kstd) around the L^4 locus. A linkage map of AC99 around the TG36 marker is located in the middle with Kosambi distances (black). A linkage map of MS around the L^4 locus is in the right side (gray). The nearest markers 189D23M and 087H3T7 were located in opposite sides of the L^4 locus

Gene annotation from contig sequence and the genomic relationships of RGAs

Shotgun sequence analysis revealed the insert size of BAC clones as 90,162, 105,856, and 114,328 bp for 043M10, 158K24, and 268G7, respectively, which overlapped with 060I2 draft sequence. Taken together, these three BAC clones makes a contig sequence of 223,989 bp. In order to predict the open reading frames, I used FGENESH program. A total of 42 genes were mined and were refined with the BLASTP program. Finally, 19 putative genes were predicted and were classified into seven groups (Table 4). Six putative genes were CC–NBS–LRR class resistance gene analogs (RGAs). Five out of six RGAs aligned very well with *R3a*, *I2*, and homologs of potato and tomato in BLAST search, but one gene containing an NBS domain aligned with very low significance than other five putative genes. The five RGAs were designated RIGAs (*R3a* and *I2* family Gene Analogs). Three RIGAs, contained CC, NBS, and LRR domains, were denoted RIGA-1, -2, and -3 (Fig.3c), but the other two RIGAs, contained only truncated CC or NBS domains, were denoted RIGA-4 and RIGA-5.

Except for truncated region, various motives (hhGRExE, P-loop, Walker B, RNBS-A to D, GLPL, and MHD) were conserved well in the NBS domain of RIGAs showing high homology with *I2* and *R3a* even inter region among

conserved domains (van Ooijen et al.2008). But, multiple nucleotide substitution, deletions, and insertions were existed in the whole amino acid sequences when compared to that of *R3a* and *I2*. Short deletion of RIGA-1 and RIGA-2 were existed in the NBS sequence even in the conserved motives. LRR domains were consisted of 27 subunits asI2. Interestingly, distinct amino acid differences starting from the 13th LRR domain: insertions at 13th LRR, deletions at 25th LRR regions, and many amino acid substitutions (Fig.6).

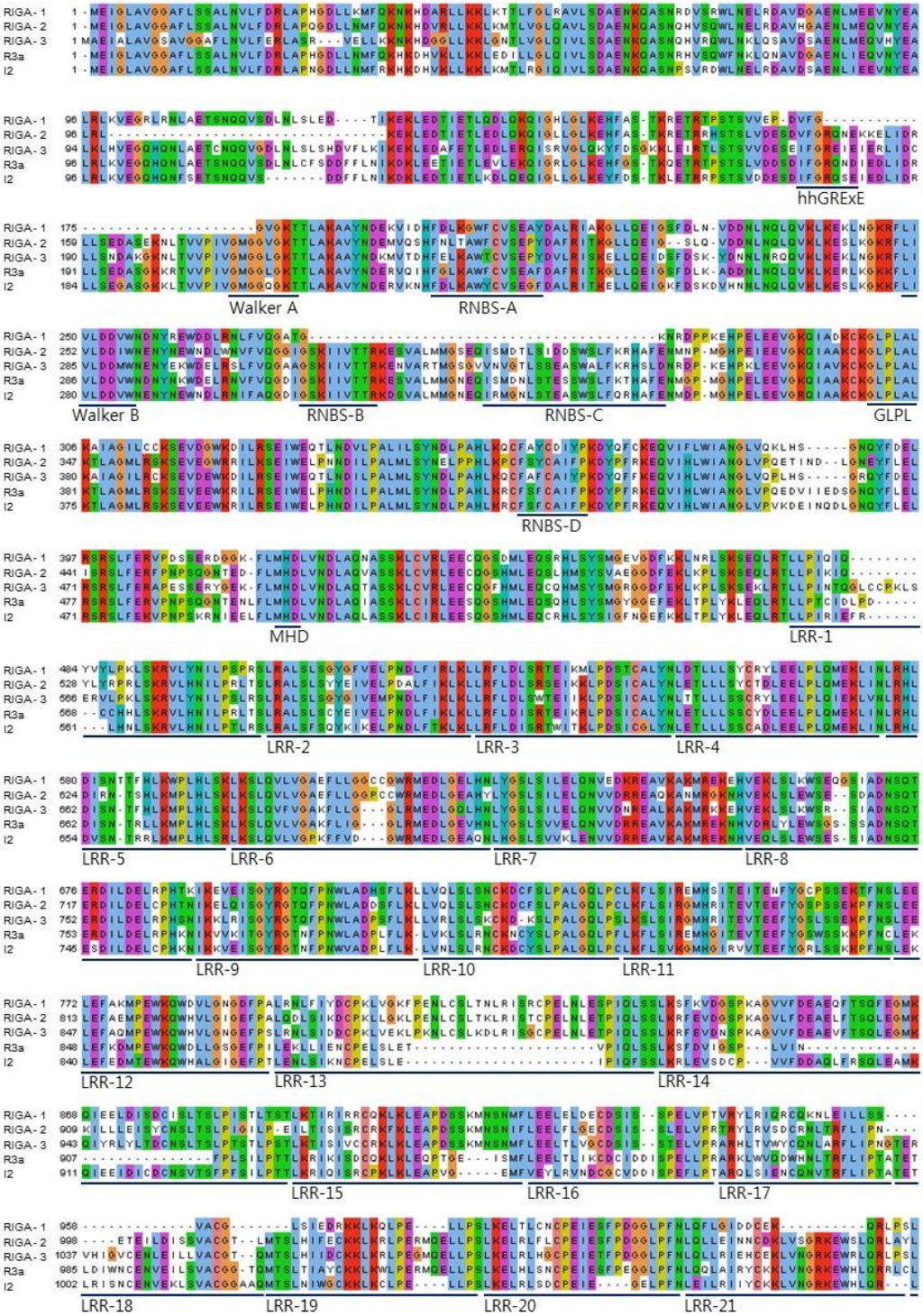
Global pairwise alignment analysis among *R3a*, *I2*, and five RIGAs was performed using Matcher program. DNA sequences of the five RIGAs shared high identities with each other (74.9–82.1% identity), and identity between *R3a* and *I2* was 83.4% with the least identities among RIGAs and *R3a* or *I2* were 68.4–77.2%. Homologies among RIGAs, *R3a*, and *I2* were higher than other resistance gene. Identities among RIGAs were higher than those between RIGAs and each of *R3a* or *I2*, but lower than those among *R3a* and *I2*.

In order to evaluate genetic relationship between RIGAs, *R3a*, and *I2*, phylogenetic analysis was conducted based on NBS region, from the P-loop motif to the GLPL motif, with other cloned resistance genes that were classified to CC–NBS–LRR in Solanaceae. Phylogenetic analysis using neighbor joining algorithm revealed RIGAs, *R3a*, and *I2* fall under the same clade, which shows the orthologous relationship among RIGAs, *R3a*, and *I2* (Fig.7a). Other

algorithms including minimum evolution, maximum parsimony, and UPGMA also confirmed that RIGAs, *R3a*, and *I2* are classified as one clade although the positions and patterns in the tip of branches in the phylogenetic clade were diverged (Fig.7b).

Table 4. Summary of gene predictions on the BAC contig sequence

Gene prediction	Number
ABC transporter family protein	1
DVF 1191 super family (unknown function)	2
CC-NBS-LRR	3
Incomplete CC-NBS-LRR	3
Retrotransposable elements with LTR	5
Retrotransposable elements without LTR	2
Incomplete retrotransposable elements	3


RIGA-1
RIGA-2
RIGA-3
R3a
I2
hhGRExE
Walker A
RNBS-A
Walker B
RNBS-B
RNBS-C
GLPL
RNBS-D
MHD
LRR-1
LRR-2
LRR-3
LRR-4
LRR-5
LRR-6
LRR-7
LRR-8
LRR-9
LRR-10
LRR-11
LRR-12
LRR-13
LRR-14
LRR-15
LRR-16
LRR-17
LRR-18
LRR-19
LRR-20
LRR-21

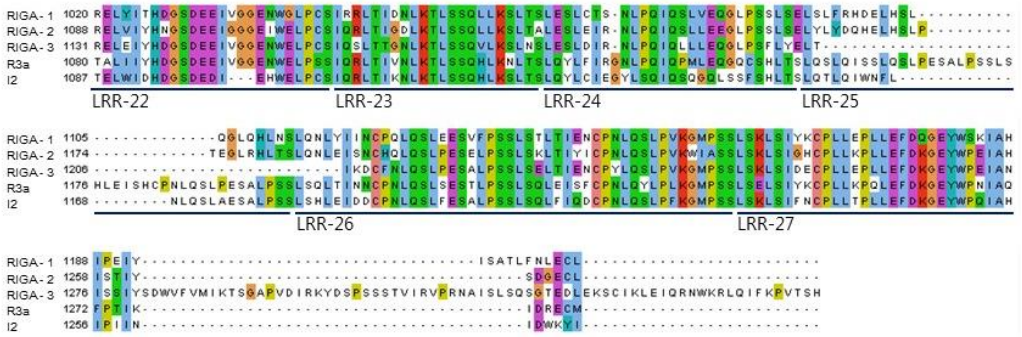


Figure 6. Multiple alignment of the *R3a*, *I2*, and three deduced RIGAs. Conserved motives in the NBS region and LRR domains are denoted with underlines as described by Ori et al. (1997) and Van Ooijen et al. (2008).

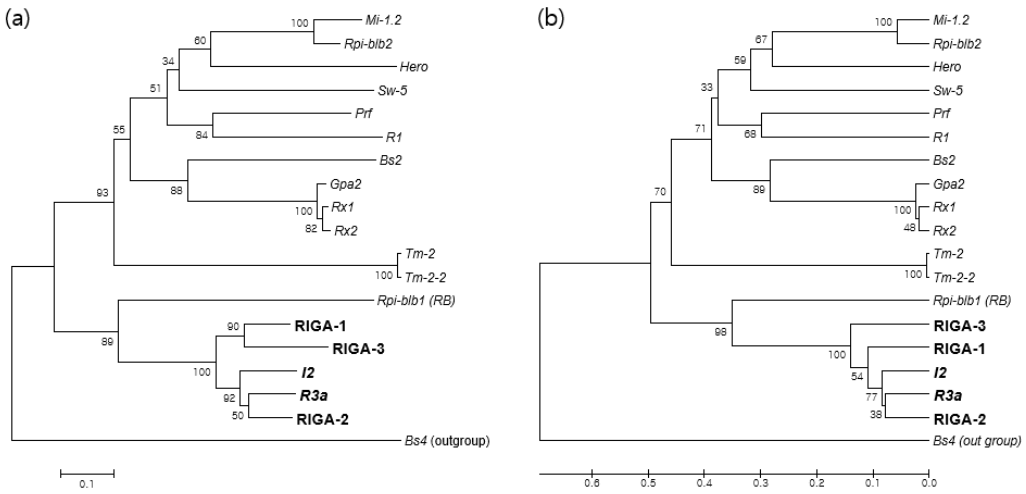


Figure 7. Phylogenetic analysis of RIGAs, *R3a*, and *I2* with other resistance genes in Solanaceae by neighbor-joining (a) and UPGMA (b) algorithm. *Bs4* is the resistance gene belonging to TIR–NBS–LRR class and was used as an out-group

DISCUSSION

Based on the hypothesis that the *L* gene in pepper conferring *Tobamovirus* resistance is an ortholog of the *R3* gene in potato and the *I2* gene in tomato conferring resistance against *Phytophthora infestans* and *Fusarium* wilt, respectively, molecular markers closely linked to the *L* locus were developed in this study by identification of *R3*/*I2* orthologs in pepper.

L^4 confers resistance to the most aggressive PMMoV pathotype P1.2.3. Development of reliable markers is important for introgression of L^4 to elite germplasm through marker-assisted-selection (MAS)/marker-assisted-backcrossing (MAB). Therefore, I have defined recently published markers linked to *L* locus together with our newly developed SNP markers (Tomita et al.2008; Kim et al. 2008) to construct a localized map around the *L* locus. Interestingly, I found that molecular markers developed for L^3 (*C. chinense*) (Tomita et al. 2008) are also applicable to L^4, although L^4 and L^3 may not be allelic. I have identified new markers specifically linked to L^4, 189D23M (Tomita et al. 2008) and our SNP marker, 087H3T7HRM, flanks the L^4 locus. Use of two markers simultaneously in a molecular breeding program is recommended; successful rate of introgression of L^4 is expected to improve when compared to the previous system using a single SCAR marker (Kim et al.2008). In addition,

our SNP markers have an additional advantage, i.e., capable of differentiating L^3 and L^4; breeders could introgress the two resistance genes to different lines and pyramid the two genes by making F_1 hybrids.

Earlier comparative genetic studies revealed the colinearity between the *R3* late blight resistance locus in potato and the *I2* locus in tomato. The perfect microsyntenic relationship between tomato and potato on chromosome 11 allowed the isolation of the potato *R3a* gene by a comparative genomics approach using the resistance gene analog of *I2*. A macrosyntenic comparison between pepper and potato/tomato illustrated a potential orthology of *L* and *R3*/*I2*. However, since the syntenic region on pepper chromosome 11 has been demonstrated to be one of the 22 genome rearrangements distinguishing tomato and pepper (Livingstone et al. 1999), it was less likely that the *L* region could be characterized by examining microsynteny and comparative gene content with the corresponding regions. Because of the complexity, instead of using the tomato and potato physical maps for fine mapping and physical mapping of *L* directly, I prioritized and employed PCR approaches using portions of the *R3* gene to identify *R3*/*I2* orthologs at the target syntenic region near TG36 in pepper (where L was previously mapped near to). With this approach, I successfully found BAC clones closely linked to *L*, other clones identified but mapped elsewhere in the genome are considered as paralogs.

Mazourek et al. (2009) demonstrated the association between R-genes with chromosome breakpoints of genome rearrangement, characteristics of the syntenic region of chromosomes 11 in pepper, tomato, and pepper further supports this observation. Tomato and potato are collinear throughout the lower arm of the chromosome 11. *I2* in tomato is located between TG36 and TG105, and *R3a* is located near TG105 with TG105 nearer to the telomere (Fig.1). Pepper is collinear with both tomato and potato in this region except TG105 is centromeric in pepper with TG36 facing the telomere (Paran et al.2004) (Fig.1). L together with disease resistance QTL against cucumber mosaic virus and *Phytophthora* spp. mapped in cluster near TG36 toward the telomere of chromosome 11 in pepper (Grube et al. 2000; Livingstone et al. 1999), apparently associates with (or located at or near) the breakpoint of chromosomal rearrangement.

I have obtained 224 kb sequences from four BAC clones, and only five *R3*/*I2* RGAs were predicted. By comparison, at the syntenic region in potato, five *I2* RGAs were identified within one BAC clone where the *R3a* gene lies (Huang et al.2005). Three resistance gene analogs (RGAs) of *R3*/*I2* with CC, NBS, and LRR domains were predicted and located at the syntenic region in pepper. Alignment of the three RGA proteins showed high similarity, indicating that the RGAs are paralogs. The roles of these RGAs are not clear. Base on our

phenotypic and genetic analysis, recombination occurs between *L* resistance and our markers, indicating that the RGAs found in this study might not be candidate genes for *L*. However, I cannot totally rule out a possible connection to resistance against *Tobamovirus*. In fact, *I2C-1* has been shown to contribute partial resistance against *Fusarium oxysporum* in tomato (Ori et al. 1997). Alternatively, these three RGAs might be functional candidates for other disease resistance genes or quantitative trait loci (QTL) associated with disease resistance. This is not surprising as it has been known that *L* is located in a R-gene cluster where disease resistance QTLs for potyvirus, *Cucumber mosaic virus* (CMV), *Tomato spotted wilt virus* (TSWV), and *Phytophthora capsici* have been reported (Lefebvre2004).

Genetic inheritance studies have demonstrated the allelic relationship of L^1, L^2, L^3, and L^4. However, our study did not confirm the expected classical allelic relationship at the molecular level. I found consistent contradiction between our study and a recent study on fine mapping of L^3. Tomita et al. (2008) also found the presence of *I2* homologs linked to L^3. However, surprisingly, a genetic distance of 0.83 cM from *L* was obtained when I tested Tomita's L^3-linked markers in our L^4-segregating populations, an almost 1.4-fold difference from what Tomita et al. (2008) found. However, it might be possible that the difference in genetic distance between L^3 and L^4 was due to occurrences of

unequal cross-over in interspecific crosses between *C. annuum*/*C. chacoense* and *C. annuum*/*C. chinense* that might have obscured the allelic relationship. In addition, use of different segregating populations and sampling error might have played a role in the differences. Alternatively, L^1, L^2, L^3, and L^4 might be different genes closely linked within the region instead of different alleles at the same locus. In fact, two closely linked genes with distinct specificities have been found for *R3* resistance to late blight in potato including the cloned *R3a* (Huang et al. 2004, 2005). Furthermore, a BAC library was developed from *C. frutescens* in this study; however, gene content and/or genome structure might be different between *C. chinense* (L^3) and *C. chacoense* (L^4), as was demonstrated for the R1 contig conferring resistance to Phytophthora infestans on chromosome V in potato, where gene content and genome structure varied between different haplotypes (Ballvora et al.2007).

Different segregation ratios were obtained for the two L^4-segregating populations. In both populations, distorted 3:1 segregation ratios were obtained, indicating preferential segregation has occurred in the region and resulting in segregation distortion. It is known that segregation distortion commonly occurs in interspecific crosses. For instance, segregation distortion has been found in other regions of chromosome 11 for crosses between *C. annuum* and *C. chinense* (L^3). However, interspecific crosses between *C. annuum* and *C. chacoense* (L^4)

have not been studied. By comparison, such segregation distortion in this region is not seen in the *C. annuum*/*C. chinense* (L^3)-segregating population, as the marker segregated according to the Mendelian pattern.

REFERENCES

Alonso E, Garcia-Luque I, de la Cruz A, Wicke B, AvilaRincon MJ, Serra MT, Castresana C, Diaz-Ruiz JR (1991) Nucleotide sequence of the genomic RNA of *Pepper mild mottle virus*, a resistance-breaking tobamovirus in pepper. Journal of General Virology 72:2875–2884

Ballvora A, Ercolano MR, Weiss J, Meksem K, Bormann CA, Oberhagemann P, Salamini F, Gebhardt C (2002) The R1 gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. The Plant Journal 30:361–371

Ballvora A, Jo¨cker A, Vieho¨ver P, Ishihara H, Paal J, Meksem K, Bruggmann R, Schoof H, Weisshaar B, Gebhardt C (2007) Comparative sequence analysis of Solanum and Arabidopsis in a hot spot for pathogen resistance on potato chromosome V reveals a patchwork of conserved and rapidly evolving genome segments . BMC genomics 8:112–124

Berzal-Herranz A, de la Cruz A, Tenllado F, Diaz-Ruiz JR, Lopez L, Sanz AI, Vaquero C, Serra MT, Garcia-Luque I (1995) The *Capsicum* L^3 gene-mediated resistance against the tobamoviruses is elicited by the coat protein. Virology 209:498–505

Boukema IW (1980) Allelism of genes controlling resistance to TMV in *Capsicum* L. Euphytica 29:433–439

Csillery G, Tobias I, Rusko J (1983) A new pepper strain of tomato mosaic virus. Acta Phytopathologica Academiae Scientiarum 18:195–200

Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833.

de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T (2005) Cartha gene: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics 21:1703–1704

de la Cruz A, Lopez L, Tenllado F, Diaz-Ruiz JR, Sanz AI, Vaquero C, Serra MT, Garcia-Luque I (1997) The coat protein is required for the elicitation of the Capsicum L^2 gene-mediated resistance against the tobamoviruses. Molecular Plant-Microbe Interactions 10:107–113

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

Huang S, Vleeshouwers VG, Werij JS, Hutten RC, van Eck HJ, Visser RG, Jacobsen E (2004) The *R3* resistance to *Phytophthora infestans* in potato is conferred by two closely linked R genes with distinct specificities. Molecular Plant-Microbe Interactions 17:428–435

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. The Plant Journal 42:251–261

Kim HJ, Han JH, Yoo JH, Cho HJ, Kim BD (2008) Development of a sequence characteristic amplified region marker linked to the L^4 locus conferring broad spectrum resistance to tobamoviruses in pepper plants. Molecules and Cells 25: 205–210

Kuang H, Wei F, Marano MR, Wirtz U, Wang X, Liu J, Shum WP, Zaborsky J, Tallon LJ, Rensink W, Lobst S, Zhang P, Tornqvist CE, Tek A, Bamberg J, Helgeson J, Fry W, You F, Luo MC, Jiang J, Buell CR, Baker B (2005) The R1 resistance gene cluster contains three groups of independently evolving, type IR1 homologues and shows substantial structural variation among haplotypes of *Solanum demissum*. The Plant Journal 44:37–51

Lefebvre V (2004) Molecular markers for genetics and breeding: development and use in pepper (*Capsicum* spp.). In: Lo ¨rz H, Wenze G (eds) Biotechnology in agriculture and forestry molecular marker systems. Springer, Heidelberg, pp 189–214

Lefebvre V, Palloix A, Caranta C, Pochard E (1995) Construction of an intra-specific integrated linkage map of pepper using molecular markers and doubled-haploid progenies. Genome 38:112–121

Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK (1999) Genome mapping in *Capsicum* and the evolution of genome structure in the solanaceae. Genetics 152:1183–1202

Matsunaga H, Saito T, Hirai M, Nunome T, Yoshida T (2003) DNA markers linked to *Pepper mild mottle virus* (PMMoV) resistant locus (L^4) in *Capsicum*. Japanese Society for Horticultural Science 72:218–220

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

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. The Plant Journal 20:317–332

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. The Plant Cell 15:809–834

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

Ori N, Eshed Y, Paran I, Presting G, Aviv D, Tanksley S, Zamir D, 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. The Plant Cell 9:521–532

Paran I, van der Voort JR, Lefebvre V, Jahn M, Landry L, van Schriek M, Tanyolac B, Caranta C, Chaim AB, Livingstone K, Palloix A, Peleman J (2004) An integrated genetic linkage map of pepper (*Capsicum* spp.). Molecular Breeding 13:251–261

Pernezny K, Roberts PD, Murphy J, Goldberg NP (2003) Compendium of pepper diseases. APS Press, St. Paul Song J, Bradeen JM, Naess SK, Raasch JA, Wielgus SM, Haberlach GT, Liu J, Kuang H, Austin-Phillips S, Buell CR, Helgeson JP, Jiang J (2003) Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. Proceedings of the National Academy of Sciences of the United States of America 100:9128–9933

Sugita T, Yamaguchi K, Sugimura Y, Nagata R, Yuji K, Kinishita T, Todoroki A (2004) Development of SCAR markers linked to L^3 gene in *Capsicum*. Breeding Science 54:111–115

Takeuchi S, Hamada H, Toyoda K, Suzuki K, Kiba A, Hikichi Y, Okuno T (2005) Discrimination between tobamoviruses and their pathotypes for *L*-gene-mediated resistance in green pepper (*Capsicum annuum* L.) by reverse transcription-polymerase chain reaction. Journal of General Plant Pathology 71:60–67

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596–1599

Tomita R, Murai J, Miura Y, Ishihara H, Liu S, Kubotera Y, Honda A, Hatta R, Kuroda T, Hamada H, Sakamoto M, Munemura I, Nunomura O, Ishikawa K, Genda Y, Kawasaki S, Suzuki K, Meksem K, Kobauashi K (2008) Fine mapping and DNA fiber FISH analysis locates the tobamovirus resistance gene L^3 of *Capsicum chinense* in a 400-kb region of R-like genes cluster embedded in highly repetitive sequences. Theoretical and Applied Genetics 117:1107–1118

Tsuda S, Kirita M, Watanabe Y (1998) Characterization of a *Pepper mild mottle tobamovirus* strain capable of overcoming the L^3 gene-mediated resistance, distinct from the resistance-breaking Italian isolate. Molecular Plant-Microbe Interactions 11:327–331

van der Vossen E, Sikkema A, Hekkert BL, Gros J, Stevens P, Muskens M, Wouters D, Pereira A, Stiekema W, Allefs S (2003) An ancient R gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. The Plant Journal 36:867–882

van der Vossen EAG, Gros J, Sikkema A, Muskens M, Wouters D, Wolters P, Pereira A, Allefs S (2005) The *Rpi-blb2* gene from *Solanum bulbocastanum* is an *Mi-1*gene homolog conferring broad-spectrum late blight resistance in potato. The Plant Journal 44:208–222

van Ooijen G, Mayr G, Kasiem MMA, Albrecht M, Cornelissen BJC, Takken FLW (2008) Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. Journal of Experimental Botany 59:1383–1397

Velasco L, Janssen D, Ruiz-Garcia L, Segundo E, Cuadrado IM (2002) The complete nucleotide sequence and development of a differential detection assay for a *Pepper mild mottle virus* (PMMoV) isolate that overcomes L^3 resistance in pepper. Journal of Virological Methods 106:135–140

Wetter C, Conti M, Altschuh D, Tabillion R, Van Regenmortel M (1984) *Pepper mild mottle virus*, a Tobamovirus infecting pepper cultivars in Sicily. Phytopathology 74:405–410

Wisser RJ, Sun Q, Hulbert SH, Kresovich S, Nelson RJ (2005) Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. Genetics 169:2277–2293

Wisser RJ, Balint-Kurti P, Nelson RJ (2006) The genetic architecture of disease resistance in maize: a synthesis of published studies. Phytopathology 96:120–129

Yi G, Lee JM, Lee S, Choi D, Kim BD (2006) Exploitation of pepper EST-SSRs and an SSR-based linkage map. Theoretical and Applied Genetics 114:113–130

CHAPTER II

Development and validation of the *L* allele specific markers in *Capsicum*

ABSTRACT

Tobamovirus is one of most destructive viruses in *Capsicum*. Accordingly, the *L* locus, a resistance gene against Tobamovirus, has been used for pepper breeding programs. Previously, the L^3 gene, one of the *L* alleles, was isolated through map-based cloning, and a L^4 gene candidate was isolated by homology-based PCR methods. Here, the L4segF&R marker was developed based on the LRR region of the L^4 candidate, and co-segregation analysis was performed using two L^4-segregating F_2 populations derived from the commercial cultivars Special and Myoung-sung. The L4segF&R marker was located within 0.3 cM of the L^4 gene but did not completely co-segregate with the L^4 gene, indicating that the

candidate is not actually L^4. To confirm the mapping result, L4segF&R genotypes of L^4-containing breeding lines from three different seed companies were analyzed, resulting in the identification of several recombinants in the breeding lines. Based on these results, I postulate several genetic models that show different introgression histories and genetic structures for the L^4-containing segment in different breeding lines. All of the models demonstrate that resistance conferred by the L^4 segment could not be explained by the L^4 gene candidate alone. Although the presence of the L^4 gene candidate could not fully explain the L^4 resistance, I was able to develop allele-specific markers for the *L* locus using the candidate sequence. To develop allele-specific markers for the *L* locus, HRM analysis was performed using primer pairs based on the LRR sequence of the L^4 gene candidate. When commercial breeding lines homozygous for L^0, L^1, L^2, L^3 or L^4 were analyzed, L4RP-3F/L4RP-3R correctly detected the *L* allele in 90 out of 91 lines. I believe that the *L* allele-specific marker developed in the study provides a solution for pepper breeders developing improved resistance lines against *Tobamovirus*.

INTRODUCTION

The genus *Tobamovirus* is one of most destructive virus groups in *Capsicum* causing considerable yield loss in sweet, hot and ornamental pepper species worldwide. The *L* locus in *Capsicum* confers resistance against *Tobamovirus* such as *Tobacco mosaic virus* (TMV), *Tomato mosaic virus* (ToMV), *Paprika mild mottle virus* (PaMMV), and *Pepper mild mottle virus* (PMMoV). Genetic factors at the *L* locus have been demonstrated to induce the hypersensitive response (HR) by recognition of the coat protein of *Tobamovirus* (Berzal-Herranz et al. 1995; de la Cruz et al. 1997). One susceptible allele (L^0) and four *L* alleles (L^1, L^2, L^3, and L^4) have been identified and classified according to resistance responses to different *Tobamovirus* pathotypes such as ToMV (P_0), PaMMV (P_0, P_1), and PMMoV of ($P_{1.2}$ and $P_{1.2.3}$) (Berzal-Herranz et al. 1995; Boukema 1980; de la Cruz et al. 1997; Tomita et al. 2011). L^1, L^2, L^3 and L^4 were identified in *C. annuum* such as cv. 'Bruinsma Wonder' and cv. 'Verbeterde Glas', *C. frutescens* cv. 'Tabasco', *C. chinense* 'PI 159236', and *C. chacoense* 'PI 2604429', respectively. Additionally, different temperature dependent alleles L^{1a}, L^{1c}, and L^{2b} were found in *C. annuum* cv. 'KC780', *C. chinense* 'KC667' and *C. baccatum* 'PI 439381-1-3', respectively (Tomita et al. 2011). Among the known alleles, L^4 is considered to have the broadest resistance spectrum, mediating

resistance against P_0, P_1, $P_{1.2}$, and $P_{1.2.3}$ pathotypes of PMMoV although it is overcome by the $P_{1.2.3.4}$ pathotype (Genda et al. 2007). The $P_{1.2.3.4}$ pathotype has been found only in Japan, but no source of resistance has yet been discovered.

Different approaches have been used for the development of molecular markers linked to the *L* locus. An earlier study developed a randomly amplified polymorphic DNA (RAPD) marker linked to the L^3 gene with a genetic distance of 4.0 cM (Sugita et al. 2004). More recently, tightly linked markers, including the 189D23M marker located within 0.1cM of the L^3 gene, were developed by bulked segregation analysis – amplified fragment length polymorphism (BSA-AFLP) and resistance gene analog (RGA) approaches (Tomita et al. 2008). In addition, SCAR markers converted from RAPD and AFLP markers have been developed 1.5 cM and 0.9 - 1.8 cM from the L^4 gene (Kim et al. 2008; Matsunaga et al. 2003). Via comparative analysis of *L* with *R3a* in potato and *I2* in tomato, several markers were developed from BAC sequences containing *R3a* and *I2* homologues (Yang et al. 2009). Among the SNP markers, 087H03T7 is located about 1.5cM from L^4 (Yang et al. 2009). Mapping experiments further demonstrated that the L^3-linked marker 189D23M and the L^4-linked marker 087H03T7 are located on opposite sides of the L^4 gene. Linkage analysis demonstrated that the linkage of 189D23M to L^4 was not as tight as that to L^3 (Yang et al. 2009).

The L^3 gene was isolated through map-based cloning (Tomita et al. 2011). Like many other disease resistance genes (R genes) isolated in plants, L^3 is belonged to the major R gene class consisting of NB-ARC and LRR domains (Martin et al. 2003; van Ooijen et al. 2007). The same group isolated candidate genes of other *L* alleles, L^1, L^{1a}, L^{1c}, L^2, L^{2b}, and L^4 through a homology-based PCR method (Tomita et al., 2011). *In vivo* interaction has been demonstrated between *L* candidate genes and the coat protein of each corresponding *Tobamovirus* pathotype, P_0, P_1, $P_{1.2}$, $P_{1.2.3}$, and $P_{1.2.3.4}$, in *Nicotiana benthamiana*. However these *L* candidates were not confirmed by genetic analysis.

The highly concerved NB-ARC domain is found in NBS-LRR R proteins in plants, which are encoded by genes located in clusters with many resistance gene analogs (RGAs) (Meyers et al. 1999; Michelmore et al. 1998). The conserved nature of the NB-ARC domain has been exploited for RGA marker development using degenerate primers and NBS profiling (van der Linden et al. 2004; Yu et al. 1996). RGA maps have been constructed to identify disease resistance linked markers in plants (Madsen et al. 2003; Pankovic et al. 2007; Sutherland et al. 2008). By contrast, LRR domains are highly variable as compared to the NB-ARC domain; marker development on the LRR domain is not common due to the lack of conserved sequences. Allele-specific markers, on

the other hand, can be developed from genetic regions such as that encoding the LRR domain where the nucleotide variation between different genotypes is high.

In this study, I developed the L^4 gene candidate based markers and carried out linkage analysis for the L^4 candidate using two F_2 segregating populations and breeding lines of several commercial seed companies. Linkage analysis revealed that the L^4 candidate did not completely co-segregated with the L^4 gene and that there are complicated genetic architectures in the *L* region. Furthermore, I developed allele specific markers for the *L* locus using the LRR domain that can distinguish between different *L* alleles. These markers will be very useful for *Tobamovirus* resistance breeding.

MATERIALS AND METHODS

Plant materials

Capsicum plants containing different *L* alleles (L^0, L^1, L^2, L^3, and L^4) were used for development of *L* allele specific molecular markers. *C. annuum* cv. 'Early California Wonder', *C. frutescens* cv. 'Tabasco', *C. chinense* 'PI 159236', and *C. chacoense* 'PI 260429' homozygous for L^0, L^2, L^3, and L^4 were used as indicator plants. The L^1 indicator plant was selected from F_2 populations derived from self-pollination of *C. annuum* cv. 'Myoung-sung' (MS) (Monsanto Korea, Chungwon, Korea) and *C. annuum* cv. 'Speical' (SP) (Enza Zaden, Enkhuizen, The Netherlands).

Breeding materials used to test the L^4 candidate gene and to determine genetic structures around the *L* locus were provided by commercial seed companies including De Ruiter (Amstelveen, The Netherlands), Monsanto Korea (Chungwon, Korea), and Nongwoo Bio (Suwon, Korea). Breeding lines of Enza Zaden were further used for validation of the *L* allele specific markers.

Virus strains and resistance screening

Tobamovirus pathotypes $P_{1.2.3}$ and P_0 were used for resistance screening of MS and SP F_2 populations, respectively. These two strains were multiplied in *N. benthamiana*. Virus resistance of F_2 plants was determined by observing the

hypersensitive response (HR) and was further confirmed by ELISA analysis. The L^1 indicator plants were selected by detached-leaf inoculation of *Tobamovirus* $P_{1.2.3}$ in F_2 populations derived from MS and SP (L^4/L^1). Individual plants no showing the HR in these populations were selected as L^1/L^1 indicator plants. Phenotype screening of commercial breeding lines was carried out in-house by each individual seed company. Progeny tests using ten to twenty $F_{2:3}$ individuals were performed for *Tobamovirus* resistance screening of Monsanto Korea F_2 breeding populations.

Genomic DNA extraction

Total genomic DNA of mapping population was extracted from leaf tissue with a hexadecyltrimethyl-ammonium bromide (CTAB) procedure (Kang et al., 2010). DNA samples were used for marker analysis.

DNA sequence analysis

PCR products were purified using the Zymoclean™ gel DNA Recovery kit (Zymo Research, USA) and sequencing was carried out by the National Instrumentation Center for Environmental Management (Seoul National University, Korea). Two programs were used for sequence analysis. The BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to perform local

nucleotide sequence alignment analysis, and Clustal W2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) was used to perform global nucleotide sequence alignment analysis.

SNP marker development and analysis

SNP markers were developed based on the sequences of the L^3 gene and other *L* candidates (L^1, AB523372; L^{1a}, AB523373; L^2, AB523375; L^3, AB523370; L^4, AB523377, GenBank). SNPs located on LRR region were identified and primers were designed manually to amplify DNA fragment including SNPs (Fig. 1; Table 1).

The high resolution melting (HRM) method was used to analyze molecular markers linked to the L^4 locus, and to analyze the genotype of individual plant, with Rotor-gene Q (QIAGEN, Germany). PCR was performed in 20 μl including 60 mM KCl, 10 mM Tris-Cl, 2.5 mM $MgCl_2$, 0.25 mM each dNTP, 5 pmol each primer, 1 unit *Taq* DNA polymerase, 1.25 M Syto9, and 50 ng genomic DNA template. The PCR program was 94°C for 1 min, (94 °C for 20 sec, 58-60 °C for 20 sec, 72 °C for 30 sec) X 40 cycles, 72°C for 5 min. Fluorescence signals were measured each 0.1°C from 70°C to 90°C and melting curve analysis was carried out using the operating software of manufacturer.

Table 1. Primers used in this study

Primer name	Primer sequence (5’ → 3’)	Reference
L4segF	TGTGAGAATCTTGAAATATTTTCGG	-
L4segR	CTTTGCGCAAGAGTGGATATTC	-
L4RP-1F	ATGTGGGACCCAGATGACGT	-
L4RP-1R	TCCCAATTCTCACCACCAACA	-
L4RP-2F	CACGATGGCAGTGACGAAGAGAT	-
L4RP-2R	GGGCAATTCCAGATCAATAGACT	-
L4RP-3F	TCTTCAGCACCTCAATTCGGTTC	-
L4RP-3R	GAAGAGGGCATCCCTTTTACT	-
3'endR	TCACAGGCATTCACAGTCAAACATAGTGCGACC	-
189D23MF	ATTGTCAGAGTCGGGAAGCA	Tomita et al. 2008
189D23MR	AACGACAAGGGTTTATTGTATGC	
087H03T7HRMF	CATGATTACATTTTATGTTGC	Yang et al. 2009
087H03T7HRMR	AAAAGGAAGGTTCTCATTGTT	

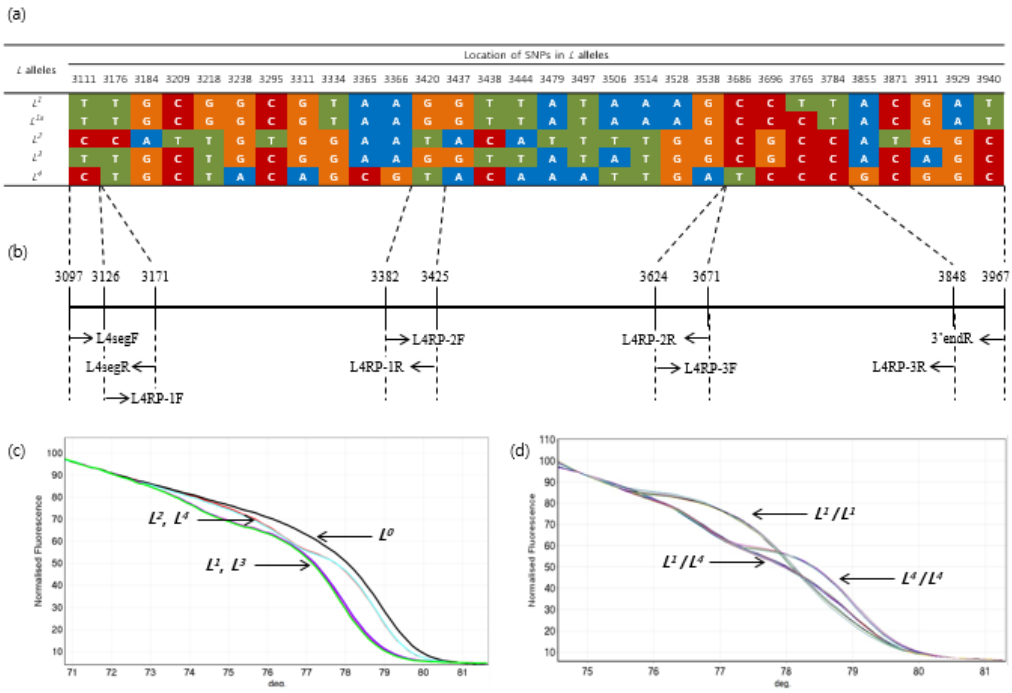


Figure 1. (a) Single nucleotide polymorphisms of L^3 and four L candidates (L^1, AB523372; L^{1a}, AB523373; L^2, AB523375; L^3, AB523370; L^4, AB523377, GenBank). SNPs of L^1 (SP) and L^1 (MS) corresponded with those of L^1 and L^{1a} candidates, respectively. (b) Black arrows show the position and direction of nine primers used in this study. Numbers above the black bars are sequences for which each primer was designed in the L^4 candidate. (c) Melting curves patterns of the L4segF&R marker. L4seqF&R showed three different melting curve patterns for 5 alleles. (d) Melting curves patterns of the L4segF&R marker in SP and MS populations.

Linkage analysis of molecular markers linked to the *L^4* gene

Co-segregation analysis was performed using SP, MS F_2 populations and breeding lines with molecular markers linked to the *L^4* gene (Table 1). Linkage analysis of molecular markers linked to the *L* locus was performed using Carthagene 1.0 program (de Givry et al. 2005) with a LOD score threshold of 4.0 and maximum distance of 30 cM.

RESULTS

Development of markers based on the L^4 candidate gene

One HRM marker set L4segF/L4segR, designated L4segF&R, was developed based on the sequence located on the 5' LRR domain of the L^4 candidate (GenBank AB523377, Tomita et al. 2011) to discriminate between the *L* alleles in commercial breeding lines and F_2 populations (Fig. 1). This marker amplifies an 80 bp DNA fragment from all *L* alleles. When HRM analysis was performed with L4segF&R using DNA from indicator plants, this marker showed 3 melting curve patterns for five *L* alleles: i) for the L^0 allele, ii) for the L^2 and L^4 alleles, and iii) for the L^1 and L^3 alleles, respectively (Fig. 1). Melting curve patterns were perfectly matched with the sequences of each allele. L4segF&R was further tested using two F_2 populations segregating L^1 and L^4 alleles. Genotypes of individuals of MS and SP F_2 populations could be clearly discriminated and showed the expected melting curve patterns such as L^1/L^1, L^1/L^4, or L^4/L^4 (Fig. 1).

Linkage analysis of the L^4 candidate

Linkage analyses were performed using SP and MS F_2 populations to test if L4segF&R perfectly co-segregated with the virus resistance. The same

populations were used to develop markers and construct a local linkage for the L^4 gene (Yang et al., 2009). Fdour molecular markers linked to the *L* locus, 087H03T7HRM, 189D23M, L4SC340, and L2kstd, were mapped in the MS F_2 population, whereas 087H03T7HRM and 189D23M were mapped in the SP F_2 population. 189D23M was the closest marker linked to the L^4 gene with a genetic distance of 0.8 cM (Kim et al., 2008; Tomita et al. 2008; Yang et al., 2009).

When a total of 631 and 858 individual plants were analyzed with L4segF&R, two and three recombinants were found in the SP and MS populations, respectively (Table 2, Table 3 type A and C). These results demonstrated that the L^4 candidate gene is located about 0.3 cM away from the resistance gene. A total of 361 commercial breeding lines provided by seed companies were also tested using L4segF&R and three recombinants were also identified in those lines (Table 2). Taken together, our results showed that the L^4 candidate might not be the L^4 gene controlling *Tobamovirus* resistance and/or that another genetic factor may be required for resistance against TMV in the tested populations.

Table 2. Analysis of cosegregation of three *L*-linked markers with *Tobamovirus* resistance

Company	Population	No of recombinants / Total No of tested plants		
		L4segF&R	189D23M	087H03T7HRM
Enza Zaden	Special F_2	2 / 631	2/ 631	2/ 631
Monsanto Korea	MyungSung F_2	3 / 858	8 / 858	11 / 858
De Ruiter	Progenygroups	2 / 249	39 / 121[z]	39 / 121[z]
Monsanto Korea	F_2 Breeding populations	0 (54)[y] / 90	0 (54)[y] / 90	3 (53)[y] / 90
Nongwoo	Breeding lines	0 (8)[y] / 22	0 (8)[y] / 22	9 (8)[y] / 22

[z] Individuals showing different genotypes for these markers in Special and MyungSung F_2 populations were not counted in these numbers

[y] Numbers in the parenthesis indicate recombinants identified by progeny test

Table 3. L^4-linked marker genotypes of recombinants

Recombinant type	Resistance phenotype	Genotype of L^4-linked marker			No. of plants
		087H02T7	L4segF&R	189D23M	
A	R[z]	S	S	S	1
B	R[z]	R	S	R	2
C	S	H	H	H	4
D	R[z]	R	R	R	37
E	R[z]	H	R	R	12
F	R	H	H	H	19

R : homozygous resistance, S : homozygous susceptible

[z] Resistance which was turned out to have the heterozygous genotype of L4segF&R by progeny tests.

Analysis of breeding lines using markers flanking the *L^4* gene

To investigate further the genetic structures of the *L^4* locus in different plant materials, genotypes of *L^4* flanking markers were analyzed using commercial breeding lines obtained from different seed companies. In additions to L4segF&R, I also used 189D23M and 087H03T7, which are located on opposite sides of the *L^4* gene (Tomita et al. 2008, Yang et al. 2009).

De Ruiter progeny groups consisted of 29 progeny groups derived from 29 lines that were all resistant against *Tobamovirus* $P_{1.2.3}$. When these plants were analyzed with 189D23M and 087H03T7 markers, thirteen and fourteen progeny groups showed identical melting curve patterns to those of individuals in MS and SP F_2 populations, respectively (Table 4). However, other progeny groups showed different melting curve patterns: fourteen and twelve additional HRM melting curve patterns were observed for 189D23M and 087H03T7HRM markers, respectively. Each progeny group showed two or three distinct melting curve patterns. The number of melting curve patterns for 189D23M and 087H03T7HRM indicate that the sequences of these markers in different breeding lines are very diverse. By contrast, the L4segF&R marker showed only three melting curve patterns in all of tested plant materials: homozygous resistance, homozygous susceptible, and heterozygous. This demonstrates that L4segF&R sequences are very well conserved in the breeding lines. Nevertheless,

two recombinants were identified using the L4segF&R marker (Table 2, Table 3 type B).

Analysis of Monsanto F_2 breeding populations and Nongwoo breeding lines with the L4segF&R marker revealed that marker genotypes were perfectly co-segregated with resistance and susceptibility phenotypes (Table 5). However, 54 out of 90 individuals from Monsanto F_2 populations and 8 out of 22 lines from Nongwoo were found to be recombinants after progeny tests (Table 1, Table 6). While the differences in recombination rates between populations were observed, the recombination rates in Monsanto and Nongwoo populations were extremely high. These results suggest that genetic structures around the L^4 gene of Monsanto F_2 breeding populations and Nongwoo breeding lines are different from those of MS and SP cultivars and De Ruiter progeny groups. To confirm if genetic structures of the breeding lines are indeed different, L^4 flanking markers 087H03T7HRM and 189D23M were tested. L4segF&R perfectly co-segregated with 189D23M, but not with 087H03T7HRM in 24 individuals from Monsanto F_2 breeding populations and Nongwoo breeding lines. After progeny tests, however, the number of recombinants was increased to 62 for L4segF&R and 189D23M markers, and 61 for 087H03T7HRM (Table 2). These results indicate that genetic structures containing these three markers are quite different. Recombinants could be classified into two different groups according to the

L4segF&R genotype and the resistance phenotype in the progeny populations: in one group the resistance phenotype was segregated and the marker genotype was fixed to resistant (Table 3 type D and E), whereas in the other group the resistance phenotype was fixed and marker genotype was segregated (Table 3 type F). Interestingly, no susceptible recombinant having homozygous or heterozygous resistance genotypes of this marker was found. This demonstrates that individuals containing the resistance haplotype of the L4segF&R region are all resistant, thus the L4segF&R region appeared to be related to the resistance in these breeding lines despite non-Mendelian inheritance patterns in progeny populations. Genotypes of all of recombinants were analyzed at least two times to confirm the results. Disease resistance response of *L* is the HR, which is clearly distinguishable from disease escape or susceptible phenotype.

Table 4. Marker analysis of De Ruiter progeny groups

Progeny groups	No. of individuals	No. of recombinants by L4segF&R	Melting curve patterns		
			L4segF&R	087H03T7HRM	189D23M
DR-1	10	1	Identical[z]	Identical	Identical
DR-2	9	0	Identical	Identical	Identical
DR-3	8	0	Identical	Identical	Identical
DR-4	7	0	Identical	Identical	Identical
DR-5	9	0	Identical	Identical	Identical
DR-6	5	0	Identical	Identical	Identical
DR-7	7	0	Identical	Identical	Identical
DR-8	10	0	Identical	Identical	Identical
DR-9	9	0	Identical	Identical	Identical
DR-10	8	0	Identical	Identical	Identical
DR-11	10	1	Identical	Identical	Identical
DR-12	8	0	Identical	Identical	Identical
DR-13	9	0	Identical	Identical	Identical
DR-14	9	0	Identical	GA[y]-1, GA-2	GB[x]-1, GB-2
DR-15	10	0	Identical	GA-3, GA-4	GB-3, GB-4
DR-16	9	0	Identical	GA-3, GA-4	GB-3, GB-4
DR-17	8	0	Identical	GA-3, GA-4	GB-3, GB-4
DR-18	10	0	Identical	GA-3, GA-4	GB-3, GB-4
DR-19	6	0	Identical	GA-3, GA-4	GB-3, GB-4
DR-20	8	0	Identical	GA-3, GA-5	GB-3, GB-5
DR-21	9	0	Identical	GA-3, GA-5	GB-3, GB-5
DR-22	8	0	Identical	GA-3, GA-5	GB-3, GB-5
DR-23	6	0	Identical	GA-6, GA-7	GB-6, GB-7
DR-24	8	0	Identical	GA-8, GA-9	GB-8, GB-9
DR-25	9	0	Identical	GA-8, GA-9	GB-8, GB-9
DR-26	10	0	Identical	GA-10, GA-11	GB-10, GB-11
DR-27	10	0	Identical	GA-10, GA-11	GB-10, GB-11
DR-28	8	0	Identical	GA-6, GA-12	GB-6, GB-12
DR-29	9	0	Identical	Identical	GB-13, GB-14

Table 5. Summary of progeny tests of Monsanto F_2 breeding populations and Nongwoo breeding lines

Plant Materials	Marker genotypes			Progeny test[z]	No. of individual
	087H02T7HRM	189D23M	L4segF&R		
Monsanto population A[y]	R	R	R	R	6
	H	R	R	R	1
	R	R	R	H	7
	H	R	R	H	12
	R	S	S	S	3
Monsanto population B[y]	R	R	R	R	8
	H	H	H	H	18
	R	R	R	H	27
	H	H	H	R	8
Nongwoo breeding lines[x]	R	R	R	R	5
	R	R	R	R	1
	S	R	R	R	8
	H	H	H	R	8

[z]R: resistance fixed; H: phenotype segregated

[y]Monsanto breeding populations were progenies derived from two different populations

[x]Nongwoo breeding lines were independent parental lines

Table 6. Progeny tests of each Monsanto F_2 breeding populations

Individuals	Phenotypes (F_2)	Genotypes			Progeny tests ($F_{2:3}$)	
		087H02T7	189D23M	L4segF&R	R	S
A^z-1	R	H	R	R	10	4
A-2	R	H	R	R	5	3
A-3	R	H	R	R	9	4
A-4	R	R	R	R	12	2
A-5	R	H	R	R	6	4
A-6	R	H	R	R	5	3
A-7	R	H	R	R	7	2
A-8	R	H	R	R	11	2
A-9	R	H	R	R	10	0
A-10	R	R	R	R	7	1
A-11	S	R	S	S	0	15
A-12	R	R	R	R	14	1
A-13	R	H	R	R	14	2
A-14	S	R	S	S	0	13
A-15	R	H	R	R	10	3
A-16	R	R	R	R	15	1
A-17	R	R	R	R	12	0
A-18	R	R	R	R	16	0
A-19	R	R	R	R	15	1
A-20	S	R	S	S	0	13
A-21	R	R	R	R	11	2
A-22	R	H	R	R	12	4
A-23	R	R	R	R	9	0
A-24	R	H	R	R	7	2
A-25	R	R	R	R	3	6
A-26	R	R	R	R	16	0
A-27	R	R	R	R	15	0
A-28	R	H	R	R	8	7
A-29	R	R	R	R	13	0
B^z-1	R	H	H	H	10	0
B-2	R	H	H	H	12	2
B-3	R	R	R	R	12	0
B-4	R	R	R	R	13	1
B-5	R	R	R	R	14	1
B-6	R	H	H	H	11	2
B-7	R	H	H	H	14	2
B-8	R	R	R	R	8	4
B-9	R	R	R	R	12	0
B-10	R	R	R	R	14	0

Individuals	Phenotypes (F_2)	Genotypes			Progeny tests ($F_{2:3}$)	
		087H02T7	189D23M	L4segF&R	R	S
B-11	R	R	R	R	11	1
B-12	R	H	H	H	14	0
B-13	R	H	H	H	12	1
B-14	R	R	R	R	9	4
B-15	R	H	H	H	13	3
B-16	R	R	R	R	12	3
B-17	R	H	H	H	13	2
B-18	R	R	R	R	13	1
B-19	R	R	R	R	9	4
B-20	R	H	H	H	15	0
B-21	R	R	R	R	13	2
B-22	R	R	R	R	8	6
B-23	R	R	R	R	13	1
B-24	R	H	H	H	14	1
B-25	R	R	R	R	14	0
B-26	R	R	R	R	13	3
B-27	R	H	H	H	11	1
B-28	R	H	H	H	10	1
B-29	R	R	R	R	12	3
B-30	R	H	H	H	12	3
B-31	R	H	H	H	13	0
B-32	R	R	R	R	13	0
B-33	R	H	H	H	11	1
B-34	R	R	R	R	10	3
B-35	R	H	H	H	15	0
B-36	R	H	H	H	13	1
B-37	R	R	R	R	15	1
B-38	R	H	H	H	12	1
B-39	R	H	H	H	12	1
B-40	R	H	H	H	15	0
B-41	R	H	H	H	13	1
B-42	R	R	R	R	11	2
B-43	R	R	R	R	14	2
B-44	R	H	H	H	11	0
B-45	R	H	H	H	14	1
B-46	R	R	R	R	8	1
B-47	R	R	R	R	15	1
B-48	R	H	H	H	10	2
B-49	R	H	H	H	15	1
B-50	R	R	R	R	9	1
B-51	R	R	R	R	15	1

Individuals	Phenotypes (F_2)	Genotypes			Progeny tests ($F_{2:3}$)	
		087H02T7	189D23M	L4segF&R	R	S
B-52	R	R	R	R	15	1
B-53	R	H	H	H	13	0
B-54	R	R	R	R	11	2
B-55	R	R	R	R	14	0
B-56	R	R	R	R	10	1
B-57	R	R	R	R	10	2
B-58	R	R	R	R	12	1
B-59	R	R	R	R	15	1
B-60	R	R	R	R	10	0
B-61	R	R	R	R	10	0

Genetic structures around the *L^4* locus in different breeding lines

Genetic structures around the *L^4* locus of different breeding lines were deduced based on genotypes of molecular markers linked to *L^4* (Table 3; Fig. 2). The L4segF&R marker derived from the *L^4* candidate sequence is expected to truly represent the *L^4* gene provided that the *L^4* candidate is the only gene conferring *Tobamovirus* resistance. Hence, homozygous or heterozygous genotypes for the L4segF&R marker are expected to give resistance, while plants carrying homozygous susceptible genotype are susceptible. In contrast to the expectation, a total of 7 recombinants were found in MS, SP F_2 populations and De Ruiter progeny groups. Two individuals of De Ruiter progeny groups and one individual of the MS F_2 population carrying the homozygous susceptible genotype for L4segF&R showed the resistance phenotype (Table 3 type A and B; Fig. 2 Model A and B), and four individuals of MS and SP F_2 populations carrying the heterozygous genotype for L4segF&R showed the susceptible phenotype (Table 3 type C; Fig. 2 Model C). These demonstrate that the *L^4* candidate is a functional homolog tightly linked to *L^4*. In addition, progenies of 49 individuals having the homozygous genotype for L4segF&R showed phenotype segregation, while progenies of 19 individuals with heterozygous genotypes for three linked markers were all resistance. Two inferences were drawn from these observations: first, the genetic background around the *L^4* candidate, probably its functional

homolog(s), might contribute *Tobamovirus* resistance; second, duplication and/or translocation might have occurred at or around the L^4 region as shown in Model D and E (Table 3 type D and E; Fig. 2). Finally, as depicted in Model F, it is possible that the genetic segment including the L^4 gene was duplicated or translocated to an unpaired chromosome region. If the L^0 allele remained and co-existed with the translocated L^4 allele in same chromosome, then heterozygous genotypes could be detected even when all progenies show homozygous resistance phenotypes (Table 3 type F; Fig. 2).

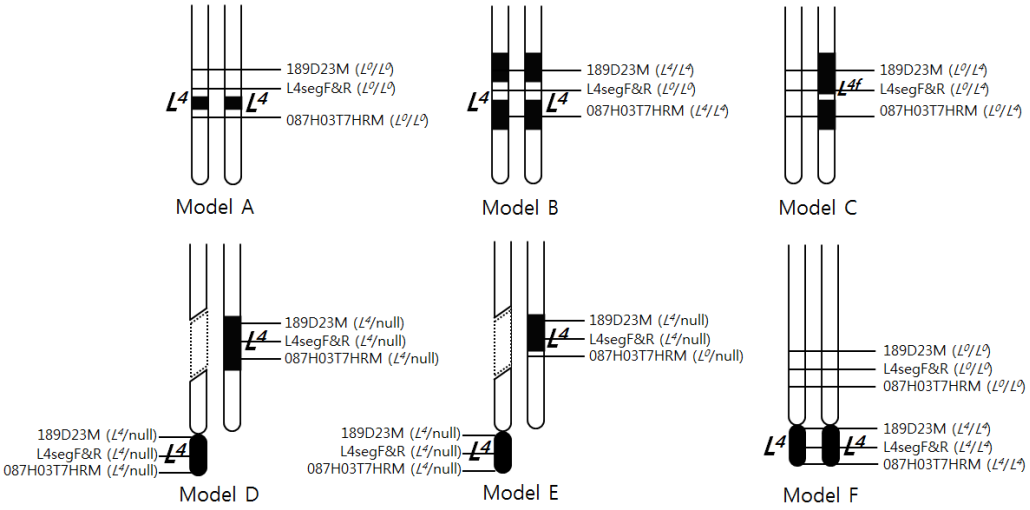


Figure 2. Putative genetic structures of the region around L^4 in different breeding lines deduced from marker genotypes (Table 3). The L^4-containing segments from *C. chacoence* (black) appear to be introgressed differently into the *C. annuum* genome. Models A, B and C are depicted under the assumption that the L^4 candidate is a functional homolog (L^{4f}) tightly linked to L^4. Models D, E and F are depicted under the assumption of translocation of L^4. There are no allelic regions corresponding to the L^4 segment in models D and E due to deletion of the allelic region (dotted lines boxes), whereas the allelic region remains in model F.

Development of *L* allele specific HRM markers

Although several molecular markers linked to L^3 or L^4 have been identified, no markers differentiating all *L* alleles have been developed. In this study, I attempted to develop *L* allele-specific markers based on full length cDNA sequences of one L^3 allele and six *L* gene candidates, L^1, L^{1a}, L^{1c}, L^2, L^{2b}, and L^4 (Tomita et al., 2011). Since the nucleotide diversity of the region encoding the LRR domains is generally higher as compared to those encoding the NBS domains, seven primers, designated L4RP-1F, L4RP-1R, L4RP-2F, L4RP-2R, L4RP-3F, L4RP-3R, and 3'end, were designed based on the LRR sequence of the L^4 candidate (Fig. 1). A total of six combinations of the primers, L4RP-1F/L4RP-1R, L4RP-2F/L4RP-2R, L4RP-3F/L4RP-3R, L4RP-1F/L4RP-2R, L4RP-2F/L4RP-3R, and L4RP-3F/3'end, were tested using indicator plants containing homozygous *L* alleles. L4RP-1F/L4RP-1R, L4RP-2F/L4RP-2R, L4RP-1F/L4RP-2R, and L4RP-2F/L4RP-3R primer sets amplified single band with expected fragment sizes. The L4RP-3F/L4RP-3R primer set amplified more than five different bands ranging from 160 to 400bp. The L4RP-3F/3'end primer set amplified only a single weak band for the homozygous L^0 genotype but showed a clear single band in all other genotypes (Fig. 3).

HRM analysis was carried out for all of the primer sets with the indicator plants. Since the melting curve patterns of the amplicons from L4RP-3F/3'end

were different between two independent sources of L^1 derived from MS and SP, we then analyzed the amplicon sequences and found that the L^1 sequences from MS and SP were indeed different, corresponding to L^{1a} (AB523373) and L^1 (AB523372), respectively. Therefore, our indicator plants represent six *L* alleles. After screening the six indicator plants with all primer combinations, two primer sets, L4RP-3F/L4RP-3R and L4RP-3F/3'end, which could discriminate all six alleles, were selected (Fig. 4). For the L4RP-3F/3'end primer set, the melting curve pattern of L^0 was slightly inconsistent due to poor amplification. Although multiple bands were obtained using L4RP-3F/L4RP-3R, the resulting HRM melting curves were very consistent and gave reproducible genotyping results.

In order to test whether the *L* allele specific markers could also differentiate heterozygous genotypes, all possible heterozygous genotypes were artificially produced by mixing DNA of different alleles. A total of 15 artificial heterozygous genotypes were generated by pairing six different *L* alleles and subjected to HRM analysis together with the corresponding homozygous parental genotypes using L4RP-3F/L4RP-3R and L4RP-3F/3'end primer sets respectively. The L4RP-3F/L4RP-3R primer set distinguished between all heterozygous types, while the L4RP-3F/3'end primer set could discriminate all of them except for the artificially heterozygous DNA mixed with L^1 from MS (Fig. 5, Fig. 6).

I further performed a blind test using L4RP-3F/L4RP-3R and L4RP-

3F/3’end primer sets using breeding materials from Enza Zaden. The breeding lines were composed of 91 individuals homozygous for each L^0, L^1, L^3, and L^4 allele. The two markers successfully detected these four different *L* alleles; only one individual of L^1 was assigned the wrong allele with these primer sets (Table 7).

Table 7. Genotype analysis of L4RP-3F/ L4RP-3R and L4RP-3F/ 3'end marker sets for Enza Zaden breeding lines

Phenotype	Recombinants / No of lines
L^0/L^0	0 / 35
L^1/L^1	1 / 13
L^3/L^3	0 / 22
L^4/L^4	0 / 21
Total	1 / 91

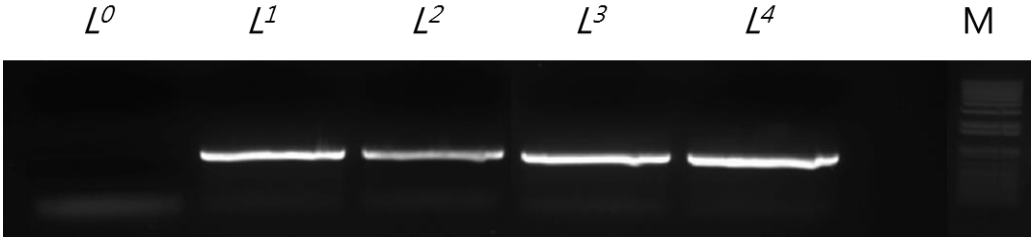


Figure 3. PCR amplification of L4RP-3R&3'end primer set in accessions containing each *L* alleles. Faint single band was amplified in L^0 genotype accessions while clear single band was amplified in other four resistant allele genotypes.

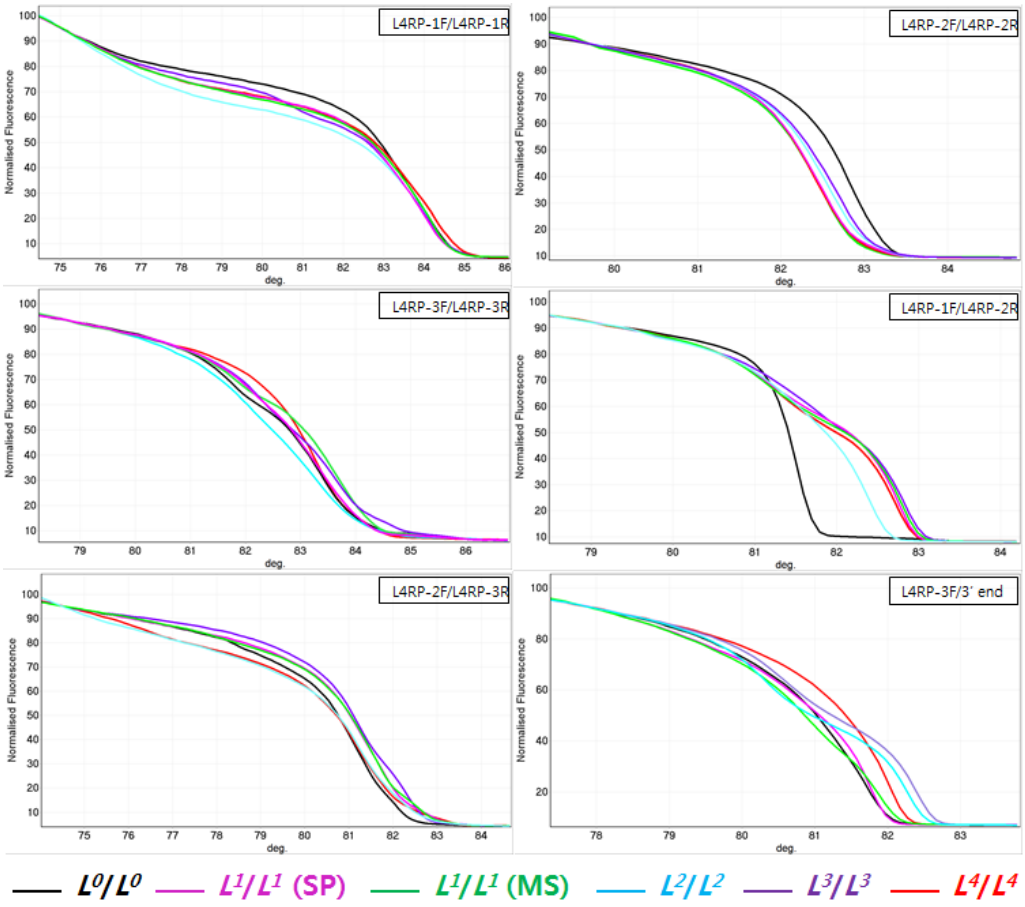
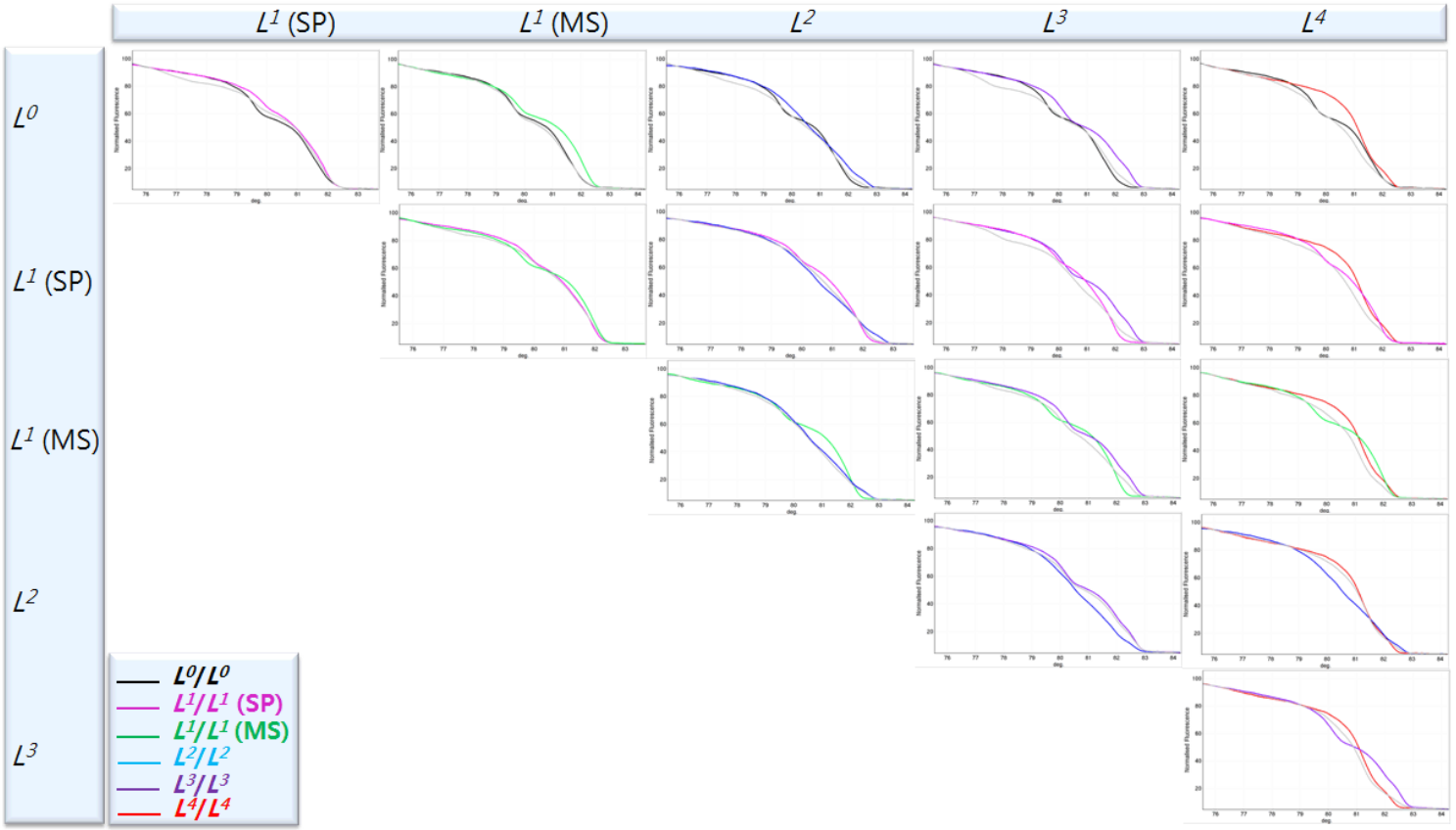


Figure 4. SNP marker profiles developed based on the 3' LRR-encoding domain of the L^4 candidate. Six different primer sets were analyzed using indicator plant gDNA homozygous for each *L* allele. Two primer sets L4RP-3F/ L4RP-3R and L4RP-3F/ 3' end were able to distinguish all of *L* alleles clearly.


L1 (SP)
L1 (MS)
L2
L3
L4
L0
L1 (SP)
L1 (MS)
L2
L3
L0/L0
L1/L1 (SP)
L1/L1 (MS)
L2/L2
L3/L3
L4/L4

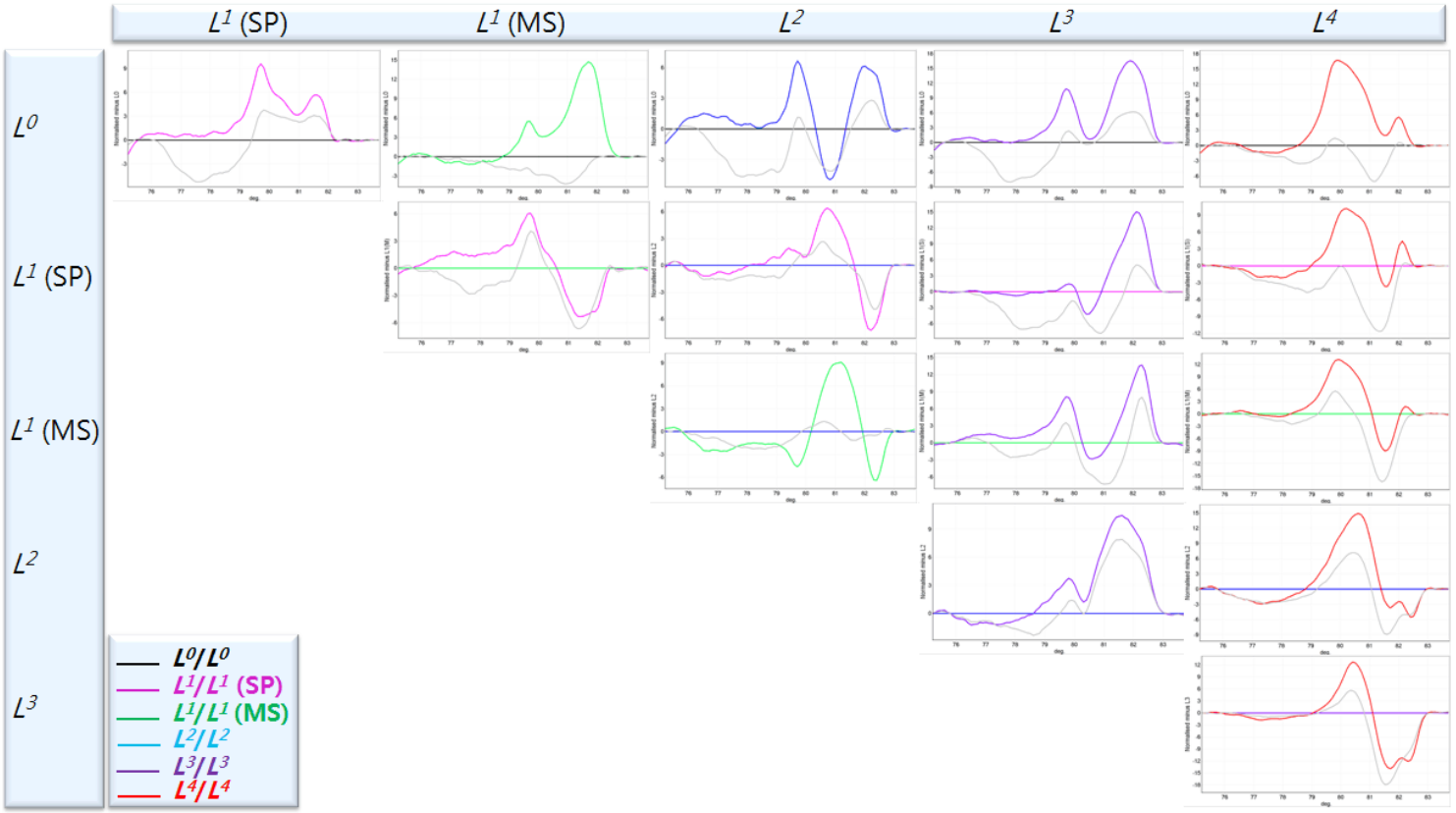
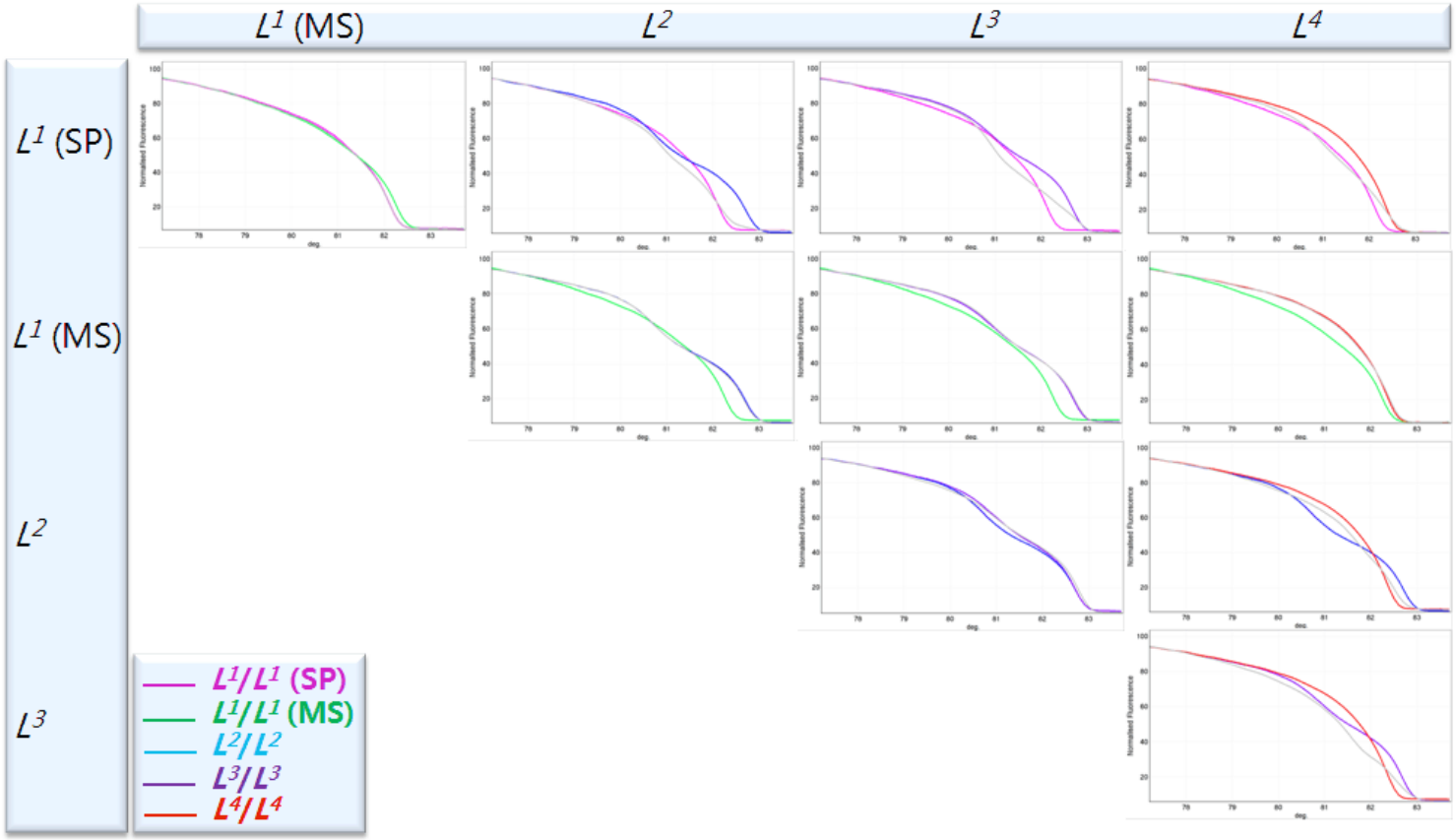

L^1 (SP)
L^1 (MS)
L^2
L^3
L^4
L^0
L^1 (SP)
L^1 (MS)
L^2
L^3
L^0/L^0
L^1/L^1 (SP)
L^1/L^1 (MS)
L^2/L^2
L^3/L^3
L^4/L^4

Figure 5. Melting curve patterns and difference graphs of homozygous genotypes of six *L* alleles, L^0, L^1 from SP, L^1 from MS, L^2, L^3, and L^4, and heterozygous genotypes of each pair of these alleles in HRM analysis of the L4RP-3F/ L4RP-3R primer set. This marker set could distinguish 15 types of heterozygous genotypes from homozygous genotypes. L^0, L^1 from SP, L^1 from MS, L^2, L^3, and L^4 alleles are depicted in black, purple, green, blue, violet, and red colors, respectively. Melting curves and difference graphs of heterozygous genotypes are in grey.


L1 (MS)
L2
L3
L4
L1 (SP)
L1 (MS)
L2
L3
L1/L1 (SP)
L1/L1 (MS)
L2/L2
L3/L3
L4/L4

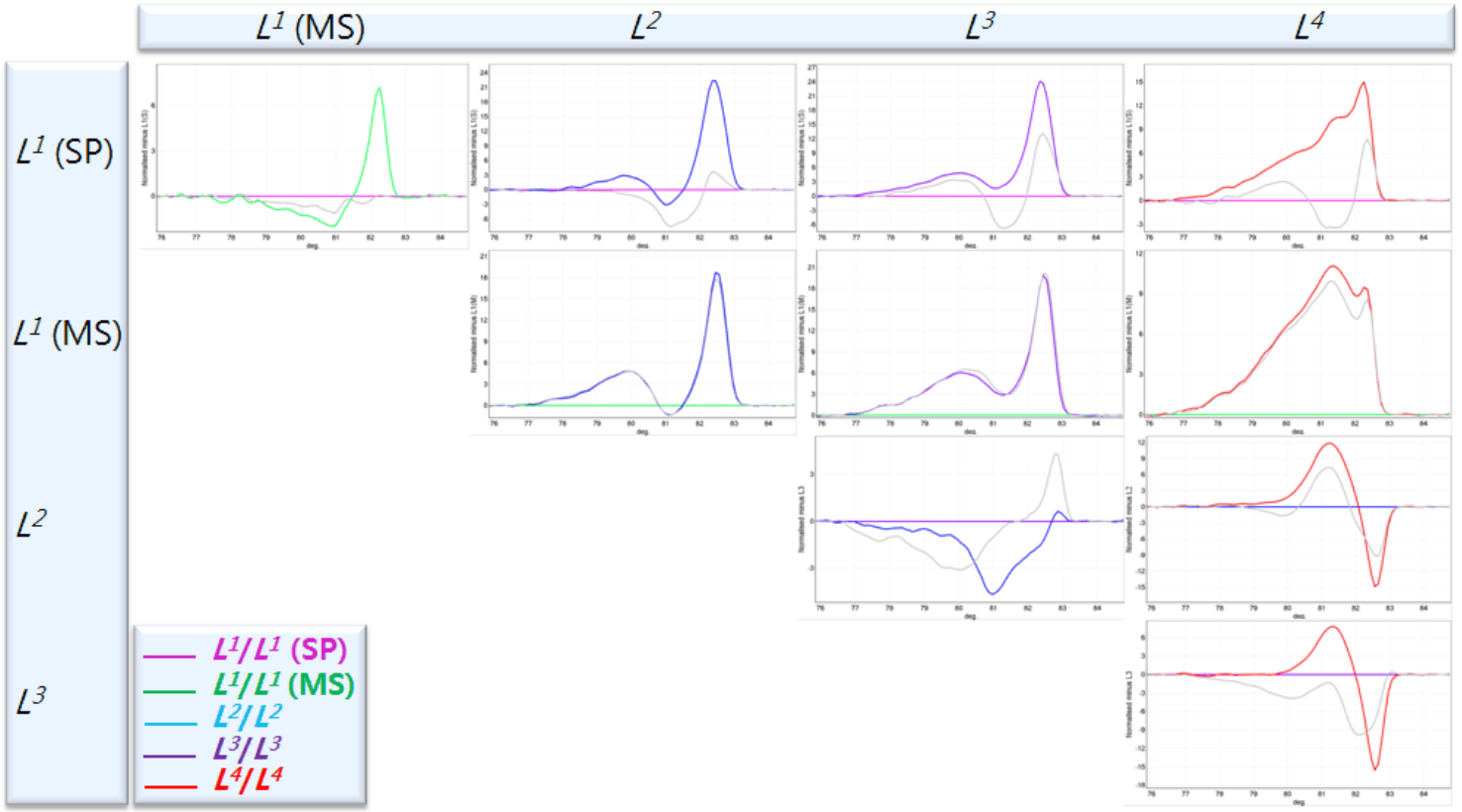

L1 (MS)
L2
L3
L4
L1 (SP)
L1 (MS)
L2
L3
L1/L1 (SP)
L1/L1 (MS)
L2/L2
L3/L3
L4/L4

Figure 6. Melting curve patterns and difference graphs of homozygous genotypes of five *L* alleles, L^1 from SP, L^1 from MS, L^2, L^3, and L^4, and heterozygous genotypes of each pair of these alleles in HRM analysis of the L4RP-3F/ 3'end primer set. This marker set could not distinguish 4 heterozygous genotypes having L^1 from MS. L^1 from SP, L^1 from MS, L^2, L^3, and L^4 alleles are shown in purple, green, blue, violet, and red, respectively. Melting curves and difference graphs of heterozygous genotypes are in grey.

DISCUSSION

Our linkage analysis of F_2 populations and breeding lines indicates that the putative L^4 candidate might not be the L^4 gene or that another genetic factor is required for full *Tobamovirus* resistance. In contrast to our expectation that the L4segF&R marker designed specifically for the L^4 candidate (Tomita et al. 2011) would perfectly co-segregate with the resistance phenotype, recombinants were found.

The existence of a functional homolog other than the L^4 candidate is plausible in all six genetic models postulated in this study. Clustering of resistance genes are commonly observed in plant genome. For instance, *R3a* and *R3b* genes found in syntenic regions of the *L* locus are functional genes clustered together (Huang et al. 2004; Huang et al. 2005). Models D, E and F are based on the hypothesis that duplication/translocation occurred in the genomes of Monsanto F_2 breeding populations and Nongwoo breeding lines (Fig. 2). Duplication/translocation could be supported by several haplotype models (Fig. 7). The L^4 gene was first introgressed from *C. chacoense* to *C. annuum* breeding lines containing the L^0 gene (model ①). However, the introgressed *C. chacoense* segment might have been unstable at the position and moved to a non-allelic region by unequal crossing over (model ②) or by translocation (model ③).

Models A, B and C are related to haplotype model ① and are explained by the existence of other functional homolog. Models D and E are related to haplotype model ① and haplotype model ②-I or II, and Model F is related to haplotype model ③. The NBS-LRR disease resistance gene family is known to form clusters in the genome together with RGAs (Hayashi and Yoshida 2009; Qu et al. 2006). Many paralogs have been reported in the genetic region around the L^3 gene and around the 087H03T7HRM marker, which is located 1cM from the L^4 gene (Tomita et al. 2008; Tomita et al. 2011; Yang et al. 2009). As was observed at the *R3a* region in potato, the region around the L^4 gene locus is expected also to contain high copies of retrotransposable elements and R gene paralogs (Michelmore et al. 1998). Frequent unequal recombination at the resistance gene clusters was also demonstrated in the *Pc* gene in *Sorghum* where the unequal recombination rate was reported to be 10^{-3} to 10^{-4} (Nagy and Bennetzen 2008). As shown in model ②-I and ②-II, I hypothesize that unequal crossover might have occurred in various regions containing repetitive sequences during the introgression of L^4. Nevertheless, haplotype model ② type introgression might have been stabilized and fixed in some breeding lines.

Interspecific karyotype variation has been commonly observed in *Capsicum* spp., particularly within the wild species. Previous karyotype analysis of *C. chacoense* and *C. annuum* showed differences in total chromosome length

between the two species (Moscone 1990; Moscone et al. 2007). Haplotype model ③ is based on the assumption that the shape and the length of Chromosome 11 of *C. chacoense* and *C. annuum* are different, and that introgression of chromosomal segment from *C. chacoense* to *C. annuum* may have resulted in structural changes to Chromosome 11. Since the *L* locus was located at the end of the long arm of the chromosome, translocation might have occurred due to difference in length of these chromosomes. The phenomenon can be further explained by chromosome breakage. Differences in length cause unpaired chromosomes to form a bubble-like loop when paring at the meiosis stage (Bosco et al. 1998). This structure may cause breakage of the chromosome, with hanging chromosome fragment then moving and attaching to the nearest end of the long arm of the chromosome (Johnson and Jasin 2001). Alternatively, model ③ can be explained by homologous recombination similar to that observed in *Solanum lycopersicoides* introgression of cultivated tomato (Canady et al. 2006). If homologous recombination occurred between chromosomes with length differences, the length of the chromosome might be changed to longer or shorter after recombination. If the longer chromosome contributed by *C. chacoense* paired with an ordinary chromosome of *C. annuum*, the end of the longer chromosome could not be paired with other chromosome. As a result, increased

homologous recombination can be observed in the genome where introgression has occurred. This could explain our observation and provide evidence for the hypothesis that the genetic region around the L4segF&R marker can be duplicated or translocated to another genetic region in the genome. If recombination between current and new genetic regions occurred during meiosis, segregation can be observed in the progeny. These models were developed based on our observations of recombinants, therefore our hypothesis of the chromosomal rearrangement at the *L* region should be confirmed by sequencing or haplotype analysis using markers saturating the *L* region.

SNPs are widely distributed in plant genomes, providing the most frequent nucleotide variation. Many techniques for SNP detection and SNP marker development have been reported (Vignal et al. 2002; Kim and Misra 2007). However, these SNP detection techniques often have limitations: restriction enzyme-based technology can detect only one specific SNP site at a time, whereas gel-based and MALDI-TOF approaches are labor-intensive and expensive. The HRM technique consisting of one round of PCR and automated post-PCR high resolution melting curve analysis provides an inexpensive and user friendly platform for SNP detection and genotyping. Furthermore, this technique can detect several SNPs in a PCR amplicon, and information about the DNA sequence of the target region and SNP position is not required. However,

HRM analysis is limited to PCR amplicon sizes shorter than 500 bp (Liew et al., 2004, Park et al., 2009).

In this study, I used HRM techniques to develop allele-specific markers for the L^4 locus. Our markers were designed from the LRR-encoding domain of the NBS-LRR disease resistance gene candidate, where the domain is characterized by high nucleotide variation and poor sequence conservation, particularly at the 3' region. I exploited the differential single nucleotide polymorphism between the four known *L* alleles and created SNP markers able to distinguish each allele. Our study demonstrates that HRM analysis is a suitable technique for allele-specific marker development; furthermore, analysis of melting curve patterns offers the potential to detect new alleles that can be isolated and confirmed by testing with different TMV pathotypes. The co-dominant nature of our *L* allele-specific marker, which can detect all homozygous and heterozygous genotypes of the *L* alleles, makes it an ideal marker for applied molecular breeding to discriminate between different alleles in breeding lines and populations. Since disease resistance spectrums of different *L* alleles overlap, it is very difficult to detect two different alleles in one individual solely by phenotypic screening and inoculation with different *Tobamovirus* pathotypes. Being the closest molecular marker identified to date, the *L* allele-specific marker developed in the study provides a solution for pepper breeders developing

improved resistance lines against *Tobamovirus* by pyramiding *L* alleles through marker assisted selection (MAS). Given the complexity of the *L* region and the presence of different alleles, markers developed in this study should be tested with each breeder's material.

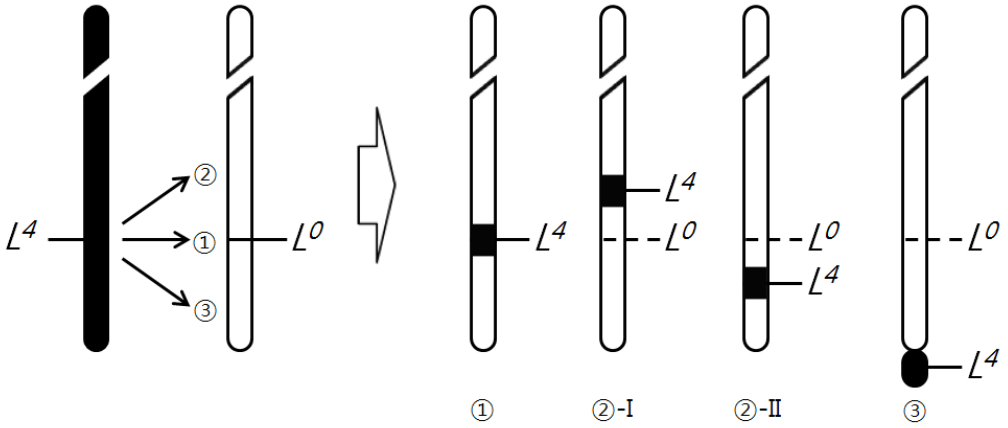


Figure. 7. Haplotypes models representing introgression of the L^4 gene from *C. chacoense* into *C. annuum*. The genomic segment including the L^4 gene was introgressed into L^0 by allelic recombination (model ①). If non-allelic homologous recombination by unequal crossing over occurred, the genomic segment including the L^4 gene could be introgressed near L^0 (model ②-I and ②-II). If the genomic segment including the L^4 gene was unstable in the *C. annuum* genetic background, it could translocate to an unpaired chromosome region (model ③). The genomic segment including the L^0 gene could remained or be deleted (dotted line) in models ② and ③.

REFERENCES

Berzal-Herranz A, de la Cruz A, Tenllado F, Diaz-Ruiz JR, Lopez L, Sanz AI, Vaquero C, Serra MT, arcia-Luque I (1995) The *Capsicum* L^3 gene-mediated resistance against the Tobamoviruses is elicited by the coat protein. Virology 209:498–505

Bosco G, Haber JE (1998) Chromosome break-Induced DNA replication leads to nonreciprocal translocations and telomere capture. Genetics 150:1037-1047

Boukema IW (1980) Allelism of genes controlling resistance to TMV in *Capsicum* L. Euphytica 29:433–439

Canady MA, Ji Y, Chetelat RT (2006) Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. Genetics 174:1775-1788

de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T (2005) Carh ta gene: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics 21:1703–1704

de la Cruz A, Lopez L, Tenllado F, Diaz-Ruiz JR, Sanz AI, Vaquero C, Serra MT, Garcia-Luque I (1997) The coat protein is required for the elicitation of the Capsicum L^2 gene-mediated resistance against the Tobamoviruses. Molecular Plant-Microbe Interactions 10:107–113

Genda Y, Kanda A, Hamada H, Sato K, Ohnishi J, Tsuda S (2007) Two amino acid substitutions in the coat protein of *Pepper mild mottle virus* are responsible

for overcoming the L^4 gene-mediated resistance in *Capsicum* spp. Phytopathology 97:787-793

Hayashi K and Yoshida H (2009) Refunctionalization of the ancient rice blast disease resistance gene *Pit* by the recruitment of a retrotransposon as a promoter. The Plant Journal 57:413-425

Huang S, Vleeshouwers VGAA, Werij JS, Hutten RCB, van Eck HJ, Visser RGF, Jacobsen E (2004) The *R3* resistance to *Phytophthora infestans* in potato is conferred by two closely linked R Genes with distinct specificities. Molecular Plant-Microbe Interactions 17:428-435

Huang S, van der Vossen EAG, Kuang H, vleeshouwers VGAA, Zhang N, Borm TJA, van Eck HJ, Baker B, Jacobsen E, Visser RGF (2005) Comparative genomics enabled the isolation of the *R3a* late blight resistance gene in potato. The Plant Journal 42:251-261

Johnson RD and Jasin M (2001) Double-strand-break-induced homologous recombination in mammalian cells. Biochemical Society Transactions 29:196-201

Kang WH, Hoang NH, Yang HB, Kwon JK, Jo SH, Seo JK, Kim KH, Choi D, Kang BC (2010) Molecular mapping and characterization of a single dominant gene controlling CMV resistance in peppers (*Capsicum annuum* L.). Theoretical and Applied Genetics 120:1587-1596

Kim HJ, Han JH, Yoo JH, Cho HJ, Kim BD (2008) Development of a sequence characteristic amplified region marker linked to the L^4 locus conferring broad

spectrum resistance to Tobamoviruses in pepper plants. Molecules and Cells 25: 205–210

Kim S and Misra A (2007) SNP Genotyping: Technologies and biomedical applications. Annual Review of Biomedical Engineering 9:289-320

Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E, Wittwer C (2004) Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clinical Chemistry 50:1156-1164

Madsen LH, Collins NC, Rakwalska M, Backes G, Sandal N, Krusell L, Jensen J, Waterman EH, Jahoor A, Ayliffe M, Pryor AJ, Langridge P, Schulze-Lefert P, Stougaard J (2003) Barley disease resistance gene analogs of the NBS-LRR class: identification and mapping. Molecular Genetics & Genomics 269:150-161

Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annual Review of Plant Biology 54:23-61

Matsunaga H, Saito T, Hirai M, Nunome T, Yoshida T (2003) DNA markers linked *to Pepper mild mottle virus* (PMMoV) resistant locus (L^4) in *Capsicum*. Japanese Society for Horticultural Science 72:218–220

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. The Plant Journal 20:317–332

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

Moscone EA (1990) Chromosome studies on *Capsicum* (Solanaceae) I. Karyotype analysis in *C. Chacoense.* Brittonia 42:147-154

Moscone EA, Scaldaferro MA, Grabiele M, Cecchini NM, Sanchez Garcia Y, Jarret R, Davina JR, Ducasse DA, Barboza GE, Ehrendorfer F (2007) The evolution of chili peppers (*Capsicum* – Solanaceae): a cytogenetic perspective. Acta Horticulturae (ISHS) 745 137-169

Nagy ED and Bennetzen JL (2008) Pathogen corruption and site-directed recombination at a plant disease resistance gene cluster. Genome Research 18:1918-1923

Pankovic D, Radovanovic N, Jocic S, Satovic Z, Skoric D (2007) Development of co-dominant amplified polymorphic sequence markers for resistance of sunflower to downy mildew race 730. Plant Breeding 126:440-444

Park SW, An SJ, Yang HB, Kwon JK, Kang BC (2009) Optimization of high resolution melting analysis and discovery of single nucleotide polymorphism in *Capsicum*. Horticulture, Environment, and Biotechnology 50:31-39

Qu S, Liu G, Zhou B, Bellizzi M, Zeng L, Dai L, Han B, Wang G (2006) The broad-spectrum blast resistance gene *Pi9* encodes a nucleotide-binding site–leucine-rich repeat protein and is a member of a multigene family in rice. Genetics 172:1901-1914

Sugita T, Yamaguchi K, Sugimura Y, Nagata R, Yuji K, Kinishita T, Todoroki A (2004) Development of SCAR markers linked to L^3 gene in *Capsicum*. Breeding Science 54:111–115

Sutherland BG, Tobutt KR, Marchese A, Paternoster G, Simpson DW, Sargent DJ (2008) A genetic linkage map of *Physocarpus*, a member of the Spiraeoideae (Rosaceae), based on RAPD, AFLP, RGA, SSR and gene specific markers. Plant Breeding 127:527-532

Tomita R, Murai J, Miura Y, Ishihara H, Liu S, Kubotera Y, Honda A, Hatta R, Kuroda T, Hamada H, Sakamoto M, Munemura I, Nunomura O, Ishikawa K, Genda Y, Kawasaki S, Suzuki K, Meksem K, Kobauashi K (2008) Fine mapping and DNA fiber FISH analysis locates the Tobamovirus resistance gene L^3 of *Capsicum chinense* in a 400-kb region of R-like genes cluster embedded in highly repetitive sequences. Theoretical and Applied Genetics 117:1107–1118.

Tomita R, Sekine KT, Mizumoto H, Sakamoto M, Murai J, Kiba A, Hikichi Y, Suzuki K, Kobayashi K (2011) Genetic basis for the hierarchical interaction between Tobamoviruses and *L* resistance gene alleles from different pepper species. Molecular Plant-Microbe Interactions 24:108-117

van der Linden CG, Wouters DCAE, Mihalka V, Kochieva EZ, Smulders MJM, Vosman B (2004) Efficient targeting of plant disease resistance loci using NBS profiling. Theoretical and Applied Genetics 109:384-393

van Ooijen G, van den BHA, Cornelissen BJC, Takken FLW (2007) Structure and function of resistance proteins in Solanaceous plant. Annual Review of Phytopathology 45:43-72

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution 34:275-305

Yang HB, Liu WY, Kang WH, Jahn M, Kang BC (2009) Development of SNP markers linked to the *L* locus in *Capsicum* spp. by a comparative genetic analysis. Molecular Breeding 24:433-446

Yu YG, Buss GR, Marrof MAS (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proceedings of the National Academy of Sciences of the United States of America 93:11751-11756

CHAPTER III

De novo assembly and digital expression profiling of the *L* homologues in *Capsicum*

ABSTRACT

The L^4 candidate was isolated by a homology-based PCR. Genetic analysis using markers based on the L^3 gene indicated that the previously reported L^4 candidate gene is a functional homologue of L^4. *De novo* assembly of RNA transcriptome expressed in *C. chacoense* 'PI260429' was performed and digital expression analysis between four resistant accessions and four susceptible accessions of *Capsicum* spp. were carried out in order to isolate contigs encoding *L* gene homologues. I optimized the parameters of *de novo* assembly of the *L*-homologous contigs. Single assembly using Velvet when k-mer is 59 showed the

most high specificity and accuracy. A total of 96 *L*-homologue contigs were isolated. These *L*-homologue contigs contained CC, NB-ARC, or LRR domains. A total of 47 contigs were mapped on five chromosomes, mainly in chromosome 11. Digital expression analysis showed expression patterns of L-homologue contigs. Eight contigs derived from four transcripts including the L^4 functional homologue were expressed only in four resistant accessions. This represents that three transcripts may be responsible for *Tobamovirus* resistance with the L^4 functional homologue. In this study, I suggests *de novo* assembly and validation strategy for highly repetitive sequences that are belonged to a multi copy gene. Our result would be a useful experimental illustration for assembly multi copy gene having high homology from using next-generation sequencing data.

INTRODUCTION

The L^3 gene was isolated through map-based cloning and other *L* allele candidates of L^1, L^{1a}, L^{1c}, L^2, L^{2b}, and L^4 were isolated by homology based PCR strategy (Tomita et al. 2011). Co-segregation analysis of molecular marker designed based on the L^4 candidate gene using two F_2 L^4-segregating populations and breeding materials demonstrated that the L^4 candidate alone was not sufficient to explain *Tobamovirus* resistance. Therefore it was assumed that another genetic factor may be required (Yang et al. 2012).

Nucleotide binding site – leucine rich repeat (NBS-LRR) was one of the major classes of plant disease resistance genes and the L^3 allele belongs to this class as well. NBS-LRR gene was known to be located in the cluster with highly repetitive homologues (Meyers et al. 1999; Michelmore and Meyers 1998). Resistance gene analog (RGA) were isolated and reported through the map-based cloning and whole genome sequencing. The function of RGA is not clear so far, but there were few cases suggesting that RGA is required for disease resistance or closely linked two NBS-LRR genes conferring resistance to same pathogen but different isolates (Huang et al 2005; Okuyama et al 2011). RGAs were found in the genetic region near the *L* locus as well. A total of eight *L*-homologue were

found in 400-kb sequence around the L^3 locus and three *L* homologues were found in genetic region 224-kb sequence apart from about 1.5cM away the L^4 locus (Tomita et al. 2011; Yang et al 2009).

Next-generation sequencing (NGS) technologies have been advanced, and many various sequencing platforms have been developed such as Roche/454, Illumina/Solexa, Life/APG, Polonator/Dover and HeliScope/Helicos BioScience companies that are available commercially (Metzker 2010). These platforms vary in short read length and total sequencing capacity per run. Among them Illumina NGS platform occupied most of the NGS market recently because of length of read and data acquisition per run. Progress on NGS techniques was accompanied by the development of bioinformatics tools. Furthermore this progress steered further development of genome and transcriptome research such as whole genome sequencing, resequencing, variant calling, metagenomics, discovering noncoding RNA and so on (Mardis 2008; Paszkiewicz and Studholme 2010; Robert et al. 2010; Wang et al. 2009).

De novo assembly programs enable to produce transcript sequence pools including RGAs without a reference genome. It is a time saving and cost effective method compared to EST analysis. Two main algorithms, overlapped graph and de Bruijn graph, were used for programming of *de novo* assembler (Clarke et al. 2013). Complexity and computational time are exponentially increased in

overlapped graph based program. But after de Bruijn graph algorithm introduced and reduced computational time, *de novo* assembly programs using this algorithm are widely used (Birol et al. 2009; Martin and Wang 2012; Schulz et al. 2012; Zerbino et al. 2008). RNA-seq provides an efficient way to measure gene expression in a single analysis (Trapnell et al. 2012). High-throughput RNA-seq which cannot be achieved by reverse transcriptase PCR (RT-PCR) and real-time PCR, furthermore RNA-seq may reduce time and cost to make array-chips.

In this study, I designed validation indexes for the optimization of *de novo* assembly of *L* homologues through Velvet program using transcriptome data analyzed by Illumina platform. I found optimal condition for highly repetitive sequences such as NBS-LRR multigene family. Digital expression analysis of four resistant and four susceptible *Capsicum* accessions and sequence comparison with currently reported RGAs were performed and putative *Tobamovirus* resistance transcripts candidate were found. This study showed possibility of *de novo* assembly of highly repetitive RGAs and suggested efficient research strategy to isolate putative resistance gene candidate through RNA-seq bioinformatics tool.

MATERIALS AND METHODS

Plant materials

Capsicum accessions containing *Tobamovirus* resistance *L* alleles, L^1, L^2, L^3, and L^4, and susceptible allele L^0 were used to analyze digital expression of *L* homologue. The allele determination of eight accessions, *C. annuum* cv. 'Yuwolcho', *C. annuum* cv. 'Criollo de Morelos 334 (CM334)', *C. chinense* PI152225, *C. chacoense* PI260429, *C. annuum* cv. 'Jeju', *C. annuum* cv. 'Tean', *C. annuum* 'LAM32' and *C. annuum* 'YCM334 (an F_6 recombinant inbred line derived from the cross of Yolo wonder x CM334)', was performed by *in vivo* resistance screening and marker analysis. P_0, P_1, $P_{1.2}$ and $P_{1.2.3}$ pathotypes of *Tobamovirus* were used for *in vivo* screening, and L4segF&R, *L* allele specific marker, was used for marker analysis (Yang et al., 2012).

Total RNA extraction and transcriptome sequencing

Total RNA of seven *Capsicum* accessions, *C. annuum* cv. 'Yuwolcho', *C. chinense* PI152225, *C. chacoense* PI260429, *C. annuum* cv. 'Jeju', *C. annuum* cv. 'Tean', *C. annuum* 'LAM32' and *C. annuum* 'YCM334', was extracted from leaf tissues using Hybrid-R™ (Geneall biotechnology, Seoul, Korea) or TRIzol® RNA Isolation Reagents (Life technologies, USA) following the procedure

described by their manufacturer. Transcriptome sequences of *C. annuum* cv. 'Jeju', *C. annuum* cv. 'Tean' and *C. annuum* 'LAM32'were analyzed using 116-bp single-end sequencing strategy in National Instrumentation Center for Environmental Management (NICEM, Seoul, Korea). Transcriptome sequence of *C. annuum* cv. 'Yuwolcho' and *C. annuum* 'YCM334' were determined using 90-bp paired-end sequencing protocol in Beijing genome institute (BGI, Beijing, China). Transcriptome sequencing of *C. chinense* PI152225 and *C. chacoense* PI260429 were performed using 101-bp paired-end sequencing method in NICEM. Transcriptome sequence of CM334 was kindly provided by Dr. Choi (Seoul National University, Seoul, Korea) (Table 1).

Table 1. Transcriptome sequences of plant materials

Accession	Species	*L* allele	Read length	Sequencing
Jeju	*C. annuum*	L^0	116-bp (single-end)	NICEM
Tean	*C. annuum*	L^0	116-bp (single-end)	NICEM
LAM32	*C. annuum*	L^0	116-bp (single-end)	NICEM
YCM334	*C. annuum*	L^0	90-bp (paired-end)	BGI
Yuwolcho	*C. annuum*	L^1	90-bp (paired-end)	BGI
CM334	*C. annuum*	L^2	101-bp (paired-end)	NICEM
PI152225	*C. chinense*	L^3	101-bp (paired-end)	NICEM
PI260429	*C. chacoense*	L^4	101-bp (paired-end)	NICEM

***De novo* assembly of transcriptome**

Low quality sequences of transcriptome were trimmed using Najoshi sickle program (https://github.com/najoshi/sickle) with a quality cutoff of 25 (phred scale). Velvet and Oases programs were used for *de novo* assembly of transcriptome (Schulz et al. 2012; Zerbino DR and Birney E 2008). k-mer value of Velvet and Oases programs were varied to find optimal condition, however the other options were default values suggested by manual. Redundant *de novo* assembly products were removed using CD-HIT-EST algorithm and clustered by GICL program, advanced new version of TGICL program (Li and Godzik 2006; Pertea et al. 2003). BLAST search was performed to find the *L* homologues in *de novo* assembly products of PI260429. Contigs matched with L^4 functional homologue sequence (AB523377, GenBank) with cutoff < e^{-100} were used for *in silico* mapping and digital expression analysis.

Annotation and *in silico* mapping

Domain of each *de novo* assembly products were identified by sequence comparison with the L^3 gene (Tomita et al. 2011). Physical locations of *de novo* assembly products in *Capsicum* genome were analyzed by BLAST search using WGS dataset ver. 1.35 which was kindly provided by Dr. Choi http://cab.pepper.snu.ac.kr). *De novo* assembly products which showed over 97%

identity with *Capsicum* WGS were admitted to be located on matching region. Short read mapping was performed using BWA and Samtools program using default options (Li and Durbin 2009; Li et al. 2009).

Digital expression analysis

De novo assembly products were used as reference sequences. RNA-seq algorithm of CLC Genomics Workbench 6 program was used for digital expression as well with 97% of identity threshold (CLC bio, Prismet, Denmark). Digital expression data was normalized and transformed by CLC genomics workbench 6 internal algorithms.

RESULTS

Optimization of *de novo* assembly of the *L* homologues

NBS-LRR class resistance genes have been known to be consisted of members of the multigene subfamily. This indicates paralogs of target NBS-LRR gene with high homology exist in genome and it is required to find optimal k-mer, a constant value to determine whether two different reads are overlapped correctly, to avoid mis-assembly between highly repetitive sequences.

De novo assembly of transcriptome sequence of *C. chacoense* 'PI260429' was performed using various k-mer value (Table2), and coverage rate (recover rates of the L^4 functional homologue sequence) and maximum contig length were calculated in order to find an optimal k-mer (Fig 1). Coverage rate increased from k-mer = 35 and declined between k-mer = 47 and 51. It increased from k mer = 53 again and reached a peak (92.2%) at k-mer = 57. It decreased slightly between k-mer = 61 and 71, and fell back to a constant (32.9%) after k-mer = 83. Maximum contig length increased to a peak (1,453 bp) at k-mer = 57 and 59 then it significantly decreased from k-mer = 63 and stayed constant (427 bp) after k-mer = 83. *De novo* assembly products of three k-mers, 55, 57 and 59, showed the most high coverage rate and a maximum contig length. Therefore these three k-mer values were chosen for optimization of Oases assembly.

Merging and reassembly of the three *de novo* assembly products with k-mer values of 55, 57, and 59, was performed using Oasis program (Table 3 and Fig 2). Coverage rates and maximum contig lengths were calculated for validation. Both coverage rate and maximum contig length showed the highest value at k-mer of 55, 59, and 63. Maximum contig length of these Oasis products was the same as that of the Velvet product (k-mer = 59) whereas the coverage rate was about 2/3 of Velvet product (k-mer = 59). Validation results of *de novo* assembly by comparing coverage rate and maximum contig length indicated that Velvet assembly with a k-mer of 59 seems to be optimal for the *de novo* assembly of *L* homologues. Therefore, assembly product of Velvet (k-mer = 59) was used for *in silico* mapping and digital expression of the *L* homologues.

Table 2. The number of contigs, N50 and maximum contig length of *de novo* assembly of transcriptome obtained from *C. chacoense* 'PI260429' according to each k-mer

k-mer	No. of contigs	N50	Max. contig length (bp)
35	138,916	1,819	17,929
37	133,953	1,804	15,791
39	130,541	1787	15,791
41	126,744	1,753	18,972
43	124,114	1,722	16,504
45	120,844	1,687	15,791
47	118,564	1,646	15,791
49	116,213	1,589	13,516
51	114,075	1,520	13,516
53	111,007	1,518	13,516
55	106,620	1,578	13,477
57	102,950	1,570	13,477
59	99,699	1,535	13,477
61	95,943	1,487	18,040
63	92,618	1,430	16,870
65	89,107	1,380	12,196
67	85,182	1,329	10,842
69	81,267	1,282	9,552
71	77,039	1,221	9,544
73	72,942	1,147	9,544
75	68,746	1,080	9,544
77	64,631	1,005	7,552
79	60,073	932	7,552
81	60,073	932	7,552
83	60,073	932	7,552
85	60,073	932	7,552

k-mer	No. of contigs	N50	Max. contig length (bp)
87	60,073	932	7,552
89	60,073	932	7,552
91	60,073	932	7,552
93	60,073	932	7,552
95	60,073	932	7,552
97	60,073	932	7,552
99	60,073	932	7,552

(continued)

Table 3. The number of contigs, N50 and maximum contig length of Oases products

k-mer	No. of contigs	N50	Max. contig length (bp)
47	95,114	1,851	16,263
51	95,285	1,848	16,263
55	95,582	1,836	16,872
59	96,298	1,817	16,872
63	96,470	1,818	16,872
67	96,493	1,818	16,872

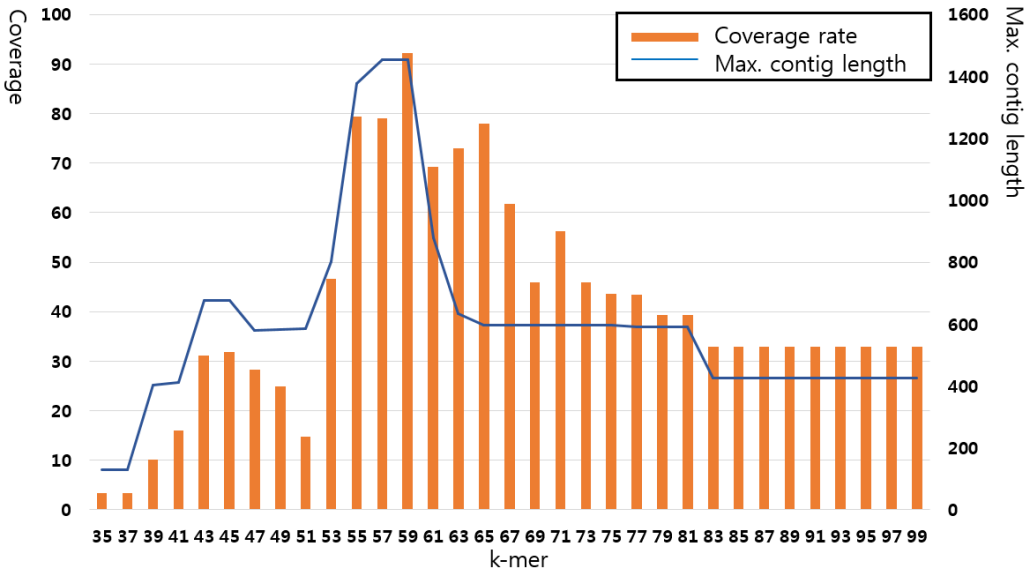


Figure 1. Maximum contig length and coverage rate of the L^4 functional homologue of *de novo* assembly of transcriptome obtained from *C. chacoense* 'PI260429' according to k-mer (35 to 99). Bars indicate coverage rate and line indicates maximum contig length. Both coverage rate and maximum contig length reached a peak when k-mer value is 59.

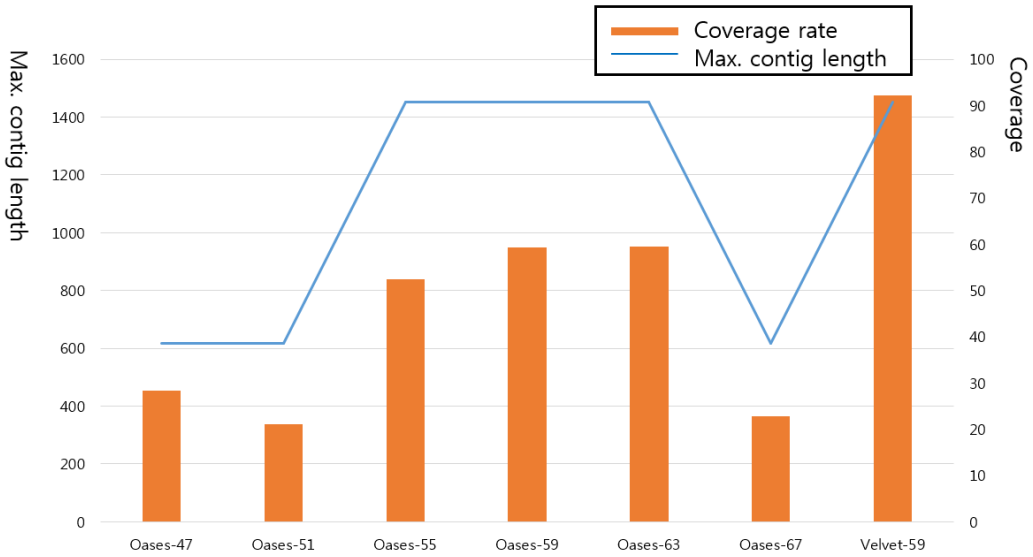


Figure 2. Comparison of coverage rates and maximum contig length of L functional homologue analyzed by Oases (k-mer = 47 - 67) and Velvet (k-mer = 59). Bars indicate coverage rate and line indicates maximum contig length. Maximum contig length of Velvet product was almost identical with Oases product when k-mer values were 55, 59 and 63. However coverage of Velvet product was higher than any other Oases products.

Annotation and short read mapping

A total of 140 contigs homologous to the L^4 candidate gene were obtained from *de novo* assembly products of PI260429. Ninety six contigs were then selected after removing redundant sequences (Table 4). These contigs showed 77 – 100 % identity with the L^4 candidate, and this result indicated that these contigs are all homologues of the L^4 candidate gene, furthermore these contigs possibly can be homologues of the L^4 gene.

Four contigs, *L*_homologue000001, *L*_homologue000101, *L*_homologue_000114, and *L*_homologue_134, perfectly matched with the L^4 candidate gene which was used for the validation of *de novo* assembly (Fig 3). These contigs contained a part of NB-ARC domain or LRR domain. Most of the contigs encode a part of CC, NB-ARC, and LRR domains. Some contigs were comprised of more than two domains and only two contigs, *L*_homologue000020 and *L*_homologue000124, contained all domains (CC, NB-ARC, and LRR). These results indicated that although Velvet assembly with high k-mer can guarantee high specificity, contigs may contained only partial sequence of whole transcripts.

Short reads mapping of transcriptome of PI260429 was performed on the L^4 candidate gene to figure out why longer contig covering whole the L^4 candidate gene was not assembled (Fig 3). *L*_homologue000001,

*L*_homologue000101, *L*_homologue_000114, and *L*_homologue_134 covered most of regions where the number of short reads reached a peak except CC domain (about 1-500 bp). Only few short reads (less than 25 reads) were mapped on sequence regions between *L*_homologue000101 – *L*_homologue114 and *L*_homologue114 – *L*_homologue000001. Short reads mapping results indicated that short reads are not evenly distributed. The region of which the number of reads was lower than other may draw deficiency interfering *de novo* assembly process.

Table 4. Domain prediction and sequence comparison of *L*-homologues *de novo* assembly products of PI260429 with the L^4 candidate gene

Contig name	Length of contig (bp)	Domain	BLAST with L^4 functional homologue		
			Matching region (bp)	Matching length (bp)	Identity (%)
*L*_homologue000001	1283	LRR	3332 - 3947	615	100.00
*L*_homologue000003	313	LRR	2888 - 3200	312	88.05
*L*_homologue000006	1793	LRR	2792 - 3934	1142	86.90
*L*_homologue000008	2334	LRR	2828 - 3934	1106	86.85
*L*_homologue000009	1545	LRR	2828 - 3934	1106	85.02
*L*_homologue000010	1531	LRR	2828 - 3934	1106	86.90
*L*_homologue000011	2545	LRR	2828 - 3921	1093	86.86
*L*_homologue000012	2256	LRR	2828 - 3934	1106	87.18
*L*_homologue000014	2053	LRR	2149 - 3938	1789	87.59
*L*_homologue000020	2904	CC, NB, ARC, LRR	1 - 2787	2786	87.21
*L*_homologue000022	1324	ARC, LRR	1288 - 2617	1329	87.15
*L*_homologue000025	624	LRR	2264 - 2787	523	89.66
*L*_homologue000026	1168	LRR	2792 - 3317	525	88.53
*L*_homologue000027	1222	CC	1 - 406	405	85.37
*L*_homologue000028	554	CC	1 - 406	405	85.37
*L*_homologue000029	1527	CC	1 - 406	405	85.37
*L*_homologue000030	1474	CC	1 - 406	405	85.37
*L*_homologue000031	2236	LRR	2805 - 3921	1116	83.39
*L*_homologue000032	1665	LRR	2805 - 3921	1116	83.39
*L*_homologue000033	2321	LRR	2805 - 3921	1116	83.39
*L*_homologue000034	723	LRR	2238 - 2930	692	89.38
*L*_homologue000035	589	LRR	2193 - 2781	588	90.56
*L*_homologue000037	708	ARC, LRR	1397 - 2113	716	89.47
*L*_homologue000039	4175	ARC, LRR	1184 - 2113	929	85.85
*L*_homologue000045	917	NB, ARC	583 - 1493	910	91.39
*L*_homologue000048	789	LRR	2279 - 2787	508	83.30

Contig name	Length of contig (bp)	Domain	BLAST with L^4 functional homologue		
			Matching region	Matching length	Identity
*L*_homologue000049	3422	NB, ARC, LRR	2810 - 3565	755	77.37
			685 - 2478	1793	83.46
*L*_homologue000050	3366	NB, ARC, LRR	2810 - 3565	755	77.37
			685 - 2478	1793	83.46
*L*_homologue000051	522	LRR	3379 - 3894	515	88.99
*L*_homologue000052	745	LRR	3361 - 3921	560	87.57
*L*_homologue000053	1182	LRR	1698 - 1960	262	87.22
			1991 - 2792	801	90.10
*L*_homologue000056	1170	CC, NB, ARC	309 - 1491	1182	86.59
*L*_homologue000058	693	LRR	3285 - 3652	367	88.74
*L*_homologue000059	1176	CC, NB	93 - 592	499	82.12
*L*_homologue000060	1141	CC,	1 - 423	422	87.26
*L*_homologue000061	1101	CC, NB	1 - 621	620	84.99
*L*_homologue000062	1791	LRR	3012 - 3496	484	87.60
			3707 - 3937	230	90.21
*L*_homologue000063	950	LRR	3185 - 3921	736	87.38
*L*_homologue000064	822	LRR	3185 - 3917	732	85.89
*L*_homologue000066	1933	LRR	2421 - 3937	1516	88.10
*L*_homologue000069	309	LRR	1796 - 2113	317	88.12
*L*_homologue000070	1154	ARC, LRR	1483 - 2639	1156	91.22
*L*_homologue000071	1092	ARC, LRR	1238 - 2319	1081	82.52
*L*_homologue000075	1179	CC, NB, ARC,	406 - 1487	1081	79.77
*L*_homologue000077	2070	CC, NB, ARC	1 - 1258	1257	83.22
*L*_homologue000079	1570	CC, NB, ARC	344 - 1447	1103	79.48
*L*_homologue000080	1465	LRR	1697 - 2619	922	87.14
*L*_homologue000081	610	LRR	1951 - 2567	616	92.39
*L*_homologue000083	906	NB, ARC	579 - 1383	804	87.47

(continued)

Contig name	Length of contig (bp)	Domain	BLAST with L^4 functional homologue		
			Matching region	Matching length	Identity
*L*_homologue000084	701	NB, ARC	579 - 1274	695	87.99
*L*_homologue000085	274	CC	176 - 449	273	91.27
*L*_homologue000086	592	LRR	2792 - 3372	580	84.85
*L*_homologue000087	1421	LRR	2695 - 2772	77	85.00
*L*_homologue000087	1421	ARC, LRR	1382 - 2619	1237	85.76
*L*_homologue000089	765	CC, NB,	1 - 626	625	96.65
*L*_homologue000090	802	CC, NB,	1 - 626	625	97.12
*L*_homologue000094	755	ARC, LRR	1407 - 2163	756	88.58
*L*_homologue000095	395	CC, NB,	278 - 634	356	87.91
*L*_homologue000097	771	ARC, LRR	1288 - 2049	761	87.07
*L*_homologue000098	996	ARC, LRR	1266 - 2262	996	86.45
*L*_homologue000099	1120	ARC, LRR	1038 - 2153	1115	84.94
*L*_homologue000101	779	NB, ARC	761 - 1539	778	100.00
*L*_homologue000102	289	CC, NB	569 - 857	288	96.54
*L*_homologue000103	1144	CC, NB	1 - 1042	1041	82.46
*L*_homologue000104	723	LRR	3384 - 3921	537	82.58
*L*_homologue000105	958	LRR	3148 - 3937	789	86.49
*L*_homologue000106	574	LRR	1845 - 2378	533	84.86
*L*_homologue000107	287	LRR	2029 - 2321	292	90.48
*L*_homologue000108	589	NB, ARC	621 - 1204	583	84.49
*L*_homologue000109	250	NB	608 - 857	249	96.80
*L*_homologue000110	1041	LRR	2977 - 3564	587	87.37
*L*_homologue000111	1145	LRR	2977 - 3738	761	85.62
*L*_homologue000112	702	LRR	2804 - 3372	568	86.24
*L*_homologue000113	508	LRR	3415 - 3912	497	87.25
*L*_homologue000114	1453	LRR	1694 - 3146	1452	100.00
*L*_homologue000115	1115	LRR	2804 - 3161	357	89.07
			1988 - 2691	703	92.92

(Continued)

Contig name	Length of contig (bp)	Domain	BLAST with L^4 functional homologue		
			Matching region	Matching length	Identity
*L*_homologue000117	741	LRR	3354 - 3911	557	85.22
*L*_homologue000118	829	LRR	1217 - 2045	828	92.80
*L*_homologue000119	414	LRR	2059 - 2475	416	89.07
*L*_homologue000120	441	LRR	2029 - 2475	446	92.84
*L*_homologue000123	1181	NB, ARC, LRR	627 - 1795	1168	80.94
*L*_homologue000124	1933	CC, NB, ARC, LRR	29 - 1795	1766	82.57
*L*_homologue000125	613	LRR	1777 - 2386	609	89.76
*L*_homologue000127	300	CC	41 - 340	299	92.67
*L*_homologue000128	463	LRR	3338 - 3858	520	87.30
*L*_homologue000129	549	ARC, LRR	1181 - 1740	559	90.91
*L*_homologue000130	340	CC, NB	269 - 608	339	87.28
*L*_homologue000131	667	LRR	2680 - 3304	624	90.23
*L*_homologue000132	285	ARC, LRR	1451 - 1735	284	92.01
*L*_homologue000133	572	CC, NB	1 - 556	555	84.56
*L*_homologue000134	275	NB	583 - 857	274	100.00
*L*_homologue000135	455	LRR	1661 - 2115	454	92.12
*L*_homologue000136	418	LRR	1698 - 2115	417	91.94
*L*_homologue000137	544	CC, NB	66 - 547	481	86.24
*L*_homologue000138	337	LRR	3148 - 3481	333	88.72
*L*_homologue000139	376	CC, NB	230 - 608	378	85.94
*L*_homologue000140	252	ARC	1288 - 1539	251	94.44

(Continue)

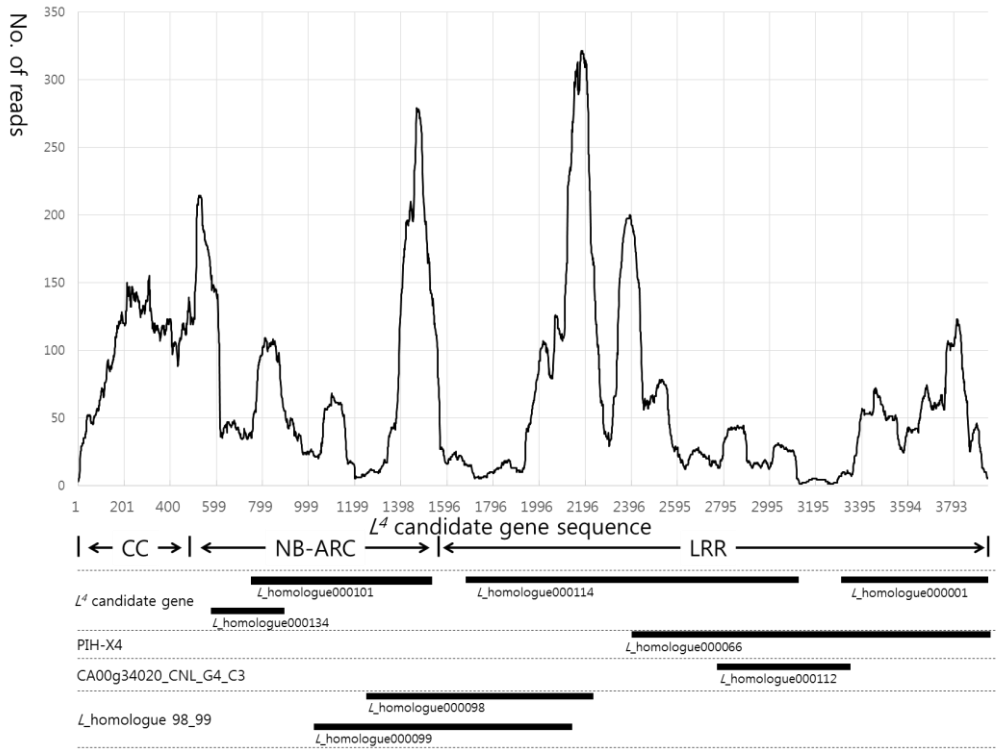


Figure 3. Short reads mapping of PI260429 to the L^4 candidate gene. Eight bars indicate the relative locations of resistance accession-specific expressed contigs derived from two transcripts, L^4 candidate gene and PIH-X4, and two putative transcripts, CA00g34020_CNL_G4_C3, and *L*_homologue 98_99.

***in silico* mapping of *L* homologues**

A total of 47 contigs were mapped on chromosome 04 (6 contigs), chromosome 05 (1 contig), chromosome 06 (4 contigs), chromosome 07 (2 contigs) and chromosome 11 (34 contigs) (Table 5). Two of contigs, *L*_homologue000060 and *L*_homologue000083, were not mapped on specific chromosome because these contigs were matched with not only chromosome 11 but also other chromosomes. This indicated that most of homologues derived from one ancestor are located in main gene complex region, however, some homologues also can be located other region. The other 47 contigs did not matched with any other sequences of chromosome WGS data set ver. 1.35. However additional 30 contigs were matched in scaffold WGS data set ver. 1.35. The other 17 contigs did not matched with any sequences in contig WGS data set ver. 1.35, as well (Table 5). This result showed that scaffold data set contains much more homologous data than chromosome data set. And 17 contigs which did not matched with any sequence of WGS might be *C. chacoense*-specific homologues or result of misassembly.

Thirty four *L*-homologous contigs anchored on chromosome 11 were located in the region from 189,753,791 bp to 244,409,999 bp of Pepper.v.1.35.Final.chr11 sequence. Some contigs such as *L*_homologue000049 and *L*_homologue000050 or *L*_homologue000031 and L_homologue000032

were mapped on same sequence regions. Contigs located on the same region shared 100 % sequence identity in most of region, but there was long insertion/deletion between them (Table 5). Three contigs such as *L*_homologue000039, *L*_homologue000133, and *L*_homologue000022 were mapped on a single location in chromosome data set whereas these contigs were matched with two different location in scaffold data set. *L*_homologue000069 was matched with more than two sequence region in chromosome and scaffold data set.

Table 5. *in silico* mapping of contigs of *L* homologues in *C. chacoense* 'PI260429'

Contig name	Chromosome ver. 1.35			Scaffold ver. 1.35		
	Chromosome	Start	end	Scaffold	start	end
*L*_homologue000061	chr11	189753791	189755804	71	4135737	4137750
*L*_homologue000049	chr11	189756476	189760625	71	4138422	4142571
*L*_homologue000050	chr11	189756476	189760625	71	4138422	4142571
*L*_homologue000048	chr11	229809770	229810558	553	391453	392241
*L*_homologue000071	chr11	229810519	229811610	553	392202	393293
*L*_homologue000103	chr11	229811793	229812893	553	393476	394576
*L*_homologue000075	chr11	236395670	244250535	309	481068	482143
*L*_homologue000077	chr11	240832579	240834403	742	43202	45026
*L*_homologue000080	chr11	240834364	240835928	742	44987	46551
*L*_homologue000039	chr11	244095590	244102090	309	327198	333698
				309	571138	580001
*L*_homologue000097	chr11	244250327	244251097	309	481935	482705
*L*_homologue000031	chr11	244251918	244253024	309	483526	484632
*L*_homologue000032	chr11	244251918	244253024	309	483526	484632
*L*_homologue000033	chr11	244251918	244253024	309	483526	484632
*L*_homologue000101	chr11	244302957	244356246	3753	970	1694
*L*_homologue000020	chr11	244323605	244325386	309	555213	556994
*L*_homologue000133	chr11	244323619	244323722	1149	216431	216535
				631	1320943	1321047
*L*_homologue000053	chr11	244355434	244385098	309	587042	587538
*L*_homologue000140	chr11	244356089	244356246	1524	248077	248328
*L*_homologue000014	chr11	244357738	244387006	309	589346	618614
*L*_homologue000090	chr11	244365517	244366313	309	597125	597921
*L*_homologue000109	chr11	244366295	244366544	309	597903	598152
*L*_homologue000045	chr11	244366723	244384628	309	615323	616236
*L*_homologue000069	chr11	244367558	244367866	309	587132	587440
		244384931	244385239	309	599166	599474
				309	616539	616847
*L*_homologue000089	chr11	244383102	244383758	309	614710	615366

Contig name	Chromosome ver. 1.35			Scaffold ver. 1.35		
	Chromosome	Start	end	Scaffold	start	end
*L*_homologue000102	chr11	244383701	244383989	309	615309	615597
*L*_homologue000134	chr11	244383715	244383989	3753	1696	1966
*L*_homologue000006	chr11	244388018	244388615	309	619626	620223
*L*_homologue000008	chr11	244388018	244409999	309	619626	641607
*L*_homologue000009	chr11	244388018	244409999	309	619626	641607
*L*_homologue000010	chr11	244388018	244388615	309	619626	620223
*L*_homologue000012	chr11	244388018	244388615	309	619626	620223
*L*_homologue000056	chr11	244400957	244401056	631	1189637	1190800
*L*_homologue000011	chr11	244409876	244409999	1848	75280	641607
*L*_homologue000051	chr07	120911512	120911685	1358	18446	18966
*L*_homologue000022	chr07	120913321	120913838	182	359507	360024
				1848	39236	40002
*L*_homologue000087	chr06	170107054	170108448	963	836907	838301
*L*_homologue000117	chr06	170108438	170109248	963	838291	839101
*L*_homologue000106	chr06	199218668	199219242	1647	50558	51132
*L*_homologue000108	chr06	199219853	199220442	1647	51743	52332
*L*_homologue000059	chr05	3399622	3399503	525	821688	826849
*L*_homologue000027	chr04	14169803	14178293	354	1831787	1840277
*L*_homologue000030	chr04	14169803	14178293	354	1831787	1840277
*L*_homologue000029	chr04	14173217	14178293	354	1831787	1840277
*L*_homologue000064	chr04	14176988	14177089	1524	225202	225767
*L*_homologue000028	chr04	14177743	14178293	354	1839727	1840277
*L*_homologue000079	chr04	14178560	14180132	354	1840544	1842116
*L*_homologue000060	Matched but not specified			1848	42131	45713
*L*_homologue000083	Matched but not specified			1848	3486	89348
*L*_homologue000001	Not detected			2377	8286	10270
*L*_homologue000026	Not detected			1209	34086	34441
*L*_homologue000062	Not detected			1524	238775	239476
*L*_homologue000066	Not detected			1632	41824	45454

(continued)

Contig name	Chromosome ver. 1.35			Scaffold ver. 1.35		
	Chromosome	Start	end	Scaffold	start	end
*L*_homologue000070		Not detected		1848	87268	88421
*L*_homologue000081		Not detected		1149	219510	220120
*L*_homologue000084		Not detected		1848	89178	89348
*L*_homologue000086		Not detected		1887	47873	47979
				631	1320628	1320734
*L*_homologue000094		Not detected		631	1188965	1189721
*L*_homologue000098		Not detected		631	1348036	1349031
*L*_homologue000099		Not detected		631	1348136	1349253
*L*_homologue000104		Not detected		1848	85790	86511
*L*_homologue000105		Not detected		1848	37715	38667
*L*_homologue000110		Not detected		1524	248722	249762
*L*_homologue000111		Not detected		1524	248722	249567
*L*_homologue000112		Not detected		1848	86523	87225
*L*_homologue000113		Not detected		1149	249415	249567
*L*_homologue000114		Not detected		3753	1	752
*L*_homologue000115		Not detected		1848	78863	79976
*L*_homologue000120		Not detected		1524	234579	235019
*L*_homologue000123		Not detected		1632	47665	48002
				1887	130821	130935
*L*_homologue000124		Not detected		1524	232435	234272
*L*_homologue000127		Not detected		631	1190769	1191068
*L*_homologue000128		Not detected		546	92374	92831
*L*_homologue000129		Not detected		1848	39945	40186
*L*_homologue000130		Not detected		1848	89319	89656
*L*_homologue000131		Not detected		1149	220231	220898
*L*_homologue000132		Not detected		1524	248240	248524
*L*_homologue000135		Not detected		2377	11487	11941
*L*_homologue000136		Not detected		1632	46810	46928
				3753	268	416

(continued)

Contig name	Chromosome ver. 1.35			Scaffold ver. 1.35		
	Chromosome	Start	end	Scaffold	start	end
*L*_homologue000003		Not detected			Not detected	
*L*_homologue000025		Not detected			Not detected	
*L*_homologue000034		Not detected			Not detected	
*L*_homologue000035		Not detected			Not detected	
*L*_homologue000037		Not detected			Not detected	
*L*_homologue000052		Not detected			Not detected	
*L*_homologue000058		Not detected			Not detected	
*L*_homologue000063		Not detected			Not detected	
*L*_homologue000085		Not detected			Not detected	
*L*_homologue000095		Not detected			Not detected	
*L*_homologue000107		Not detected			Not detected	
*L*_homologue000118		Not detected			Not detected	
*L*_homologue000119		Not detected			Not detected	
*L*_homologue000125		Not detected			Not detected	
*L*_homologue000137		Not detected			Not detected	
*L*_homologue000138		Not detected			Not detected	
*L*_homologue000139		Not detected			Not detected	

(continued)

Digital expression of *L* homologues

Digital expression analysis of 96 *L*-homologues of PI260429 on four *Tobamovirus*-resistance accessions (Yuwolcho, CM334, YCM334, PI152225 and PI260429) and four *Tobamovirus*-susceptible accessions (YCM334, Lam32, Jeju and Tean) was performed (Table 6). Expression levels of contigs were clearly distinguished into expressed group (expression level > 2) and un-expressed group (expression level = 0). Most of *L*-homologues were not expressed in the three susceptible accessions, LAM32, Jeju and Tean while these were expressed in the two resistant accessions, CM334 and PI260429. And about two third of *L*-homologues were expressed in the two resistance accessions, Yuwolcho and PI152225, and one susceptible accession YCM334.

Sequence of eight contigs were expressed only in four resistance accessions, Yuwolcho, CM334, PI152225, and PI260429 (Table 6, Fig 3). Four contigs, *L*_homologue000001, *L*_homologue000101, *L*_homologue_000114, and *L*_homologue_134, shared sequence perfectly with the L^4 functional homologue. The others four contigs, *L*_homologue000066, *L*_homologue000112, *L*_homologue000098, and *L*_homologue000099, showed 85-86% identity with the L^4 functional homologue. Currently isolated NBS-LRR of *Capsicum* and published *L*-homologue were used for the identification of transcript of these contigs (Tomita et al., 2008; Yang et al. 2009). *L*_homologue000066 and

*L*_homologue 000112 were derived from PIH-X4 and CA00g34020_CNL_G4_C3, respectively. However *L*_homologue000098 and *L*_homologue000099 were not matched with any NBS-LRR and *L* homologue sequences found in *Capsicum*. These two contigs shared sequence over 99.5%, therefore these were supposed to be originated from a novel putative transcript designated to *L*_homologue 98_99. PIH-X4, CA00g34020_CNL_G4_C3, the L^4 candidate and *L*_homologue 98_99 were expressed only in *Tobamovirus*-resistant accessions. One of transcript, the L^4 candidate was known to be an important factor to induce Tobamovirus resistance. The other three transcripts also may related to the function of *Tobamovirus* resistance.

Table 6. Digital expression of *L* homologues in eight *Capsicum* accessions. Non-expressed contigs in the table are shaded.

Feature ID	Digital expression (common logarithm value)								Chr
	Yuwolcho	CM334	PI152225	PI260429	YCM334	LAM32	Jeju	Tean	
*L*_homologue000006	2.4	3.7	2.8	3.8	2.6	0	0	0	chr11
*L*_homologue000011	4	3.9	2.3	3.9	3.1	0	0	0	chr11
*L*_homologue000014	3.1	3.7	3.9	4.1	3.7	0	0	0	chr11
*L*_homologue000027	3.7	3.6	3.7	4	3.6	0	0	0	chr04
*L*_homologue000028	3.7	3.9	3.3	3.2	4	0	0	0	chr04
*L*_homologue000029	4.3	3.9	4.2	3.9	4.3	0	0	0	chr04
*L*_homologue000030	3.7	3.1	3.3	3.3	3.4	0	0	0	chr04
*L*_homologue000031	3.1	3	3	3.6	3.4	0	0	0	chr11
*L*_homologue000032	3.3	3.6	3.2	3.7	3.7	0	0	0	chr11
*L*_homologue000033	3.1	3.7	3.6	3.6	3.5	0	0	0	chr11
*L*_homologue000048	4.6	4.2	4.6	3.5	4.4	0	0	0	chr11
*L*_homologue000049	4.1	3.9	4.1	3.9	4.4	0	0	0	chr11
*L*_homologue000050	3.3	3.1	3.7	3.6	3.6	0	0	0	chr11
*L*_homologue000051	5	4.4	3.3	3.8	4.9	0	0	0	chr07
*L*_homologue000039	2.6	2.4	2.8	3.5	2.8	0	0	0	chr11
*L*_homologue000056	3.5	3.6	3.9	4	4.2	0	0	0	chr11
*L*_homologue000059	3.9	4.2	3.6	3.9	4.5	0	0	0	chr01
*L*_homologue000061	3.9	4.1	3.9	3.7	4.2	0	0	0	chr11
*L*_homologue000062	3.4	3.7	3.4	3.8	3	0	0	0	-
*L*_homologue000064	4.7	3.8	5	3.7	4.7	0	0	0	chr04
*L*_homologue000070	3.3	3.9	2.6	3.8	3.4	0	0	0	-
*L*_homologue000071	4.7	4.2	4.8	3.9	4.5	0	0	0	chr11
*L*_homologue000075	3.8	4.1	3.3	3.8	3.8	0	0	0	chr11
*L*_homologue000077	4.2	4.2	4.1	3.9	4.4	0	0	0	chr11
*L*_homologue000079	4.7	4.3	4.7	4.5	4.5	0	0	0	chr04
*L*_homologue000080	4.6	4.4	4.6	3.9	4.7	0	5.4	0	chr11
*L*_homologue000083	2.7	3	3.1	4.1	2.9	0	0	0	chr05
*L*_homologue000084	3.7	3.7	3.5	4	3.7	0	0	0	-
*L*_homologue000087	3.4	3.9	3.7	3.8	4.1	0	0	0	chr06
*L*_homologue000089	3.9	4.3	3.3	3.9	3.6	0	0	0	chr11
*L*_homologue000090	2.8	4.1	3.5	3.9	2.9	0	0	0	chr11
*L*_homologue000094	3.7	3.9	3.5	4	3.7	0	0	0	-
*L*_homologue000097	3.6	4.3	3.4	3.6	4	0	0	0	chr11
*L*_homologue000103	4.4	4	4.8	3.8	4.2	0	0	0	chr11
*L*_homologue000106	3.4	3.9	4	3.9	4.2	0	0	0	chr06

Feature ID	Digital expression (common logarithm value)								Chr
	Yuwolcho	CM334	PI152225	PI260429	YCM334	LAM32	Jeju	Tean	
*L*_homologue000108	2.9	3.6	4	3.6	3.8	0	0	0	chr06
*L*_homologue000110	4.2	4.2	4.2	3.6	4.3	0	0	0	-
*L*_homologue000111	4.1	3.8	2.7	3.4	4	0	0	0	-
*L*_homologue000113	4.7	4.3	4.3	3.4	4.8	0	0	0	-
*L*_homologue000115	3.8	3.9	2.7	3.7	3.3	0	0	0	-
*L*_homologue000117	3.5	3.4	4.1	3.7	4.3	0	0	0	chr06
*L*_homologue000120	4.4	4.5	4.1	3.5	4.3	0	0	0	-
*L*_homologue000124	4.4	4.4	4.2	3.7	4.3	0	0	0	-
*L*_homologue000129	3.2	4.2	3.7	3.4	3.8	0	0	0	-
*L*_homologue000135	4.1	4.4	4	3.2	4.2	0	0	0	-
*L*_homologue000123	3.5	3.8	3.1	0	3.6	0	0	0	-
*L*_homologue000022	3.4	3.6	0	3.6	3.2	0	0	0	chr07
*L*_homologue000026	3.4	3.7	0	3.7	4.1	0	0	0	-
*L*_homologue000060	4	3.7	0	3.7	3.8	0	0	0	chr04
*L*_homologue000105	4.4	4	0	4	4.5	0	0	0	-
*L*_homologue000127	3.7	4.7	0	3.9	4	0	0	0	-
*L*_homologue000001	3.5	3.9	4.5	3.5	0	0	0	0	-
*L*_homologue000008	2.3	2.3	2.3	4.2	0	0	5.2	0	chr11
*L*_homologue000066	4.5	4.2	4.2	3.8	0	0	0	0	-
*L*_homologue000098	3.8	3.6	3.6	3.2	0	0	0	0	-
*L*_homologue000099	4.3	4	4.1	3.5	0	0	0	0	-
*L*_homologue000101	3.8	4.5	4.5	3.8	0	0	0	0	chr11
*L*_homologue000112	3.1	3.1	2.9	3.7	0	0	0	0	-
*L*_homologue000114	4.3	4.4	4.5	3.9	0	0	0	0	-
*L*_homologue000134	3.2	4.4	3.3	3.2	0	0	0	0	chr11
*L*_homologue000025	0	4	3.2	3.9	3	0	0	0	-
*L*_homologue000035	0	3.8	2.9	4	3.1	0	0	0	-
*L*_homologue000052	0	3.3	4.1	4.2	3	0	0	0	-
*L*_homologue000053	0	3.9	3.1	3.6	3.1	0	0	0	chr11
*L*_homologue000069	0	4.3	3.9	3.5	3.6	0	0	0	chr11
*L*_homologue000081	0	4	2.9	3.4	3.5	0	0	0	-
*L*_homologue000104	0	3.3	3.2	4.1	3.4	0	0	0	-
*L*_homologue000133	2.9	4.3	0	3.4	0	0	0	0	chr11
*L*_homologue000012	0	2.5	0	2.3	2.5	0	0	0	chr11
*L*_homologue000131	0	3.8	0	3.3	3.7	0	0	0	-
*L*_homologue000137	0	3.4	0	3.4	3.4	0	0	0	-
*L*_homologue000128	3.6	0	3	3.5	0	0	0	0	-
*L*_homologue000102	0	4.1	3.7	2.9	0	0	0	0	chr11
*L*_homologue000139	0	3.5	3.6	3.5	0	0	0	0	-

(Contignued)

Feature ID	Digital expression (common logarithm value)								Chr
	Yuwolcho	CM334	PI152225	PI260429	YCM334	LAM32	Jeju	Tean	
*L*_homologue000140	0	3.9	3.3	3.2	0	0	0	0	chr11
*L*_homologue000003	0	3.3	0	3.7	0	0	0	0	-
*L*_homologue000009	0	2.8	0	3.8	0	0	0	0	chr11
*L*_homologue000020	0	3.4	0	3.8	0	0	0	0	chr11
*L*_homologue000034	0	3.7	0	4.2	0	0	0	0	-
*L*_homologue000037	0	3.7	0	3.1	0	0	0	0	-
*L*_homologue000045	0	3.2	0	3.5	0	0	0	0	chr11
*L*_homologue000058	0	3.2	0	4.2	0	0	0	0	-
*L*_homologue000085	0	4.1	0	3.2	0	0	0	0	-
*L*_homologue000086	0	3.3	0	4.1	0	0	0	0	-
*L*_homologue000095	0	2.7	0	3.6	0	0	0	0	-
*L*_homologue000107	0	3.2	0	3.8	0	0	0	0	-
*L*_homologue000109	0	3.4	0	2.9	0	0	0	0	chr11
*L*_homologue000118	0	3.9	0	3.6	0	0	0	0	-
*L*_homologue000119	0	3.2	0	3	0	0	0	0	-
*L*_homologue000125	0	3.9	0	3.4	0	0	0	0	-
*L*_homologue000130	0	3.5	0	3.4	0	0	0	0	-
*L*_homologue000132	0	4.1	0	3.5	0	0	0	0	-
*L*_homologue000136	0	3.9	0	3	0	0	0	0	-
*L*_homologue000138	0	3.3	0	3.5	0	0	0	0	-
*L*_homologue000010	0	0	0	4.7	0	0	0	0	chr11
*L*_homologue000063	0	0	0	4.1	0	0	0	0	-

(Continued)

DISCUSSION

One of the major features of plant resistance gene is that belongs to the multigene-family. For that reason of redundant sequences exist in the genome (Meyer et al. 2003). Because of high similarity between paralogs, high specificity and accuracy are required for precise *de novo* assembly of a resistance gene. Although many kinds of *de novo* assemblers for transcriptome have been developed, only few assembler based on de Bruijn graph system have been used recently (Birol et al. 2009; Martin and Wang 2012; Schulz et al. 2012; Zerbino et al. 2008). A k-mer is a standard length of de Bruijn algorithm determining whether two short sequences are overlapped or broken (Wang et al. 2009). Low k-mer is not proper for de novo assembly of highly redundant sequences derived from multiple paralogs. The k-mer of Trinity assembler is fixed at 25 which are not appropriate for this purpose. The k-mer of Velvet and Abyss can be altered by user, manually. Abyss program produce final product by merging and reassembly of wide range of k-mer assembled products automatically. Velvet-Oases program follows a similar *de novo* assembly strategy, however, single k-mer assembly and merging/reassembly step are separated. Therefore, Velvet-Oases is chosen for finding optimal conditions for *de novo* assembly of specific

highly-repetitive sequences because detailed adjustment of parameters and steps were possible.

Validation of two indexes of single k-mer assembly of PI260429 demonstrated that k-mer = 59 is an optimal condition for *de novo* assembly of *L*-homologue sequences (Fig 1). The average length of short reads of Illumina platform is less than 100 bp. k-mer = 59 means that the short reads shared over 60% of their own length can be assembled. Higher k-mer (more than 59) may bring higher specificity, but assembly products can be lost because short reads sharing more than 60% of sequence identity are few. In spite of high specificity of the PI260429 when k-mer is 59, k-mer 59 may not be suitable for other cases of *de novo* assembly of NBS-LRR genes or even *de novo* assembly of *L*-homologues in other accessions depending on the length of target resistance gene or depth of transcriptome. However higher k-mer (more than 53, lower than 61) would be required for *de novo* assembly to get high specificity for any kind of multigene family sequences. It needs to be validated by sequencing analysis, however I showed a possible strategy to assemble a multigene family gene using transcriptome data.

Oases program is designed for merging and reassembly, however it was not possible to get higher indexes compared to single k-mer Velvet assembly products (Fig. 2). Maximum contig length of Oases products was identical with

Velvet products while coverage rate of Oases was decreased. This suggests that relatively short sequence fragments were lost during the assembly step using Oases.

Four out of 96 *L*-homologue contigs obtained by *de novo* assembly product of PI260429 perfectly matched with the L^4 candidate gene. A single contig sequence covering entire the L^4 functional homologue sequence could not be obtained. Short read mapping of PI26429 transcriptome into the L^4 candidate may give us why a single contig could not be obtained (Fig 3). Short reads were not evenly distributed on the L^4 candidate gene. This result indicated that short read information in some region may not be enough to assemble a entire gene sequence. Highly conserved region may also disturb assembly. If highly conserved region is longer than 100 bp, it is very difficult to assemble using de Bruijn graph because k-mer is less than 100.

The number of matching *L*-homologuess contigs mapped on pepper chromosome ver. 1.35 and scaffold ver 1.35 were 47 and 77, respectively. This indicates that numerous *L*-homologue contigs were lost during pseudo-molecule construction. Highly repetitive sequence such as NBS-LRR class could be lost because of its redundancy. Therefore, scaffold data set would be proper for isolation of multigene family than chromosome data set. However physical location cannot be found in scaffold data set alone. So, both of scaffold data set

and contig data set should be used for isolation and detection of physical location of highly repetitive sequences.

Digital expression of *L*-homologue contigs of eight accessions demonstrated that the expression patterns of eight accessions could be clearly distinguished into three groups; 1) most of *L*-homologue contigs expressed (PI260429 and CM334), 2) most of the *L*-homologue contigs were not expressed (LAM32, Jeju, and Tean), 3) about two third of *L*-homologue contigs were expressed (Yuwolcho, PI152225, and YCM334). Phylogenetic analysis result between L^3 and other orthologs was consistent with digital expression pattern of eight accessions (Fig 4). Expression of *L*-homologue contigs of PI260429 (L^4) and CM334 (L^{2b}) that are belonged to group 2 were closely related in the phylogenetic tree as well. Although I have not analyzed all of accessions containing *L* alleles, this demonstrates that expression patterns of *L* homologues are correlated with the phylogenetic relationship of resistance genes.

Eight contigs were expressed only in resistance accessions. Sequence comparison with NBS-LRR and *L* homologues in *Capsicum* demonstrated that these eight contigs derived from four transcripts. Four contigs perfectly matched with the L^4 candidate. This indicates that the other three *Tobamovirus* resistance specifically expressed transcripts, PIH-X, CA00g34020_CNL_G4_C3, and *L*_homologue 98_99, can also play important role for *Tobamovirus* resistance.

However *in vitro* or *in vivo* experiments should be performed to analyze the function of these transcripts. Co-segregation analysis using the L^4 candidate specific marker demonstrated that the L^4 candidate and resistance candidate gene were closely linked to each other (Yang et al. 2012). Therefore if three transcripts identified in this study are responsible for *Tobamovirus* resistance, these should be located near L^4 functional homologue. PIH-X is known to be located in a same contig with the L^4 candidate (Tomita et al 2011). However the physical location of other two transcripts, CA00g34020_CNL_G4_C3 and *L*_homologue 98_99, were not determined yet.

RGAs were known to be generated by tandem duplications induced by transposable elements activities (Ratnaparkhe 2010). The function of RGAs is not clear so far, however few cases demonstrating that RGAs can be a responsible factor for the disease resistance were reported in rice (Hayashi and Yoshida 2009; Okuyama et al 2011). A long terminal repeats (LTRs) retrotransposon is inserted in the upstream region of unexpressed 'sleeping' rice blast resistance gene, *Pit* (Hayashi and Yoshida 2009). It turned on expression of *Pit,* and then resistance was activated. This indicates that homologues can be 'awoken'. Multifaceted genomic research revealed that the two NBS-LRR class RGAs that are tightly linked to blast disease resistance gene, *Pia,* are necessary for the full resistance (Okuyama et al 2011). This showed an example that RGA assists resistance of a

R gene.

Interestingly, one of susceptible accession, YCM334 belonged to group 3. YCM334 is susceptible accession but it has totally different history with other three susceptible accessions. YCM334 is one of the individual of F_6 RIL population derived from the cross of Yolo Wonder and CM334 containing resistance alleles, L^1 and L^2, respectively. The other susceptible accessions are landraces of Korea and India (Bonnet et al 2007; Gebre-Selassie and Marchoux 1991). LAM32, Jeju, and Tean might not have crossed with resistant accessions while genetic background of YCM334 were derived from resistance accessions. It is not certain when YCM334 had lost *Tobamovirus* resistance, however *L* homologues which were expressed in the past are remained until now.

Total RNA of eight accessions were extracted from leaf which were not inoculated by *Tobamovirus*. Therefore the expression level of the L^4 candidate and other *L* homologue were expressed constitutively. In this regard, what *L* homologues were not expressed in LAM32, Jeju, and Tean demonstrated that *L* homologues of these accessions were not constitutively able to be expressed or *L* homologues were not included in their genome. Expression of *L* homologue is not regulated by environmental or *Tobamovirus* infection. Internal genetic factors rather than outer factors seems to affect the expression of *L* homologues. An evolutionary event such as speciation appears to affect not only expression of

resistance genes but also resistance gene homologues. Regulation of gene expression may have been maintained after the evolutionary event.

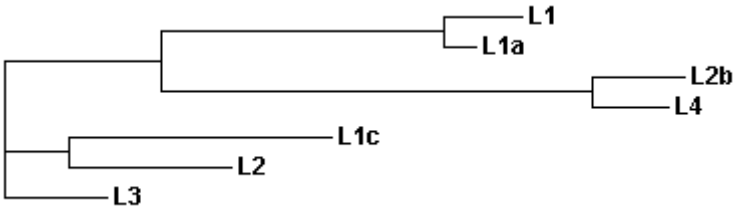


Figure 4. Phylogenetic tree of *L* alleles. L^4 and L^{2b} alleles and L^1 and L^{1a} alleles were closely related.

REFERENCES

Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R, Stazyk G, Morin RD, Zhao Y, Hirst M, Schein JE, Horsman DE, Connors JM, Gascoyne RD, Marra MA, Jones SJM (2009) *De novo* transcriptome assembly with ABySS. Bioinformatics 25:2872-2877

Bonnet J, Danan S, Boudet C, Barchi L, Sage-Palloix AM, Caromel B, Palloix A. & Lefebvre V (2007). Are the polygenic architectures of resistance to *Phytophthora capsici* and P-parasitica independent in pepper? Theoretical and Applied Genetics 115: 253–264

Clarke K, Yang Y, Marsh R, Xie LL, Zhang KK (2013) Comparative analysis of *de novo* transcriptome assembly. Science China Life Science 56:156-162

Gebre-Selassie K and Marchoux G (1991) Identification and characterization of Tobamoviruses strains infecting *L*-resistant genotype of pepper in France. Journal of Phytopathology 131:275-289

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N & Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29:644-652

Hayashi K and Yoshida H (2009) Refunctionalization of the ancient rice blast

disease resistance gene *Pit* by the recruitment of a retrotransposon as a promoter. The Plant Journal 57:413-425

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.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome project data processing subgroup (2009) The sequence alignment/Map format and SAMtools. Bioinformatics 25:2078-2079

Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754-1760

Li W and Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends in Genetics 24:142-149

Martin and Wang (2012) Next-generation transcriptome assembly. Nature Reviews Genetics 12:671-682

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant Journal 20:317-332

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell 15:809-834

Metzker ML (2010) Sequencing technologies - the next generation. Nature Reviews Genetics 11:31-46

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

Okuyama Y, Kanzaki H, Abe A, Yoshida K, Tamiru M, Saitoh H, Fujibe T, Matsumura H, Shenton M, Galam DC, Undan J, Ito A, Sone T, Terauchi R (2011) A multifaceted genomics approach allows the isolation of the rice *Pia*-blast resistance gene consisting of two adjacent NBS-LRR protein genes. The Plant Journal 66:467-479

Paszkiewicz K and Studholme DJ (2010) De novo assembly of short sequence read. Briefings in Bioinformatics 11:457-372

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651-652

Ratnaparkhe MB, Wang X, Li J, Compton RO, Rainville LK, Lemke C, Kim C, Tang H and Paterson SH (2010) Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny. New Phytologist 192:164-178

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086-1092

Tomita R, Sekine KT, Mizumoto H, Sakamoto M, Murai J, Kiba A, Hikichi Y, Suzuki K, Kobayashi K (2011) Genetic basis for the hierarchical interaction between Tobamoviruses and *L* resistance gene alleles from different pepper species. Molecular Plant-Microbe Interactions 24:108-117

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7:562-578

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall1 K, Lee S, Okada1 HM, Qian JQ, Griffith M, Raymond A, Thiessen N, Cezard T, Butterfield YS, Newsome R, Chan SK, She R, Varhol R, Kamoh B, Prabhu AL, Tam A, Zhao YJ, Moore RA, Hirst M, Marra MA, Jones SJM, Hoodless PA, Birol I (2010) *De novo* assembly and analysis of RNA-seq data. Nature Methods 7:909-912

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Review Genetics 10, 57-63

Yang HB, Liu WY, Kang WH, Jahn M, Kang BC (2009) Development of SNP markers linked to the *L* locus in *Capsicum* spp. by a comparative genetic analysis. Molecular Breeding Breed 24:433–446

Yang HB, Liu WY, Kang WH, Kim JH, Cho HJ, Yoo JH, Kang BC (2012)

Development and validation of *L* allele-specific markers in *Capsicum*. Molecular Breeding 30:819-829

Zerbino DR and Birney E (2008) Velvet: Algorithms for *de novo* short read assembly using de Bruijin graphs. Genome Research 18:821-829

초 록

토바모바이러스(Tobamovirus)는 고추에서 가장 큰 피해를 주는 바이러스 중 하나이다. 바이러스는 화학적 약제에 의한 방제가 거의 불가능하고 식물 위생만으로는 토바모바이러스 감염을 막는 데 한계가 있다. 토바모바이러스를 방제하는 가장 좋은 방법은 저항성 품종을 재배하는 것으로 방제에 필요한 노력이나 비용을 절감할 수 있다. 토바모바이러스 저항성 유전자에 대한 분자유전학 연구는 토바모바이러스 저항성에 대한 깊은 이해를 가능케하며 저항성 육종에 유용하게 활용될 수 있다.

첫 번째 장에서는 토마토의 *I2*, 감자의 *R3*, 고추의 *L* 사이의 비교유전학적 연구를 통해 *I2C-1*과 *R3/7*에 상응하는 BAC 클론을 선발했다. 선발한 BAC을 PCR, 염기서열 분석, 유전자 지도 작성을 통해 분류했고, 그 중 082F03 BAC이 TG36 근처에 위치했다. 082F03 염기서열을 이용하여 추가적인 BAC 클론을 선발했고, 이들로 구성된 224kb의 컨티그(contig)를 작성했다. 컨티그 염기서열로부터 개발한 3개의 분자표지는 L^4 유전자와 약 1.2 cM 거리에 위치했다. L^3 유전자와 완전하게 공분리하는 분자표지를 L^4 분리 집단에서 연관 분석을 수행한 결과 L^3와 L^4가 같은 유전자좌에 위치하는 대립유전자가 아닌 가깝게 연관된 서로 다른 유전자일 가능성이 있음을 확인했다.

두 번째 장에서는 동정된 L^3 유전자의 동질성을 이용한 PCR로 찾은 L^4 후보유전자로부터 개발한 분자표지(L4segF&R)의 연관

분석을 통해 L^4 후보유전자를 검정했다. L4segF&R은 L^4 분리집단에서 완벽하게 공분리하지 않고 약 0.3cM 거리에 위치했다. 육종 재료의 L4segF&R 분자표지 유전형을 분석한 결과 추가적인 재조합체를 발견했다. 재조합체의 정보를 바탕으로 유전적 구조 모델을 고안했고, 토바모바이러스 저항성이 L^4 후보 유전자에 의해서 완전히 설명되지 않다는 것을 확인했다. 마지막으로 L^4 후보 유전자의 LRR 연기서열로부터 L4RP-3F/L4RP-3R HRM 분자표지를 개발했다.

세 번째 장에서는 L^4를 보유한 PI260429의 전체 전사체 염기서열을 이용하여 *L*과 상동성이 높은 전사체의 *de novo* assembly를 수행했다. *De novo* assembly 결과를 검증하기 위하여 2개의 지표(컨티그 수, 최대 컨티그 길이)를 고안했다. k-mer가 59일 때 Velvet 전사체는 2개의 지표에서 모두 가장 높은 점수를 얻어 추후 분석에 사용되었다. 총 96개의 컨티그가 *L*과 상동성을 보였고, 4개의 토바모바이러스 저항성 계통과 4개의 이병성 계통의 전체 전사체 염기서열을 이용하여 디지털 발현 분석을 수행했다. 8개의 contig가 저항성 계통에서만 특이적으로 발현되었는데, 이들 contig는 4개의 전사체인 L^4 후보유전자, PIX-4, CA00g34020_G4_C3, *L*_homologue 98_99로부터 유래한 것으로 추정된다. 이 결과는 저항성에 중요한 역할을 하는 L^4 후보유전자를 제외한, 다른 3개의 전사체 역시 토바모바이러스 저항성에 중요한 역할을 할 가능성이 있음을 시사한다.

본 연구의 이상의 성과는 고추 토바모바이러스 저항성의 분자유전학적인 이해를 돕고, 더 나아가 토바모바이러스 저항성 분자

육종에 기여할 수 있을 것으로 기대한다.

주요어: 고추(*Capsicum*), 토바모바이러스, 병 저항성, 분자표지, *de novo* assembly