GENETIC MAPPING OF RESISTANCE GENE ANALOGS IN AN INTERSPECIFIC POPULATION OF TOMATO SEGREGATING FOR EARLY BLIGHT RESISTANCE

A Thesis in

Genetics

By

David Orlando Niño-Liu

© 2004 David Orlando Niño-Liu

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

May 2004
The Thesis of David O. Niño-Liu was reviewed and approved* by the following

Majid R. Foolad
Professor of Plant Genetics
Thesis Advisor and Chair of Committee

John E. Carlson
Associate Professor of Molecular Genetics

Richard Craig
J. Franklin Styer Professor of Horticultural Botany

Mark J. Guiltinan
Professor of Plant Molecular Biology

Seogchan Kang
Associate Professor of Plant Pathology

Richard W. Ordway
Associate Professor of Biology
Chair of the Intercollege Graduate Degree Program in Genetics

*Signatures are on file in the Graduate School
Abstract

Limited sources of genetic variation exist within the cultivated tomato, *Lycopersicon esculentum* Mill., for resistance to one of its most destructive foliar diseases, early blight (EB), which is caused by the imperfect fungus *Alternaria solani* Sorauer. Current breeding strategies are trying to identify specific chromosomal regions in wild tomato relatives carrying genes that significantly contribute to an enhanced resistance response. The introgression of quantitative trait loci (QTLs) for EB resistance into the cultivated varieties is greatly facilitated by innovative molecular approaches such as marker-assisted selection (MAS).

A recent trend in crop breeding is the utilization of gene-markers with assigned functions or putative roles in the expression of the trait under evaluation. Plant resistance genes (*R*-genes) are some of the most attractive gene-markers because of their direct involvement in vertical resistance and their presumed contribution to quantitative and non-specific resistance. In the past few years the structural similarities between *R*-gene proteins have been efficiently exploited in the identification, mapping and cloning of new *R*-gene analog (RGA) sequences and specificities.

A purpose of this study was to develop PCR-based markers by using the conserved domains of proteins encoded by *R*-genes as template for primers capable of detecting RGAs in tomato. Ninety-nine polymorphic marker loci were detected after the separation of the total amplification products of these primers by high-resolution polyacrylamide gel
electrophoresis. These markers were added to a genetic map of tomato which was previously developed in our laboratory based on a BC₁ population of a cross between a L. esculentum (EB susceptible) and a L. hirsutum (EB resistant) lines. The mapped markers were initially considered as RGAs because of their observed resemblance to R-genes’ genomic organization and their coincidence with the locations of QTLs for EB resistance in the genetic map. A second purpose of this study was to determine whether the markers mapped in the tomato BC₁ population corresponded to regions of R-genes or R-gene analogs. For this purpose, all markers were isolated from dried gels, cloned, sequenced and compared to the thousands of nucleotide and protein sequences stored in the GeneBank. Blast homology searches revealed that, despite their genomic organization and coincidence with R-genes and QTLs, very few of the markers were genuine RGAs.

These results indicate that, contrary to what has been continuously assumed and reported by other investigators, most markers produced by R-gene-derived primers are not in fact R-gene analogs. However, the R-gene primer approach appears as a valuable tool for the development of new molecular markers and saturation of targeted chromosomal regions such as QTLs for disease resistance. Furthermore, while the nature of clustering of these markers remains unclear, the R-gene primer approach may provide useful markers associated with QTLs for use in MAS.
Table of Contents

LIST OF TABLES ................................................................. viii

LIST OF FIGURES ............................................................. x

ACKNOWLEDGEMENTS ......................................................... xi

CHAPTER 1
Introduction ................................................................. 1
Early Blight of Tomato ..................................................... 2
Breeding for Early Blight resistance ................................. 5
Molecular markers, tools for genetic linkage analyses and marker-assisted selection ................................................................. 7
Molecular markers targeting gene function ....................... 8
R-genes and R-gene analogs as DNA markers ................ 9
The thesis research .......................................................... 10
Objectives ....................................................................... 12
References ....................................................................... 13

CHAPTER 2. Genetic Mapping of Resistance Gene Analogs and their co-localizations with Resistance Genes and QTLs in Tomato
Abstract ........................................................................ 19
Introduction ..................................................................... 20
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant $R$-genes, characteristics, classes, conserved peptide domains</td>
<td>20</td>
</tr>
<tr>
<td>$R$-genes, sequence homologies</td>
<td>24</td>
</tr>
<tr>
<td>Use of degenerate primers for the discovery of $R$-gene candidates</td>
<td>25</td>
</tr>
<tr>
<td>Use of PAGE and the search for polymorphism</td>
<td>27</td>
</tr>
<tr>
<td>Material and Methods</td>
<td>28</td>
</tr>
<tr>
<td>Plant material</td>
<td>28</td>
</tr>
<tr>
<td>RFLP analysis, map construction and QTL identification</td>
<td>29</td>
</tr>
<tr>
<td>RGA analysis</td>
<td>29</td>
</tr>
<tr>
<td>Selection of primers</td>
<td>29</td>
</tr>
<tr>
<td>PCR amplification</td>
<td>29</td>
</tr>
<tr>
<td>Electrophoresis and silver staining</td>
<td>30</td>
</tr>
<tr>
<td>Size determination and cloning of RGA fragments</td>
<td>31</td>
</tr>
<tr>
<td>Cloning and sequencing of PAGE monomorphic markers</td>
<td>31</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>31</td>
</tr>
<tr>
<td>Map construction</td>
<td>31</td>
</tr>
<tr>
<td>QTL identification</td>
<td>32</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>32</td>
</tr>
<tr>
<td>Genetic linkage map and genome distribution in the BC$_1$ population</td>
<td>32</td>
</tr>
<tr>
<td>Identification of RGAs</td>
<td>34</td>
</tr>
<tr>
<td>PAGE-polymorphic RGAs</td>
<td>34</td>
</tr>
<tr>
<td>PAGE-monomorphic RGAs</td>
<td>35</td>
</tr>
<tr>
<td>Mapping of RGAs</td>
<td>36</td>
</tr>
<tr>
<td>Identification of QTLs for EB resistance</td>
<td>38</td>
</tr>
<tr>
<td>Relationship between genomic locations of RGAs and QTLs</td>
<td>38</td>
</tr>
<tr>
<td>Use of RGA markers for MAS</td>
<td>42</td>
</tr>
<tr>
<td>References</td>
<td>43</td>
</tr>
</tbody>
</table>

**CHAPTER 3.** Sequence analysis of $R$-gene-derived markers

Abstract                                                                       | 72   |
Introduction                                                                   | 73   |
Material and Methods ................................................................. 75
  Isolation, cloning and sequencing of RGA fragments ..................... 75
  Sequence analysis and homology searches ................................. 76
  Sequence alignment and phylogenetic analysis ............................ 76
Results and discussion .............................................................. 77
  Cloning and sequencing of RGA markers ................................... 77
  RGA sequence analysis .......................................................... 78
    PAGE-polymorphic markers ................................................. 79
    PAGE-monomorphic markers .............................................. 81
  Associations between PAGE-monomorphic and
    PAGE-polymorphic markers ................................................ 86
  Evaluation of the degenerate primer approach ............................ 89
References .................................................................................. 91

CHAPTER 4.
Conclusions ................................................................................ 118
# List of Tables

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Early Blight (EB) disease severity (final percent defoliation, ±SE) and the area under the disease progress curve (AUDPC, ± SE) for the parental lines and BC₁ and BC₁S₁ progeny of a cross between <em>L. esculentum</em> breeding line NC84173 (EB susceptible) and <em>L. hirsutum</em> accession PI126445 (EB resistant)</td>
<td>59</td>
</tr>
<tr>
<td>2 Oligonucleotide primers designed from conserved peptide domains encoded by various R genes</td>
<td>60</td>
</tr>
<tr>
<td>3 QTLs detected for Early Blight (EB) resistance based on simple interval mapping (SIM) in BC₁ and BC₁S₁ populations of an interspecific cross between <em>L. esculentum</em> (NC84173; EB susceptible) and <em>L. hirsutum</em> (PI126445; EB resistant)</td>
<td>61</td>
</tr>
<tr>
<td>4 One-thousand genotype-shuffle permutation test performed for RGA markers associated with QTLs for EB resistance based on the BC₁-1998 disease evaluations</td>
<td>62</td>
</tr>
<tr>
<td>5 One-thousand genotype-shuffle permutation test performed for RGA markers associated with QTLs for EB resistance based on the BC₁S₁-1999 disease evaluations</td>
<td>63</td>
</tr>
<tr>
<td>6 One-thousand genotype-shuffle permutation test performed for RGA markers associated with QTLs for EB resistance based on the BC₁S₁-2000 disease evaluations</td>
<td>64</td>
</tr>
</tbody>
</table>
7 Final count of amplification products obtained with degenerate RGA-primers
8 Twenty-nine PAGE-polymorphic markers genetically mapped in the tomato BC1 population map
9 Seventy-five PAGE-monomorphic markers isolated from dried polyacrylamide gels
10 PAGE-polymorphic markers and their most significant homologies to non-R-genes
11 PAGE-monomorphic markers with their most significant homologies to R-genes, RGAs and other plant genes
# List of Figures

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65</td>
</tr>
<tr>
<td>A linkage map of tomato chromosomes based on a BC₁ population of a cross between <em>Lycopersicon esculentum</em> (breeding line NC84173; pistillate and recurrent parent) and <em>L. hirsutum</em> (accession PI126445) and 241 DNA markers (142 RFLPs and 99 RGAs)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>68</td>
</tr>
<tr>
<td>A linkage map of tomato displaying chromosomal locations of resistance gene analogs (RGAs) and QTLs for Early Blight (EB) resistance</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>70</td>
</tr>
<tr>
<td>Permutation analysis for the BC₁S₁-1999 disease evaluation</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>116</td>
</tr>
<tr>
<td>Jukes-Cantor UPGMA trees displaying phylogenetic relationships between PAGE-polymorphic and PAGE-monomorphic markers amplified with <em>R</em>-genes degenerate primers</td>
<td></td>
</tr>
</tbody>
</table>
I would like to express my sincere appreciation and gratitude to my academic advisor Dr. Majid Foolad, for the priceless time and effort devoted to the direction of this research project. I extend my gratitude to Drs. John Carlson, Mark Guiltinan and Seogchan Kang, who served as members of my doctoral committee and provided with significant assistance during the course of my investigation. I would especially like to thank Dr. Richard Craig for his precious professional and personal counsel and for his invaluable friendship.

I also would like to thank former and present members of Dr. Foolad’s *Tomato Genetics and Molecular Breeding Program*, Drs. Liping Zhang, Aftab Khan and Arun Sharma, and Mr. Guo-Yang Lin, for their important contribution to this investigation; to all greenhouse, lab and field personnel, and to the graduate and undergraduate students involved in the several years of experiments and evaluations.
Chapter 1

Introduction

The tomato (*Lycopersicon esculentum* Mill.) is one of the world’s most important vegetables and a well-established plant model system in genetics and physiology. It is a short-lived perennial plant with perfect flowers and of easy growing cultures and breeding due to its diploidy and self-fertility.

The tomatoes of the genus *Lycopersicon* are all native to South America. They originated in South-Central Perú, where eight species of the genus still grow wild in the Andes. The current range of wild tomato relatives extends from the north of Chile and west of Bolivia to Ecuador, reaching inland from the Pacific 100-200 miles, and the Galápagos Islands (Rick 1976; Gould 1983).

The tomato was domesticated and extensively grown by the Inca, Mayan and Aztec empires. Its original names, variants of the word “xitomatl” of the Meso-American languages, were latinized by the conquerors to “tomate”. It was brought to Europe by the Spaniards in the early 16th century as a botanical curiosity to become part of scientific collections. For nearly four hundred years the tomato was grown only as an ornamental and known by different names. For example, the Latin name of the cultivated tomato
derived from a German toxic variety with hallucinogenic properties, and it translates as “edible juicy wolf peach”. In the mid-nineteenth century it was discovered that tomato fruits, after ripening, were as good a food as most other fruits or vegetables. Soon after this discovery, it gradually became very popular all over the world. The tomato is now grown virtually anywhere and it holds second place among the vegetables produced by humans, only exceeded by the potato (Gould 1983; Rick et al. 1990; Hancock 1992).

The initial domestication by the Native Americans and the subsequent breeding efforts and exchange of germplasm between Europeans and North Americans represented a profound and rapid event in the evolution of the tomato, altering the distribution of the species and enabling dramatic changes in its anatomy to fit human needs and demand. The most clear example is the selection towards attractive fruit sizes, shapes and vivid colors (Rick et al. 1990). It has been hypothesized that the majority of the anatomical changes are attributed to a few loci selected over years of improvement by our ancestors, which caused alterations in the overall genomic diversity of the genus. At least three population bottlenecks in the history of modern tomato are currently being reflected in a lack of genetic diversity among the cultivated varieties. This genetic uniformity places the cultivated tomato in a very vulnerable position against the threat of biotic and abiotic stresses that may affect its production and food security (Rick 1976; Grandillo et al. 1999; Nesbitt and Tanksley 2002).

**Early Blight of Tomato**

It has been proposed that an average of one hundred diseases can affect each crop plant (Agrios 1988). In tomato, there are nearly 200 characterized parasitic and non-parasitic diseases with diverse causes and etiologies. Parasitic diseases in tomato can be caused by insects, opportunistic phanerogams, and pathogenic microbes including viruses, viroids, bacteria, nematodes, mycoplasms and fungi (Agrios 1988; Jones 1991).

The imperfect fungus *Alternaria solani* Sorauer is the causal agent of Early Blight (EB) of tomato and potato (*Solanum tuberosum* L.), and is one of the best known and
economically important members of the genus *Alternaria* Nees (Rotem 1994). EB of tomato is distributed worldwide; however, high epidemic incidence occurs in many regions of North, Central and South America, South Asia, and Africa, which have favorable climatic conditions for the disease. In the United States, EB is the most important defoliation disease of tomatoes, causing damage from New England and the Middle Atlantic to the South East and central states (Jones 1991). Reported annual losses in the Commonwealth of Pennsylvania include yield reduction of 20%–30%, fruit size reduction of 10%, and a considerable increase in anthracnose fruit rot which is associated with premature defoliation (Hoffmann 1996).

Two distinct types of disease phases can be distinguished in tomato EB: the collar rot and leaf blight. The collar rot phase occurs when infected seedlings with stem lesions are transferred to the field as transplants, such lesions will eventually encircle the stem forming "collars" that inhibit vascular processes. The leaf blight phase, commonly referred to as early blight, manifests itself as coalescing concentric circular necrotic lesions on the foliar surface of generally lower leaves, progressing upward as the plant matures (Jones 1991; Rotem 1994). In heavily infected leaves the pathogen resides in the center of the necrotic lesions surrounded by a non-invaded chlorotic halo and a green pre-halo of healthy appearance. Infected leaves eventually fall from the plant, which reduces the photosynthetic rate, increases the respiration of healthy tissue, and thus disrupts the equilibrium of supply and demand of nutrients. Defoliated plants often result in sun-scalded fruits. Under field conditions, EB infection of tomato plants usually occurs after flowering; although after artificial inoculation in an optimal controlled environment, symptoms are observed at all stages of plant growth. Necrotic lesions may also appear in lower floral parts and fruits, both green and ripe. (Jones 1991; Strandberg 1992; Rotem 1994).

Temperature and humidity affect the production of conidiophores, sporulation and germination, both *in vivo* and *in vitro*. The infection is favored by a relative humidity above 80% (conditioned by frequent periods of rain and night and morning dew) and moderate temperatures, with the optimum being 27.5º C. Under favorable conditions,
spores of *Alternaria* are capable of germinating on leaves of host (resistant or susceptible to the disease) and nonhost plants (Rotem 1994). After germination, the germ tubes emerge from spores and spread on the leaf surface. Infection occurs by direct hyphal penetration of the foliar surface through stomata, dead epidermal cells, or wounds caused by insects or by sand particles carried by strong winds (Jones 1991; Strandberg 1992).

Enzymatic action and the production of toxins after spore germination play very important roles in the pathogen entry and colonization of the susceptible plant. Cellulases break down plant cell walls, and their action is typically complemented by enzymes such as pectin methyl galacturonase, pectin lyase and pectin methylesterase. In some cases, the simultaneous action of pathogen toxins, deleterious low-molecular weight compounds, suppress defense mechanisms at any or all stages of infection by causing abrupt histological and physiological changes in the host. Spore germination is also enhanced by materials leached from infected tissue, composed mainly of organic and amino acids, vitamins and carbohydrates (Rotem 1994; Prell and Day 2001).

*A. solani* overwinters and survives as conidia or mycelium on buried host debris and tubers, particularly in fields with poor cultural practices such as continuous cropping to potato or tomato. Thick-walled chlamydospores resistant to low temperatures buried on organic debris may also serve as the primary inoculum for the following growing season. Therefore, control methods for EB incorporate lengthy non-host crop rotation, debris removal, sanitation and the utilization of aseptic seeds and transplants. Additional methods of disease control may include the increase in nitrogen and phosphorus fertilizers and the reduction of wetness on foliar surface through soil-directed irrigation systems. Nevertheless, widespread application of preventive fungicides soon after seedlings emerge or are transplanted is the most frequent control measure, although it still has limited success in humid or rainy cultivated areas (Agrios 1988; Rotem 1994; Narayanasamy 2002).
Breeding for Early Blight resistance

The development and use of cultivars with genetic resistance to pathogens is the most desirable and attractive alternative for disease management. Control by resistance implies the search and utilization of the genetically-conditioned differences in susceptibility to virulent pathogens from current cultivars and related exotic species. Sources of genetic resistance to EB have never been known to exist within the cultivated species of tomato, mainly because of the narrow genetic diversity and high cultivar uniformity, products of the intensive selection and breeding (Rick 1976). Due to the continuous introduction and heavy use of chemical fungicides, the search for accessions resistant to EB had not been a priority until the mid-1950s when, for the first time, germplasm collections were screened in the United States and Canada. Accessions resistant to EB were reported only within related wild species of tomato in the early 1970s. Since then, several breeding and commercial lines with satisfactory levels of resistance have been developed, mainly from crosses between the cultivated tomato and the resistant accessions of *L. pimpinellifolium* (Jusl.) P. Mill., *L. hirsutum* Humb. & Bonpl., and *L. peruvianum* (L.) P. Mill. (Barksdale and Stoner 1973; Maiero et al. 1991; Gardner and Shoemaker 1999; Foolad et al. 2000).

Resistance to EB in tomato is a complex quantitative trait, controlled by the additive effects and epistatic interactions of many genes and highly influenced by the environment (Nash and Gardner 1988). EB resistance does not follow the gene-for-gene model of vertical qualitative resistance proposed by Flor (1971), where the specificity of a resistant host against its avirulent pathogen is determined by the interaction of their respective dominant *R* and *avr* genes. Since the host recognition of the fungus occurs in a non-specific manner, early events of the defense response in resistant plants depend on constitutively expressed and rapidly accumulating pathogenesis-related (PR) proteins in an elicitor-releasing mechanism. PR-proteins consist mostly of catalytic enzymes that accumulate in intracellular compartments and leak from the original dead epidermal cells during pathogen attack. Once released, PR-proteins are capable of creating an anti-fungal environment, for example by hydrolyzing the pathogen’s cell walls. The timely recognition of non-specific elicitor molecules from the total fungal cell debris activates a
chain of defense mechanisms that form the overall plant resistance phenotype (Linthorst 1991; Lawrence et al. 1996).

EB resistance is also influenced by the physiological changes that accompany the plant during aging and developmental stages. It has been reported for example that resistance to *A. solani* may be different for the collar rot phase that affects seedlings versus the foliar phase that affects adult plants (Maiero et al. 1991; Rotem 1994). Studies of sporulation in *Alternaria* indicated that photosynthesis and high contents of sugars in younger uninfected leaves play an inhibitory role in the production of spores at early stages of pathogenesis, thus the preference of the pathogen for lower senescence leaves and the “apparent” resistance in late-maturing or indeterminate plants that exhibit a low yield-to-foliage ratio (Maiero et al. 1990; Rotem 1994).

Current resistance breeding and integrated pest management programs are designed to significantly decrease fungicide spray applications to economically benefit growers and consumers and to reduce the negative ecological impact on the environment. Much progress has been achieved in the identification of EB resistant accessions and the development of breeding lines. However, enhanced levels of disease resistance transferred from wild tomato species are very often associated with undesirable agronomic traits such as late-maturity and low fruit-yielding ability (Maiero et al. 1990; Foolad et al. 2000). This results from loci controlling such unwanted traits remaining in the genotype of developed lines that have been selected phenotypically during plant evaluations due to their resistant appearance. Late-maturity and low-yielding ability are ill-suited traits to tomato production, and become particularly important in geographic areas where tomato production is seasonal and subjected to short-season field growing conditions. Moreover, in some cases even the earliest maturing breeding lines developed for the market may still have a lower fruit yield or possess a level of resistance not sufficient under most epidemic field conditions (Maiero et al. 1990; Rotem 1994).

To facilitate breeding efforts, research has been initiated to better understand the genetic basis of EB resistance in tomato. Such efforts are directed to the identification of
genes or quantitative trait loci (QTL) conditioning EB resistance in exotic tomato germplasm, estimation of the contribution of each QTL to the overall expression of the resistance, and their transfer to the cultivated backgrounds (Christ and Foolad 1996; Foolad et al. 2002).

**Molecular markers, tools for genetic linkage analyses and marker-assisted selection**

The development of independently inherited molecular markers represents the most powerful method for the identification of QTLs affecting important agricultural traits. Molecular markers are phenotypically neutral physical entities in the form of DNA or protein polymorphisms that exhibit simple Mendelian inheritance. The molecular basis of DNA marker polymorphisms is formed by genetic point mutations, insertions or deletions that differentiate the individual genomes of all members of a species. Perhaps the most common use of molecular markers is the estimation of their chromosomal locations with respect to their physical associations with other markers or genes. Such associations, the sequential order of markers and genes in chromosomes and the relative distances that separate them, provide the scaffold for the construction of genetic linkage maps (Lander and Botstein 1989; Paterson et al. 1991).

The sophisticated DNA marker technology applied to quantitative EB resistance studies is of much potential use for breeders. For example, QTLs significantly influencing the overall expression of the resistance are identified by simultaneous statistical analyses of phenotypic data (EB disease scores) in combination with the segregation of markers in a population of individuals. Further, with markers already positioned in moderate- or high-density genetic maps, it is also possible to detect inter- and intra-allelic interactions and to ascertain the influence of different environments on the phenotype (Young 1996; Foolad et al. 2002; Zhang et al. 2003).

Molecular markers’ application in breeding is also realized by marker-assisted selection (MAS), that is, the use of markers to select for linked neighboring genes or QTLs, which are difficult to manipulate by other means. Polymorphic genetic markers
associated with EB resistance represent potential genetic variation. By analyzing the segregation and combinations of markers, one can readily estimate relatedness between different plants and consequently predict which matings are capable of producing individuals with superior resistance. Similarly, plants with particular combinations or higher number of alleles controlling unwanted traits can also be detected and selected against. A clear example is the elimination of alleles for late-maturity, self-incompatibility and indeterminate growth habit in populations derived from interspecific crosses between *L. hirsutum*, utilized as the source of EB resistance, and the cultivated tomato (Bernacchi et al. 1998; Foolad et al. 2000). Because DNA markers are phenotypically neutral, one can predict the presence or absence of an interesting gene or loci even before the phenotype is actually determined. In this regard, MAS saves time by conducting the screenings at early stages of plant development, and provides a safer experimental environment by reducing dependence on unreliable artificial pathogen inoculum for disease screenings.

**Molecular markers targeting gene function**

Tomato has one of the best-defined genetic maps in the plant kingdom, probably ranking third after *Arabidopsis* and rice, with over 1100 characterized “anchor” markers along its 12 chromosomes. The “anchor” markers compose the framework of the tomato genome and are formed almost entirely by complementary DNA (cDNA) and, to a greater extent, by anonymous genomic sequences, capable of displaying restriction fragment length polymorphisms (RFLP). Interestingly however, a few dozen morphological traits continuously mapped since the early 1900s and a few hundred isozymes, now considered archaic markers, are still being utilized as anchors (Bernatsky and Tanksley 1986; Tanksley et al. 1992).

One great advantage of DNA markers is their apparent unlimited availability. In tomato, for example, a collection of over 140,000 expressed sequence tags (ESTs) is publicly available in addition to the current molecular and morphological anchors. The total number of ESTs in that collection represents approximately 28,000 unique
consensus sequences, with a prediction of a minimum number of 35,000 single-copy genes existing in the tomato genome (Van der Hoeven et al. 2002; TIGR 2003).

However, as opposed to model plants like tomato or Arabidopsis, many less-characterized plant species do not possess well-developed collections of ESTs, and thus the lack of functional markers is still a limiting factor in genetics and mapping. Nonetheless, different innovative approaches have been adopted to increase the number of markers while maintaining and exploiting their functional aspects.

**R-genes and R-gene analogs as DNA markers**

In the mid 1990s, a handful of plant resistance genes (R-genes) from maize, tomato, Arabidopsis, tobacco, flax and rice were cloned and sequenced. Surprisingly, most of the predicted sequences and structures of the encoded proteins shared a striking similarity, even though they governed resistance to different viral, bacterial and fungal pathogens. The similarities among the predicted R-proteins provided the knowledge for the identification of new gene families and their analogs in several other plant species (Hammond-Kosack and Jones 1997; Hulbert et al. 2001).

Initially, the first cloned R-genomes were utilized as RFLP probes for DNA cross-hybridization, and with some success, multiple copies of R-gene analogs (RGAs) were detected in all species evaluated. However, the DNA homology across distant species was rather low and proved to be insufficient for gene isolation (Bent 1996a; Hulbert et al. 2001). Nevertheless, a more powerful approach arrived with the utilization of degenerate oligonucleotides capable of annealing to the flanking coding regions of R-genes and prime DNA amplification by the polymerase chain reaction (PCR) (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996).

Literature on R-gene degenerate primers is extensive. They have been mainly used for the production of RGAs in the form of conventional RFLP probes for genetic mapping
and for the isolation of new \( R \)-gene specificities. Because the degenerate primers were originated from \( R \)-protein sequences, it is possible to assign putative biological function to their PCR-amplified markers and to use them for \( R \)-loci tagging.

The manipulation of informative functional markers follows a recent trend in genetic mapping, especially in map-based cloning, that greatly modifies earlier marker technology by departing from its original reliability on anonymous or arbitrary sequences. Because each annotated RGA potentially represents a functional allele of a “candidate” \( R \)-gene, a mapped RGA statistically associated with an increase in resistance implies that such candidate may be directly affecting the trait or at least genetically linked to a locus that is (Ballabio 1993; Rothschild and Soller 1997; Pflieger et al. 2001). In the past few years, RGA markers have been used to identify QTLs for disease resistance in more than a dozen of crop plants. Furthermore, they also have explained the effects and interactions of QTLs by correlating putative gene role in resistance or defense with their map positions, and in reference to other potentially involved candidate genes and loci (Chen et al. 1998; Faris et al. 1999; Pflieger et al. 2001; Quint et al. 2001; Wang et al. 2001).

However, besides being used as RFLP probes for genetic mapping, very few alternative uses for the RGA primers have been developed and reported. Among these alternatives are selective genotyping, DNA fingerprinting and phylogenetics, as well as direct genetic mapping as PCR-based markers (Chen et al. 1998; Trognitz et al. 2002; Zhang et al. 2002; Ramalingam et al. 2003).

The thesis research

This research involved the application of an alternative experimental method to the conventional use of degenerate \( R \)-gene-derived primers for the production of new DNA markers. The primary goal was to determine whether the chromosomal locations of RGA-derived markers coincide with the locations of QTLs identified in our laboratory for tomato EB resistance, and to address their relevance and applicability in MAS and
crop improvement programs.

Briefly, the experimental approach began with the PCR-amplification of genomic DNA with degenerate \( R \)-gene primers, followed by the separation of the total heterogeneous PCR products on a high-resolution gel matrix. The segregation patterns of polymorphic RGA markers displayed on gels were directly scored as dominant loci for genetic mapping. This strategy omits the traditional and time-consuming steps of ‘cloning–sequencing–screening’, which most RGA markers, originated by the same primers, usually undergo before they are selected for mapping. Very few reports exist on the direct utilization of \( R \)-gene primers for genetic mapping, since most of the attention has been directed to RGAs selected from the total PCR product after their homologies to \( R \)-genes are confirmed by sequencing (Chen et al. 1998; Zhang et al. 2002; Ramalingam et al. 2003). In the latter approach, the preliminary sequencing and screening normally loses dozens of unselected amplified fragments because no significant homology to any gene in the genebanks is found, and therefore their position in genetic linkage maps is never resolved.

The molecular approach followed in this thesis research intended to produce informative markers with a possible \( R \)-gene or \( R \)-gene-like function. These PCR markers represent a more useful substitute to, for example, randomly amplified polymorphic DNAs (RAPDs), because they are more likely to contain sequences encoding protein domains similar to those in \( R \)-proteins. Nevertheless, a final sequencing step was integrated into the experimental procedure to reveal whether mapped markers, which were coinciding with EB resistance QTLs, were genuine \( R \)-genes.

Research conducted in maize, rice, wheat, tomato and potato have all resulted in the non-random genomic distribution of RGAs often clustering around \( R \)-genes and/or within QTLs for resistance to different pathogens (Faris et al. 1999; Li et al. 1999; Wang et al. 2001; Ramalingam et al. 2003). This evidence supports the hypothesis that \( R \)-genes can explain a significant proportion of the quantitative variation for resistance, even in populations evaluated with non-specific virulent pathogens (Freymark et al. 1993;
Leonards-Schippers et al. 1994; Wang et al. 1994; Michelmore 1996). In Chapter 3, the sequence identities of the mapped RGAs are presented and compared with the sequences available in public gene databanks. Further, in the future, for markers that mapped to positions of QTLs accounting for significant variation in EB resistance, we intend to determine whether they are \( R \)-genes, RGAs, other known genes, or simply non-informative amplified DNAs.

### Objectives

This thesis research was initiated with the following objectives:

1. To identify DNA markers using degenerate primers based on the conserved protein domains encoded by various \( R \)-genes.

2. To genetically map the DNA markers in an interspecific population of tomato segregating for EB resistance.

3. To determine whether the chromosomal locations of RGAs coincide with the location of any known \( R \)-genes or EB-resistance QTLs identified in our laboratory.

4. To clone and sequence the \( R \)-gene-derived markers and resolve their nucleotide and translated-peptide sequence homology to \( R \)-genes and \( R \)-gene products.

5. To assess the utility of \( R \)-gene-derived markers as DNA markers for QTL discovery and marker-assisted selection.
References


Kanazin, V., Marek, L.R., and Shoemaker, R.C. 1996. Resistance gene analogs are


Chapter 2

Genetic Mapping of Resistance Gene Analogs and their Co-localizations with R-Genes and QTLs in Tomato

Abstract

Sequence conservation in functional protein domains encoded by different plant resistance genes (R-genes) has facilitated the use of polymerase chain reaction (PCR)-based approaches in isolating and cloning novel R-genes. The purpose of this research was to identify and genetically map resistance gene analogs (RGAs) in tomato by PCR amplification of genomic DNA using degenerate primers for the conserved nucleotide binding site (NBS), leucine-rich repeat (LRR) and serine/threonine kinase (STK) domains of several known R-genes. Previously, a molecular linkage map of tomato was constructed in our laboratory based on a BC1 population (N = 145) of a cross between Lycopersicon esculentum Mill. breeding line NC84173 (maternal and recurrent parent) and L. hirsutum Humb. and Bonpl. accession PI126445. The framework map included 142 restriction fragment length polymorphism (RFLP) markers mapped onto the 12 tomato chromosomes. The RFLP markers, which were based on random genomic and
cDNA clones, were used as anchor markers. In the present study, 99 RGAs were added to the framework map, of which 29 were PCR-based and 70 were hybridization-based (i.e. RFLP type) markers. The RGA markers were mapped to 11 of the 12 tomato chromosomes, often associating in clusters of lengths no longer than 28 cM. The BC$_1$ population (consisting of 145 individuals) and its self-progeny (145 BC$_1$S$_1$ families) were previously evaluated for Early Blight (EB) disease symptoms in 1998 (BC$_1$), and 1999 and 2000 (BC$_1$S$_1$) and over 10 QTLs were identified for EB resistance. These QTLs were mapped to 9 of the 12 tomato chromosomes. In the present study, the chromosomal locations of RGAs were compared with those of QTLs for EB resistance as well as locations of several tomato $R$-genes. The results indicated that chromosomal locations of 27 RGA markers coincided with locations of 6 EB-resistance QTLs and several known $R$-genes. These results suggested that RGA markers might be involved with genes controlling the expression of resistance to different diseases. Furthermore, RGAs linked to significant QTLs represent a powerful source of genetic markers, which can be used in marker-assisted selection and breeding projects.

Introduction

Plant $R$-genes, characteristics, classes, conserved peptide domains

In many plant-pathogen interactions, incompatibility (i.e., resistance to disease) is controlled by the interaction of plant-encoded resistance genes ($R$-genes) and their corresponding pathogen-encoded avirulence genes ($Avr$-genes) of matched specificity. The “gene-for-gene” model predicts that the pathogen $Avr$-gene products act as elicitors and are directly or indirectly recognized by specific receptors in plant $R$-genes products. Such specific recognition triggers a signal transduction pathway that induces a strong resistance response and activates an ample array of host defense mechanisms (Flor 1971; Gabriel and Rolfe 1990; Lamb 1996).
The R-genes constitutively expressed in a plant create a refined surveillance and recognition system that detects the presence of a pathogen by interacting with its elicitor molecules during infection. The distinguishing hallmark of most R-avr gene interactions is the activation of the hypersensitive response (HR), in which cells surrounding the sites of attempted pathogen ingress undergo immediate death to prevent further colonization. HR can be phenotypically diverse, but it usually appears as very small necrotic flecks within hours of contact with the pathogen (Dangl et al. 1996). Other features of the R-gene-mediated response include the induced synthesis of antimicrobial metabolites (often referred to as phytoalexins), pathogenesis-related (PR) proteins and enzymes harmful to invading pathogens, reinforcement of the plant cell wall, generation of toxic amounts of reactive oxygen species, and the initiation of a distinct signal transduction pathway known as a systemic acquired resistance (SAR) (Bent 1996b; Ryals et al. 1996).

During the past decade, over 40 R-genes have been cloned in different plant species for resistance to a wide spectrum of pathogens, including bacteria, viruses and fungi. The predicted protein products (R-proteins) of most of these R-genes, however, share common sequence motifs typically involved in signal transduction and protein-protein interactions (Michelmore 1996; Hammond-Kosack and Jones 1997; Martin et al. 2003). Based on their deduced amino acid sequences and the combination of R-protein structural domains, R-genes have been classified into six major classes:

Class 1 has only one member, the tomato Pto gene, whose product carries a serine/threonine kinase (STK) catalytic region. Pto governs resistance to the bacterial speck pathogen Pseudomonas syringae pv. Tomato, expressing the corresponding avirulence gene avrPto. After the intracellular recognition of the avrPto product, Pto is activated and involved in signal transduction via self- and specific phosphorylation of additional protein substrates that leads to disease resistance responses (Martin et al. 1993a; Bogdanove 2002).

The second and third classes include the majority of R-genes whose encoded proteins carry a region of leucine-rich repeats (LRRs) and a putative nucleotide binding site
(NBS). These classes, however, differ in the N-termini of their proteins. Class 2 has a leucine-zipper (LZ) or other coiled-coil sequence (CC) whereas the class 3 has a region referred to as TIR because of its similarity to the *Drosophila Toll* and the mammalian Interleukin 1 receptor (Bent 1996b; Hammond-Kosack and Jones 1997; Hulbert et al. 2001).

LRR domains are hypothesized to be the most important determinants in the specificity of pathogen recognition in classes 2 and 3 (although N-terminal CC or TIR domains might also be involved to a much lesser extent). This is confirmed by the high levels of allelic polymorphism found among the encoding genes, and the presence of LRR’s solvent-exposed residues in many other plant proteins, not related to resistance or defense responses but required for protein-protein communication. Additional experiments such as domain swapping, yeast-two-hybrid and in-vitro binding analysis have also supported this hypothesis (Ellis and Jones 1998; Hulbert et al. 2001).

NBS regions similar to those of *R*-proteins are also present in several other protein families that affect their function through nucleotide binding or hydrolysis. Based on structural predictions and protein alignment with NBS of cell death-genes in *Caenorhabditis elegans* and humans, plant NBS-containing *R*-products are suggested to control HR cell death after they are activated via LRR-recognition of pathogen *Avr* genes (Biezen et al. 2002).

The *R*-genes belonging to classes 2 and 3 include, among others, *RPS2* and *RPM1* conferring resistance to the bacterial blight pathogen *Pseudomonas syringae* and *RPP5* conferring resistance to the fungus *Peronospora parasitica* in *Arabidopsis* (Bent et al. 1994; Mindrinos et al. 1994; Grant et al. 1995; Parker et al. 1997), the tobacco *N* gene conferring tobacco mosaic virus (TMV) resistance (Whitham et al. 1994), the *L6* and *M* genes conferring flax rust resistance caused by the fungus *Melampsora lini* (Lawrence et al. 1995; Anderson et al. 1997), the *I2C-1* locus conferring resistance to race 2 of the tomato vascular wilt fungus *Fusarium oxysporum f sp lycopersici* (Ori et al. 1997) and the *Xa1* gene conferring resistance to the bacterial blight pathogen *Xanthomonas oryzae*
in rice (Yoshimura et al. 1998).

The \(R\)-genes of the first three classes have no physical associations with the cell membrane and are all thought to be localized in the cytoplasm. The intracellular localizations of these \(R\)-proteins have serious implications in understanding the interception of pathogen \textit{avr}-gene products, as well as the mechanism used to deliver such products into the plant cell at the beginning of pathogenesis (Hammond-Kosack and Jones 1997; Hulbert et al. 2001).

Class 4 of \(R\)-proteins is formed by the tomato \textit{Cf}-glycoproteins, which lack a NBS but possess an anchor transmembrane domain and large extracellular LRRs. This group includes the gene families \textit{Cf}-9, \textit{Cf}-2, \textit{Cf}-4 and \textit{Cf}-5 conferring race-specific resistance to the leaf mold pathogen \textit{Cladosporium fulvum} by recognizing the fungus’ secreted \textit{avr}-gene products in plant intercellular spaces (Jones et al. 1994; Dixon et al. 1996).

Class 5 consists of only the rice \textit{Xa21} gene that governs resistance to the bacterial leaf blight pathogen \textit{Xanthomonas oryzae pv. oryzae} (Song et al. 1995). \textit{Xa21} product contains an extracellular LRR, a transmembrane domain and a cytoplasmic serine/threonine kinase. The architecture of the \textit{Xa21} protein suggests that its LRR domain recognizes extracellular pathogen \textit{avr}-gene products whereas its intracellular protein kinase is involved in signal transduction to trigger plant resistance mechanisms by phosphorylation-mediated activation (Hulbert et al. 2001).

The last, class 6, is represented only by the newly discovered \textit{Arabidopsis RRS1-R} gene, which confers an atypical broad-spectrum resistance to the bacterial wilt pathogen \textit{Ralstonia solanacearum} in which HR is not developed. \textit{RRS1-R} is a unique gene that contains a putative nuclear localization signal (NLS) and a WRKY domain (Tryptophan-Arginine-Lysine-Tyrosine), in addition to the standard TIR-NBS-LRR domains. The presence of a putative WRKY, a transcriptional activation domain for defense gene products such as PR-proteins, suggests a possible dual function for \textit{RRS1-R}: the recognition of a pathogen-derived signal and a direct nuclear activation of defense-related
genes. (Deslandes et al. 2002).

Very few $R$-proteins do not fit into any of the six classes described above, mainly due to the lack of obvious NBS-LRR or $Pto$-like structures for protein recognition and signal transduction. The most clear example is the maize $Hm1$, the first cloned $R$-gene, that confers resistance to Cochliobolus carbonum. $Hm1$ encodes a NADPH-dependent reductase that inactivates the potent plant toxin. The detoxification strategy of $Hm1$ does not involve recognition of pathogen $avr$-gene products and thus, events such as HR are not present in this type of genetic resistance (Johal and Briggs 1992).

**$R$-genes sequence homologies**

$R$-genes are found in diverse plant species and are active against the full taxonomic range of plant pests, however, the basic functional LRR, NBS and STK domains contained by their proteins are structurally similar. Such protein similarity suggests a high degree of mechanistic conservation among the common signal transduction pathways that plants use to trigger defense responses initiated after the specific recognition of the pathogen (Staskawicz et al. 1995; Bent 1996b).

In the early 1970s, a specific “elicitor-receptor” recognition model was proposed to account for Flor’s gene-for-gene resistance theory. And although this model was valid in most cases where incompatible interactions occurred, it did not facilitate the isolation of either the $R$-gene or its encoded protein receptor needed for elicitor recognition (Flor 1971; Gabriel and Rolfe 1990). No success in the isolation of the first $R$-genes was achieved until the early 1990s, when technologies for gene cloning based on their phenotype and genetic location became available.

In general, plant $R$-genes have been isolated by three main methods: (1) cloning based on their phenotypes and chromosomal locations, (2) PCR amplification with degenerate primers, and (3) gene prediction from databases of plant genome sequencing initiatives.
The original R-genes were isolated, despite their unknown structure and function, by either positional cloning, e.g. the tomato Pto gene (Martin et al. 1993a), or transposon tagging, e.g. the maize Hm1 gene (Johal and Briggs 1992). These two approaches led to the first wave of cloned R-genes in Arabidopsis, RPS2 and RPM1 (Bent et al. 1994; Grant et al. 1995), tomato, Cf-9, Cf-2 and Prf (Jones et al. 1994; Dixon et al. 1996; Salmeron et al. 1996), tobacco, N (Whitham et al. 1994), flax, L6 (Lawrence et al. 1995), and rice, Xa21 (Song et al. 1995). Surprisingly, the sequences of these new R-genes, after being translated to peptides, showed a remarkable similarity to each other, but no resemblance to the original Pto or Hm1. Based on their deduced peptide sequences and predicted protein structures, the new R-genes were classified as NBS-LRR (classes 2 and 3 described above). The NBS-LRR classification of these R-genes enhanced the detection of additional candidate R-gene sequences that were PCR-amplified using degenerate oligonucleotide primers designed from such conserved domains (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996). Today, the extensive use of degenerate primers in combination with gene mining from databanks of several plant genome sequencing initiatives are the most common methods to uncover novel genomic and cDNA sequences related to the R-genes classes.

Use of degenerate primers for the discovery of R-gene candidates

In the past several years, the use of degenerate primers based on the conserved domains of plant R-proteins has offered the possibility to identify new R-genes and R-gene analogs (RGAs) from many plant species. Amplified products have often been useful as genetic markers for the evaluation of genetic diversity in germplasm, genetic and physical mapping, QTL discovery and analysis, genomic library screening, and positional cloning (Feuillet et al. 1997; Chen et al. 1998; Seah et al. 1998; Spielmeyer et al. 1998; Mago et al. 1999). RGA identification with degenerate primers is a powerful approach to gene discovery. This is particularly important when low or insufficient DNA sequence homology exists across species; in such cases, many strong R-gene candidates may never be detected by using previously cloned R-genes as RFLP probes.
The first use of primers with different degrees of degeneration and lengths was reported in the mid 1990s. Kanazin et al. (1996) and Yu et al. (1996) designed primers based on the NBS domains of the Arabidopsis RPS2, tobacco N, and flax L6 genes and identified up to 11 different classes of RGAs in soybean. These RGAs were localized in the vicinity of previously mapped soybean R-genes conferring resistance to potyviruses, root rot and powdery mildew. Similarly in potato, Leister et al. (1996) deduced primers from the RPS2 and N genes and isolated PCR-products that not only were homologous to known R-genes, but that were also linked without genetic recombination to nematode and late blight resistance loci. Moreover, map positions revealed that these potato RGAs were also related to resistance loci present in the tomato and tobacco genomes. In wheat, Feuillet et al. (1997) used oligonucleotides corresponding to the conserved subdomains of serine/threonine kinases and isolated the wheat Lr10 gene that confers resistance to the leaf rust pathogen Puccinia recondita.

In the following years the number of new R-gene and RGAs sequences significantly increased. The degenerate primer approach provided a rapid and efficient method to identify, map and clone new uncharacterized resistance candidates to diverse pathogens and from virtually any plant species. For example, Collins et al. (1998) used primers based on eight regions of amino acid similarity in and around the NBS-LRR from dicots to identify 11 unique classes of RGAs in maize. Some of the maize RGAs cosegregated perfectly with two rust disease resistance loci, rp1 and rp3, and even detected deletion events in rp1 mutants. Interestingly, Timmerman-Vaughan et al. (2000) used different primer combinations to identify genomic regions in pea that contain NBS-class genes involved in resistance and plant-microbe interactions. Nine RGAs were cloned, characterized and their inferred genomic organization identified loci in the proximity of known pea R-genes and also of symbiosis genes needed for pea-rhizobia interactions.

Use of PAGE and the search for polymorphism

To date, most reports on the cloning of RGAs have used agarose gel electrophoresis to resolve PCR products, mainly to confirm satisfactory amplification with any particular
primer pair and to estimate the predicted RGA length in base pairs. The agarose gel electrophoresis is followed by direct cloning and sequencing, and finally the clones are used as RFLP probes to search for genotypic polymorphism among individuals. For example, Kanazin et al. (1996) and Mago et al. (1999), working on soybean and rice respectively, determined by restriction digestion the presence of multiple amplification products, of approximately the same length, in what appeared to be a single band resolved on 1–1.5% agarose gels. Under less stringent PCR conditions, the presence of one or two single large bands was accompanied by either extra unclear bands of variable size or low-molecular-weight smear. In any scenario, amplified RGA products show high heterogeneity which can be detected often at the agarose resolution level (Yu et al. 1996; Seah et al. 1998).

The use of RGAs as RFLP probes for genetic mapping research is generally laborious and time consuming. The experimental procedure initially involves PCR amplification, selection and purification of PCR products within the expected length range, cloning into bacterial vectors and screening of numerous colonies and inserts. Subsequently, all inserts are sequenced and DNA homology analyses and translated peptide comparisons reveal the identities of some of the inserts as genuine RGAs. Finally, RGA clones are used to search for informative polymorphisms among individual genotypes before they are tested on segregating populations. Additional delicate protocols such as radioisotope DNA labeling may be necessary (Yu et al. 1996; Donald et al. 2002; Madsen et al. 2003).

A faster and more straightforward approach for the analysis of PCR products from degenerate primers was first presented by Chen et al. (1998). After confirming positive PCR-amplifications in wheat, rice and barley genomic DNA, the PCR products were separated directly on high-resolution polyacrylamide gel electrophoresis (PAGE). The PCR-fragments resolved on PAGE were very heterogeneous and showed a variable but relatively high degree of polymorphism. Such polymorphism was applied to evaluate genetic relationships among germplasm. Thus, in the three monocot species the RGA markers were capable of differentiating near isogenic lines backgrounds, and furthermore, the RGA banding patterns were consistent with the known agronomic
characteristics and geographic origins of the genotypes. PAGE-polymorphic RGAs segregating in a wheat F₆ recombinant inbred line population were inherited as single locus markers and used for genetic mapping. Almost 85% of the RGAs detected genetic linkages with R-genes controlling resistance to the wheat stripe rust pathogen *Puccina striiformis*.

One significant advantage of RGAs over anonymous DNA markers [e.g. RAPDs, amplified fragments length polymorphisms (AFLPs)] is that they are more likely to amplify genes containing any of the R-gene domains, or others with close similarity. DNA sequencing analyses might still be needed in the screening and selection of RGAs from the total pool of amplified fragments, in particular to confirm the nature of the most interesting and useful markers whose chromosomal locations coincide with R-loci.

The purpose of this study was to identify and map RGAs in tomato and compare their chromosomal locations with those of known R-genes and QTLs for EB resistance. A second goal was to generate new PCR-based DNA markers that could be potentially used for mapping and marker-assisted selection purposes in tomato.

**Materials and methods**

**Plant material**

Backcross (BC) populations which had been previously developed in our laboratory from a cross between *L. esculentum* breeding line NC84173 (pistillate parent) and *L. hirsutum* accession PI126445 (Zhang et al. 2002) were used in this study. NC84173 is a horticulturally superior, advanced breeding line with a determinant growth habit, mid-season maturity, and susceptibility to EB (Foolad et al. 2000). PI126445 is a self-incompatible accession with a indeterminate growth habit, vigorous vine and extremely
late maturity that was previously identified to be highly resistant to tomato EB (Gardner 1988; Foolad et al. 2000). For the production of the BC populations, a single F_1 hybrid plant was used as the pollen parent to hybridize plants of NC84173 and produce BC_1 seeds. The BC_1 population was used in our laboratory for marker analysis and map construction (Zhang et al. 2002). The BC_1 and BC_1S_1 (self-pollinated progeny of BC_1) progeny were evaluated for EB resistance in 1998 (BC_1), and 1999 and 2000 (BC_1S_1) and used for QTL mapping (Foolad et al. 2002) (Table 1).

**RFLP analysis, map construction and QTL identification**

The procedures for RFLP analysis and map construction (Zhang et al. 2002) as well as for QTL mapping were previously described (Foolad et al. 2002) (Table 1).

**RGA analysis**

**Selection of primers**

Twenty-two oligonucleotide primers with different degrees of degeneracy, designed from the conserved LRR, NBS and STK domains of several known R-genes were chosen for this study (Table 2). Degenerate primers containing inosine or a mixed base at the third codon position fit the consensus sequences of corresponding amino acids. Slight alterations at the third nucleotide of each codon (for mixed bases) select for different amino acid variants and lead to the amplification of different fragments. Thus, in order to cover an extended range of potential candidate sequences encoding the consensus amino acid motifs, variants of these primers were used.

**PCR amplification**

Nuclear DNA was extracted from approximately 10 g of leaf tissue from each BC_1 plant using standard protocols for tomato (Bernatzky and Tanksley 1986). Standard PCR
conditions were used for the amplification of RGAs (Foolad et al. 1995). Briefly, each PCR reaction was performed in a 25 µl volume consisting of 300 µM each of dATP, dCTP, dGTP and dTTP, 5 mM of MgCl₂, one unit of Taq DNA polymerase, 2.5 µl of 10X buffer (PCR Core System I; Promega, Madison, WI), 2 µM of each primer, and 40 ng of genomic DNA that was used as template. For control reactions, the template was substituted by sterile nuclease-free water to ensure that there was no contamination. Each PCR reaction was overlaid with mineral oil and carried out in a Perkin Elmer DNA Thermal Cycler 480, programmed for 4 min at 94°C for an initial denaturation, and 36 cycles of 1 min at 94°C, 1 min at 50°C and 1.5 min at 72°C, followed by a final 7 min extension at 72°C.

_Electrophoresis and silver staining_

Denaturing polyacrylamide gel electrophoresis (PAGE) was used to separate the PCR-amplified DNAs. Gels (7M Urea, 6% polyacrylamide) were prepared on a sequencing apparatus (420 x 330 x 0.4 mm, Fisher Biotech, Springfield, NJ) using Bind- and Repel-Silane as recommended by the manufacturer (Promega). After polymerization, gels were pre-run in 1X TBE buffer for 30 min at 40W (~1400 V) to reach a gel temperature of 50°C before loading the samples. Following PCR amplification, 12 µl of loading solution (10M Urea and 0.08% xylene cyanole) was added to each 25-µl sample, heated at 94°C for 4 min to denature the amplified DNA, and immediately put on ice. After cleaning the gel loading area, a 0.4-mm-thick shark tooth comb was inserted into the gel. Subsequently, 7 µl of each PCR-amplified sample were loaded. Each gel accommodated 60 PCR-amplified DNA samples and three DNA size markers as positional references between different runs (non-denatured 1Kb, 50bp, and PCR Low). Gels were run at 35W (~1350 V) for 3.5-4 hr, and stained using a standard silver staining protocol (Promega). After electrophoresis, gels, fixed to the Bind-Silane surface of one glass plate, were silver-stained following the manufacturer’s protocol (Promega). Finally, gels were air dried overnight at room temperature and stored in the dark for future scoring and scanning.
**Size determination and cloning of RGA fragments**

To determine the size of polymorphic RGA fragments, DNA bands were isolated directly from the denaturing PAGE and re-amplified, using the needle scratching and PCR re-amplification method (Stumm 1997). The re-amplified products were separated on 1.5% agarose gels and stained with ethidium bromide. The approximate size of RGA fragments was determined using DNA size markers.

Cloning was performed by excising the re-amplified bands from the ethidium bromide-stained agarose gels, cleaning with the QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA) and inserting them in pSTBlue-1 vectors (Novagen, Madison, WI). Five randomly selected colonies were chosen for every cloning event to verify by sequencing either the presence of a single band or several bands of approximately the same length.

**Cloning and sequencing of PAGE monomorphic markers**

DNAs were digested with 5 restriction enzymes, *DraI, EcoRI, EcoRV, HindIII* and *XbaI*, according to the manufacturer’s instructions (Amersham Pharmacia Biotech, Piscataway, NJ). Agarose gel electrophoresis, Southern blotting, DNA hybridizations and autoradiography were as described by Foolad and Jones (1993). Probes were labeled with $^{32}$P-dCTP by the hexanucleotide oligo extension (Feinberg and Vogelstein 1983).

**Statistical analysis**

**Map Construction**

The construction of the framework map was described by Zhang et al. (2002). In this study, however, RGA markers were added to the map following the standard procedure using the MAPMAKER Program v3.0 (Lander et al. 1987). Briefly, the “group” command was used to assign marker loci into linkage groups ($\log_{10}$ of odds, LOD ≥ 4.0).
Subsequently, 3-point and multipoint linkage analyses and the "compare" and "try" commands were used to find the best order of the loci within each group. The "ripple" command was used to verify the order (LOD ≥ 3.0). Once the linear arrangement of marker loci along each chromosome was determined, the recombination frequencies between markers were estimated using multipoint analyses with the "map" command. The Kosambi mapping function (Kosambi 1944) was used to convert recombination frequencies to map distances in centimorgans (cM).

**QTL identification**

QTLs for EB resistance in the BC₁ and BC₁S₁ populations were determined previously as described by Foolad et al. (2002). In the present study, however, computer analysis for QTL identification was repeated as more markers (i.e. RGAs) were added to the original map. Briefly, simple interval mapping (SIM) analysis, using computer programs MAPMAKER/QTL v. 1.1 (Lincoln et al. 1992) and QGene (Nelson 1997), was employed to identify marker intervals on the tomato chromosomes that contained QTLs.

**Results and discussion**

**Genetic linkage map and genome distribution in the BC₁ population**

Addition of RGAs to the framework map, which was previously developed in our laboratory (Zhang et al. 2002), resulted in the construction of a new linkage map with a total of 241 DNA markers, including 142 anchor RFLPs and 99 RGAs. The map spanned approximately 1,872 cM of the tomato genome with an average distance of 7.7 cM between markers, although some chromosomes were more marker-saturated than others (Figure 1). The number of markers per chromosome ranged from 9 (chr. 8) to 43 (chr. 1) with an average of 20 markers per chromosome. Chromosomes 1 and 2 had the largest
linkage groups, whereas chromosomes 4, 10 and 11 were among the smallest ones, consistent with previous linkage maps of tomato developed based on different interspecific populations. With the exception of chromosome 1 (see below), the length of the map and the order of the markers were in agreement with the high-density RFLP map of tomato, which was previously constructed based on a *L. esculentum* × *L. pennellii* F$_2$ population (Pillen et al. 1996).

Of the 241 marker loci scored in the BC$_1$ population, 152 (63%), located on 8 of the 12 tomato chromosomes, exhibited significant deviation from the expected 1:1 segregation ratio for a backcross population. In all cases, there was an excess of *L. esculentum* homozygotes. Genome composition of the 145 BC$_1$ plants ranged from 68% to 98% from the *L. esculentum* parent, with an average of 82.4%, larger than the expected 75% for a BC$_1$ population. This was, however anticipated because of (1) the use of an interspecific cross, and (2) the selections which were made against self-incompatibility, indeterminant growth habit and late maturity in the BC$_1$ population (Zhang et al. 2002), all three traits contributing from the *L. hirsutum* genome. The extent of skewness was rather large (63%), although only markers on chromosomes 1, 5 and 6 were severely skewed. Selections against self-incompatibility, growth habit and late maturity contributed to a high level of skewness in this population (Zhang et al. 2002). Such selections, however, were necessary in order to identify QTLs with true effects on EB resistance (Zhang et al. 2002). The highest level of skewness was observed in markers on the short arm of chromosome 1, where the $S$ locus controlling self-incompatibility has been mapped ( Tanksley and Loaiza-Figueroa 1985; Pillen et al. 1996). Interestingly, 26 RGA loci, exhibiting a significantly skewed segregation, mapped on the same region of chromosome 1 where the tomato *R*-gene clusters *Cf-* and *rx* have been previously located ( Pillen et al. 1996) (the potential effects of this RGA cluster is further discussed below). Similar skewed segregation for markers on chromosome 1 was previously reported in other interspecific crosses of tomato where marker-assisted selection (MAS) was performed for self-compatibility before conducting QTL mapping (Fulton et al. 1997b; Bernacchi et al. 1998). The second highest level of skewness was observed for markers on chromosome 6 on which the self-pruning (*sp*) locus, controlling the indeterminate
growth habit, is located (Paterson et al. 1988; Grandillo and Tanksley 1996; Fulton et al. 1997a). In tomato, self-incompatibility and indeterminate growth habit are undesirable characteristics for genetics and breeding studies (including QTL mapping) as well as for commercial production under field conditions (Stevens and Rick 1986; Bernacchi and Tanksley 1997; Fulton et al. 1997a).

Identification of RGAs

Depending on the primer pair used, between approximately 10 (for primer pair S1-AS1) and 100 bands (for primer pair PtoKin1-PtoKin2) were detected on PAGE. A total of approximately 600 bands was detected from all primer pairs used. This indicated that sequences containing potential NBS, LRR and STK coding regions are abundant in tomato, and support the previous suggestion that PCR-amplification with RGA primers results in products of many different lengths and reproducibility (Leister et al. 1996; Chen et al. 1998; Mago et al. 1999).

PAGE-polymorphic RGAs

Of the total of 600 bands produced by all primer pairs, about 130 had strong reproducible signals in the two parents, of which 29 exhibited direct scorable segregations in the BC$_1$ population, and thus were mapped (Figure 1) (additional RGA markers produced by the same primer combinations were detected and mapped in a different population of the same cross (Zhang et al. 2003)). The fragment size for the PAGE-polymorphic RGAs ranged from 140 bp to 950 bp, with an average size of 340 bp. However, only 4 markers (13%) had fragment size larger than 550 bp and the remaining ranged between 140 bp and 550 bp, as determined on agarose gels. In the absence of introns in $R$-genes, the degenerate primers are expected to amplify DNA fragments in the range of 150 and 550 bp (Leister et al. 1996). However, in previous studies, using other plant species such as soybean, potato and Arabidopsis, several RGAs were reported larger than 550 bp (Kanazin et al. 1996; Leister et al. 1996; Chen et al. 1998; Speulman et al. 1998). Larger fragments could be due to different causes, for example, the presence of introns between
the primer annealing sites, the amplification of fragments from two adjacent resistance genes, the amplification of pseudogenes, and PCR artifacts. Among the 29 PAGE-polymorphic RGAs, the highest number (6) was obtained for primer pairs ANo.2-ANo.3 and PtoKin1-PtoKin2, and the lowest (1) for primer pairs S2-AS3 (Table 7). No PAGE-polymorphic markers were produced by the primer pairs NBS-F1–NBS-R1, S1-AS3, S2-AS1, PtoFen-S–PtoFen-AS and LRKfor-LRKrev; nevertheless, reproducible monomorphic bands were obtained from these combinations.

### PAGE-monomorphic RGAs

In addition to the 29 polymorphic bands, 95 amplification products were monomorphic in the BC₁ population, as determined by denaturing PAGE. To investigate the nature of these monomorphic bands, and to determine whether they could be converted into informative polymorphic RFLP markers, they were isolated from dried gels, cloned and sequenced (as described in Chapter 3). Sequence alignment demonstrated that these monomorphic clones (hereafter referred to as PAGE-monomorphic RGAs) could be ordered in 75 unique groups that shared less than 80% nucleotide identity and ranged from 100 bp to 860 bp, with an average length of 340 bp. The clones were used as RFLP probes and hybridized to blots containing restriction-digested genomic DNAs of the parents of the BC₁ population. Of the 75 clones, 37 produced useful polymorphism and were retained for genetic mapping. The remaining clones produced either multiple-copy banding patterns or remained monomorphic, and thus were not useful for mapping.

It is interesting to note that, when used as RFLP probes, these PAGE-monomorphic RGAs were capable of displaying a higher number of polymorphic bands than the PAGE-polymorphic RGAs. The 37 PAGE-monomorphic clones that remained for genetic mapping produced a total of 70 segregating loci.
Mapping of RGAs

Linkage analysis indicated that the 29 PAGE-polymorphic and 70 PAGE-monomorphic segregating loci were distributed throughout the tomato genome, mapping to 11 of the 12 chromosomes (Figure 1). These results are comparable to previous reports suggesting that PAGE-polymorphic RGAs can be used as genetic markers for genome mapping alone (Chen et al. 1998) or in combination with RGA-RFLP probes (Ramalingam et al. 2003). In several cases, RGA marker loci appeared in clusters, similar to that previously reported for other R-genes and RGAs in other plant species (Kanazin et al. 1996; Yu et al. 1996; DeJong et al. 1997; Aarts et al. 1998; Ashfield et al. 1998; Simons et al. 1998; Speulman et al. 1998; Thomas et al. 1998; Ramalingam et al. 2003). For example, in all chromosomes except 3, 6, 8 and 11, three or more RGA loci, which were detected by the same or different primer pair (or PAGE-monomorphic probe), mapped to the same or nearby positions (Figure 1). This observation indicates that different primers may initiate amplification of closely-linked RGA marker loci, which might be members of the same or different gene families. Similarly, hybridizations with the same or different PAGE-monomorphic RGAs also detected loci clustering with themselves only (e.g. Chromosome 10) or with PAGE-polymorphic RGAs (e.g. Chromosome 9).

The highest numbers of mapped RGAs were on chromosomes 1 (32) and 2 (18), whereas chromosome 11 had only one mapped RGA and chromosome 8 had none (Figure 1). It is interesting to note the high density of skewed RGAs ($3.91 < \chi^2 < 98.32$) on the short arm of chromosome 1 in proximity to the $S$ locus. For the three traits against which this population was selected and developed, RGAs showed the strongest skewed linkage towards SI, causing their strange clustering around the $S$ locus. Additionally, the skewed RGAs in chromosome 1 produced an abnormal increase of 145 cM in the total chromosomal length, doubling the original length determined based on RFLP anchor markers alone. Interestingly, two $R$-gene families were previously located on the same chromosomal region of the $S$ locus, the $Cf$- (Haanstra et al. 1999) and the $rx$-gene families (Yu et al. 1995) (Figure 1). Because RGAs have the tendency to cluster around $R$-genes, it may be possible that some of these markers showing skewed segregation on
chromosome 1 might also be linked or be part of the Cf- and rx gene families. However, due to the extreme degree of skewness and the alterations in the genetic distances, it is not possible in this study to determine the authenticity of the observed co-localization in this chromosomal region. Map saturation with specific RFLPs anchors and a larger population size are necessary to assess the marker associations, order and relative distances in the vicinity of the S locus.

This BC1 population was also selected for indeterminate growth habit (i.e. sp/sp genotype) and, as expected, markers on chromosome 6, including two mapped RGAs, were greatly skewed towards the determinate phenotype of the L. esculentum parent; however, no alteration in the genetic distances was detected.

Despite the dubious results observed in chromosome 1, overall inspection of the RGAs chromosomal locations indicated that their distribution might not be random. Using previously-published information, approximate positions of known tomato R-genes were inferred based on linkage to reference markers (Figure 1). In this study, several RGA loci were mapped to genomic regions that were known to contain tomato resistance loci, including regions on chromosomes 2 (Tm-1), 3 (Asc), 5 (Pto, Fen and Prf), 6 (OI-1, Mi, Ty-1, Cm6.1, Cf-2, CF-5, Bw-5 and Bw-1), 7 (I-1, I-3, and Ph-1), 9 (Ve, Tm-2a and Fr1), 10 (Cm10.1) and 12 (Lv). These apparent positional associations and clustering of RGAs suggest their possible involvement with disease resistance, either as linked DNA marker sequences or as parts of R-genes loci. However, before establishing any functional relationship, it is necessary to map, clone and characterize the associated RGAs in other populations segregating for the various R-genes. Similar associations were previously identified in several other plant species and a few functional relationships were established, as described in Chapter 1. Because of their reproducibility, the identified RGAs can also be used as genetic markers for germplasm evaluation and genetic fingerprinting.

This study did not identify RGAs coinciding with several other known R-genes in tomato as illustrated in the genetic map (Figure 1). This may reflect the fact that the
search for RGAs was not exhaustive, probably because of the relatively low similarity between some primer combinations and genomic DNA or the high astringency conditions used in the PCR reactions. Additional primers and mapping populations may be necessary to detect a larger number of RGA marker loci that provide wider genome coverage. As new R-gene specificities are discovered (e.g. the WRKY domain of the new RRS1-R gene), more peptide sequences will become available as templates for new primer designs.

**Identification of QTLs for Early Blight resistance**

Fourteen QTLs controlling EB resistance in tomato were previously identified in the BC$_1$ and BC$_1$S$_1$ populations as described by Foolad et al. (2002) (Table 3, Figure 2). However, recent addition of more RGAs (PAGE-monomorphic markers), somewhat changed the landscape of the QTL LOD plot for several chromosomes (Figure 2).

**Relationship between genomic locations of RGAs and QTLs**

In this study, the 26 significantly skewed RGAs and one RFLP probe that mapped to the short arm of chromosome 1 (Figure 1) may have had confounding effects on the QTL analyses. To avoid these effects, a modified version of chromosome 1 was derived from the original by removing such ambiguous chromosomal region associated with the $S$ locus. Thus, the new genetic map of chromosome 1 is only 109 cM and with an acceptable coverage as limited by the two farthest RFLP anchors (CT197 and CT137) positioned on the two chromosomal ends (Pillen et al. 1996) (Figure 2).

RGA markers were assessed for their positional association with QTLs identified for EB resistance. Among the 99 RGA loci, the positions of 27 coincided with QTLs located in chromosomes 5, 9, 10, 11 and 12 (Figure 2). Furthermore, using anchor markers as positional references, RGAs within QTLs were also in proximity to tomato $R$-genes governing specific resistance to bacterial speck ($Pto$ and $Prf$, chromosome 5), tobacco...
mosaic virus, late blight, spotted wilt virus (*Tm-2a*, *Ph-3* and *Sw-5* respectively, chromosome 9), bacterial canker (*Cm10.2*, chromosome 10), fusarium wilt, yellow leaf curl virus (*I-2* and *Ty-2* respectively, chromosome 11) and powdery mildew (*Lv*, chromosome 12) among others (Figure 1, Figure 2).

The most interesting associations of RGAs with QTLs for EB resistance and *R*-genes were found in chromosomes 5 and 9. In chromosome 5 a cluster of 7 RGAs mapped within a 29-cM region (20% of the total chromosome length) and co-localized with the *Pto* and *Prf* *R*-genes and the QTL *EBR5.2*. An even denser cluster is the one on chromosome 9 containing 11 RGAs within a 18.7-cM region (10% of the total chromosome length), contained in the QTL *EBR9.1* and in close proximity to the *Tm-2a* and *Fr1* *R*-genes. Both QTLs *EBR5.2* and *EBR9.1* independently explain 12.0% and 25.7%, respectively, of the phenotypic variation for EB resistance in the BC1S1-1999, whereas the combined actions of both QTL account for 34.1% of the total phenotypic variation. MAS using RGAs in those target regions on chromosome 5 and 9 would demonstrate their potential utility for improving tomato EB resistance cultivars (Table 3).

A different scenario was observed in chromosome 2, with 17 RGAs in a cluster spanning 21 cM (11% of the total chromosome length). Although no QTL was detected in that region, these RGAs co-localized with the *R*-gene *Tm-1*. In order to further evaluate the nature and potential use of similar RGA clusters, it may be necessary to consider the same set of mapped markers when evaluating the population against other pathogens and track any possible appearance of new QTLs for different disease responses in the region. For example, Wang et al. (2001) mapping candidate *R*-genes in rice, found common QTLs against the blast fungus, bacterial blight and sheath blight in rice, and also other different QTLs that were specific for only one or two of such diseases. It is possible that RGA clusters are capable of mapping *R*-gene regions that are significantly involved in the expression of resistance to a particular pathogen but not to others, and thus fail to detect some QTLs.

Whether the RGA markers are part of genes within QTLs controlling EB resistance
could not be determined solely by this mapping study (see marker sequencing and comparisons in Chapter 3). Because of the consistency of the QTL LOD scores and the similar landscape of the QTL peaks for the three years of phenotypic evaluation, it is expected that the positional association between RGA markers and QTL did not arise by chance. However, no statistical or probabilistic analysis could be implemented at this point to estimate the odds for the random occurrence of these associations, mostly because such tests would have to assume certain parameters, which would be in disagreement with the real genotypic and phenotypic background of the population. For example, for such analyses, it would be necessary to assume fixed and equal chromosomal lengths, invariable genome size, consistent homologous recombination for equal chances of crossing-over events (and equal distribution of polymorphic sites for their detection), and uniform distributions and spacings of marker loci along the genome. Furthermore, in such assumptions some QTL LOD values would be affected, in particular those controlled by polymorphic clustered alleles that in reality are difficult to separate by genetic recombination events (CTC 2003).

A permutation analysis performed with the QGene computer program tested the association of 27 RGA markers with QTLs for EB resistance (i.e. RGA markers that crossed the established significant threshold of LOD 2.4) in every generation and year of phenotypic evaluation (BC$_1$-1998, BC$_1$S$_1$-1999, BC$_1$S$_1$-2000, Tables 4, 5 and 6 respectively). The analysis was intended to confirm that the genetic associations between RGA and the QTLs did not occur as random events, rather they confirmed genetic linkage to genes contributing to the overall expression of the trait. The permutation analysis is a random resampling procedure for linked and non-independent markers. It basically shuffles repeatedly the genotypic data one thousand times in order to break any non-random association between the originally linked markers. The analysis output calculates the range of the statistical LOD scores under the assumption that there was no initial association between the marker genotype and phenotype (i.e. null hypothesis accepted) (Churchill and Doerge 1994). The results of the analysis shown in Tables 4, 5 and 6 present the LOD scores for each marker that are greater than 95% and 99% of the 1000 LODs obtained after the random shuffling of the phenotypic data. The results
indicate that for most of the RGA markers there is a difference between their calculated and permutation LODs, and because the permutations do not favor any of the parental lines as the original phenotypic data does, (and regardless of the parental line that contributed the polymorphic markers) the permutation LOD scores may be greater or lesser than the calculated scores, with a more uniform and less deviated distribution. This uniform distribution of scores is illustrated in Figure 3, where the obtained QTL LODs are graphically compared with the 1000-shuffle permutation LODs only for the BC$_{1}$S$_{1}$-1999 evaluation (Table 5). By observing Figure 3 it can be inferred that if the associations between RGA markers and QTLs for EB resistance had been due to random events, the scores for calculated and permutation LODs would have approximated and shared the shape of QTL peaks. However, in certain chromosomal regions, calculated QTLs with large effects (significant LODs with unshuffled data) depart considerably from the permutation LODs, for example, QTLs in chromosomes 5, 9 or 12 (Figure 2, Figure 3). Nevertheless, while the permutation analysis only confirmed the non-random occurrence of the QTLs for EB resistance (and including the markers that detected them) for this particular backcross population, it did not calculate any probability of such occurrences for any particular marker individually.

Genetic mapping and positional comparisons provide strong initial evidence of the nature of the segregating loci and their associations with candidate genes for EB resistance. Several previous investigations have demonstrated linkage associations between genomic locations of RGAs and $R$-genes in different plants including soybean (Kanazin et al. 1996; Yu et al. 1996), potato (Leister et al. 1996), lettuce (Shen et al. 1998), maize (Collins et al. 1998), rice (Chen et al. 1998; Mago et al. 1999), and wheat and barley (Chen et al. 1998; Seah et al. 1998). In $Arabidopsis$, for example, RGAs amplified by the primer pair ANo.2-ANo.3 have actually derived from functional members of an $R$-gene family (Aarts et al. 1998; Speulman et al. 1998), proving that the approach has the potential to identify viable candidate loci. In tomato, vertical specific resistance against EB does not exist; however, the tagging of $R$-genes by the RGA approach has the potential to reveal the positions of the most attractive QTLs where candidate genes for defense against EB may be located. $R$-gene tagging ultimately tests
the hypothesis that R-genes account for a significant proportion of the quantitative variation of resistance (Leonards-Schippers et al. 1994; Michelmore 1996). Therefore, localizing R-genes against other pathogens of tomato will be of great value to determine the locations of the most important QTLs controlling resistance against virtually any disease, EB among them. Genome sequencing, QTL cloning and molecular characterization of the RGAs are necessary before any additional functional relationship can be established (see Chapter 3).

Use of RGA markers for MAS

Previous studies in our laboratory used MAPMAKER/QTL and QTL Cartographer computer programs to identify the most suitable combinations of QTLs that could account for a significant proportion of the total phenotypic variation. Six QTLs, more or less independent of each other and exhibiting additive effects, were selected as the most suitable combination for MAS to transfer resistance from PI126445 to the cultivated tomato (Foolad et al. 2002). The selection of each QTL combination was based on several criteria, including the QTL expression in different generations and years, as well as the independence and magnitude of their effects. Because these QTLs were identified after the elimination of confounding factors (SI, indeterminant growth habit, late-maturity), their effects on EB resistance should be real and of significant value for the transfer of resistance via MAS (Foolad et al. 2002).

In this study, the associations confirmed between the locations of RGAs and QTLs for EB resistance support the use of RGAs as molecular tools for conducting MAS. Twenty-seven RGA markers coincided with the QTLs identified on chromosomes 5, 9, 10, 11 and 12 (LOD > 2.4, Figure 2), and eighteen of them exhibited genetic associations with two of the QTLs selected for MAS (EBR5.2 and EBR9.1, Figure 2) (Foolad et al. 2002). At this stage, only the most important QTL-linked RGA markers are considered for direct breeding purposes, and thus the application of RGAs to MAS appears less laborious than conventional genetic mapping. For example, after the identities of the RGAs are revealed by DNA sequencing (as discussed in Chapter 3), it is possible to
develop specific sequence characterized amplified regions (SCAR, Paran and Michelmore 1993) to test individual plants in MAS projects for the presence of one QTL-linked RGA marker at a time.

Because in most MAS programs the backcross breeding approach is used (where QTLs controlling the trait of study are originated from the donor parent), the marker-assisted introgression of different QTL combinations is greatly facilitated by the simple dominant inheritance of the RGA markers. This allows the breeder to test only two possible allelic arrangements per locus, homozygotes of *L. esculentum* (*E/E*) and heterozygotes (*E/H*); where only the heterozygote combination will carry the segregating marker from the donor parent, *L. hirsutum* (*H*). Simultaneous introgression of selected QTLs into the cultivated tomato by MAS using SCARs is feasible, providing opportunities to rapidly develop cultivars with enhanced and durable EB resistance.

**References**


Grant, M.R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W.,
and Dangl, J.L. 1995. Structure of Arabidopsis RPM1 gene enabling dual

The Cf-ECP2 gene is linked to, but not part of, Cf-4/Cf-9 cluster on the short arm
of chromosome1 of tomato. Mol Gen Genet 262: 839-845.


Hanson, P.M., Bernacchi, D., Green, S., Tanksley, S.D., and Muniy-Appa, V. 2000.
Mapping a wild tomato introgression associated with tomato yellow leaf curl

Hoffmann, M. 1996. IPM for Diversified fresh market vegetable producers in NJ, NY,
and PA: Progress Report., pp. 127. National IPM Implementation Program,
Northeast Region.


Johal, G., and Briggs, S. 1992. Reductase activity encoded by the HM1 disease resistance

Jones, D.A., Thomas, C.M., Hammond-Kosack, K.E., Balint-Kurti, P.J., and Jones,
J.D.G. 1994. Isolation of the tomato Cf-9 gene for resistance to Clasosporium
fulvum by transposon tagging. Science 266: 789-793.


Table 1. Early Blight (EB) disease severity (final percent defoliation, ±SE) and the area under the disease progress curve (AUDPC, ± SE) for the parental lines and BC$_1$ and BC$_1$S$_1$ progeny of a cross between *L. esculentum* breeding line NC84173 (EB susceptible) and *L. hirsutum* accession PI126445 (EB resistant). Data taken from Foolad et al. 2002.

<table>
<thead>
<tr>
<th>Generation</th>
<th>Growth Habit$^a$</th>
<th>n$^b$</th>
<th>Final Percent Defoliation</th>
<th>AUPDC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Average</td>
<td>Range</td>
</tr>
<tr>
<td>P$_1$ (NC84173) (1999)</td>
<td>D</td>
<td>150</td>
<td>99.7±0.6</td>
<td>99.0–100.0</td>
</tr>
<tr>
<td>P$_2$ (PI126445) (1999)</td>
<td>I</td>
<td>150</td>
<td>5.0±0.0</td>
<td>5.0–5.0</td>
</tr>
<tr>
<td>F$_1$ (P$_1$ × P$_2$) (1999)</td>
<td>I</td>
<td>30</td>
<td>8.3±1.5</td>
<td>7.0–10.0</td>
</tr>
<tr>
<td>BC$_1$ (1998)</td>
<td>D</td>
<td>146</td>
<td>50.1±23.4</td>
<td>15.0–100.0</td>
</tr>
<tr>
<td>BC$_1$S$_1$ (1999)</td>
<td>D</td>
<td>146</td>
<td>59.2±18.7</td>
<td>25.7–100.0</td>
</tr>
<tr>
<td>BC$_1$S$_1$ (2000)</td>
<td>D</td>
<td>146</td>
<td>69.5±17.8</td>
<td>25.0–100.0</td>
</tr>
</tbody>
</table>

$^a$ D = Determinant, I = Indeterminant

$^b$ n = Total number of plants (for the P$_1$, P$_2$ and BC$_1$) or families (for the BC$_1$S$_1$) evaluated for EB resistance. For the BC$_1$S$_1$-1999 population 30 plants of each family and for BC$_1$S$_1$-2000 population 20 plants of each family were evaluated

$^c$ Not available
Table 2. Oligonucleotide primers designed from conserved peptide domains encoded by various R-genes.

*Code for mixed bases: Y=C/T, N=A/G/C/T, R=A/G and D=A/G/T

<table>
<thead>
<tr>
<th>Group</th>
<th>Primers</th>
<th>Sequences (5' - 3')*</th>
<th>Design Basis</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>I LRR</td>
<td>CLRR for</td>
<td>TTTTCGTGTTCACAGACG</td>
<td>LRR domain of the tomato Cf-9 gene conferring resistance to Cladosporium fulvum</td>
<td>Chen et al. 1998</td>
</tr>
<tr>
<td></td>
<td>CLRR rev</td>
<td>TAACGTCTATCGACTTCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RLRR for</td>
<td>CGCAACCACTAGAGTAAC</td>
<td>LRR domain of the Arabidopsis RPS2 gene conferring resistance to Pseudomonas syringae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RLRR rev</td>
<td>ACACTGGTCCATGAGGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XLRR for</td>
<td>CCGTTGGACAGGAGGAG</td>
<td>LRR domain of the rice Xa21 gene conferring resistance to Xanthomonas campestris pv oryzae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XLRR rev</td>
<td>CCCATAGACCGACCTTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ANo. 2</td>
<td>TATAGCGGCCGCIARIGCIARIGGIARNCC</td>
<td>Conserved P-loop and hydrophobic NBS regions of the N and RPS2 genes from tobacco and Arabidopsis respectively</td>
<td>Speulman et al. 1998</td>
</tr>
<tr>
<td></td>
<td>ANo. 3</td>
<td>ATATGCGGCCGCGGIGGIGRTIGGIAARACNAC</td>
<td>Conserved peptide sequence of two NBS domains present in the N y RPS2 genes of tobacco and Arabidopsis respectively</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NBS F1</td>
<td>GGAATGGGNGGNGTNGGNAARAC</td>
<td>Conserved hydrophobic and P-loop NBS domains of the N and RPS2 genes from tobacco and Arabidopsis and the L6 gene from flax conferring resistance to flax rust</td>
<td>Yu et al. 1996</td>
</tr>
<tr>
<td></td>
<td>NBS R1</td>
<td>YCTAGTTGRTYDATDYYTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S1</td>
<td>GGTTGGGTTGGGAAGGAAACG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>GIGIGGIGGIGIGGIGGIAICAC</td>
<td>Conserved hydrophobic and P-loop NBS domains of the N and RPS2 genes from tobacco and Arabidopsis and the L6 gene from flax conferring resistance to flax rust</td>
<td>Leister et al. 1996</td>
</tr>
<tr>
<td></td>
<td>AS1</td>
<td>CAACGCTAATGGCAATCC</td>
<td></td>
<td>Mago et al. 1999</td>
</tr>
<tr>
<td></td>
<td>AS3</td>
<td>IAGIGCIAAGGIGGACC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ptokin1</td>
<td>GCATTGGAACAAAGGTGAA</td>
<td>Serine/threonine protein kinase domain of the tomato Pto gene conferring resistance to the bacterial pathogen Pseudomonas syringae pv tomato</td>
<td>Chen et al. 1988</td>
</tr>
<tr>
<td></td>
<td>Ptokin2</td>
<td>AGGGGAGCACCACCTAG</td>
<td></td>
<td>Leung, H. (pers. commun.)</td>
</tr>
<tr>
<td></td>
<td>Ptokin3</td>
<td>TAGTTGCACTGTTTACAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ptokin4</td>
<td>AGTGCTTGTAGGTTATC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LRK for</td>
<td>GAYGTNAARCCIGARAA</td>
<td>Serine/threonine kinase sequence subdomains of the wheat Lr10 gene conferring resistance to Puccinia recondita.</td>
<td>Feuillet et al. 1997</td>
</tr>
<tr>
<td></td>
<td>LRK rev</td>
<td>TCYGGYCRATRTNACCNGGIGICCC</td>
<td>Conserved protein kinase domains of the tomato Pto and Fen gene family, the latter conferring susceptibility to the insecticide fenthion.</td>
<td>Leister et al. 1996</td>
</tr>
<tr>
<td></td>
<td>PtoFen-S</td>
<td>ATGGGAAGCAAGATTCAAGGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PtoFen-AS</td>
<td>TGGGCACAAATCTCATCAAGC</td>
<td>Conserved protein kinase domains of the tomato Pto and Fen gene family, the latter conferring susceptibility to the insecticide fenthion.</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. QTLs detected for Early Blight (EB) resistance based on simple interval mapping (SIM) in BC₁ and BC₁S₁ populations of an interspecific cross between *L. esculentum* (NC84173; EB susceptible) and *L. hirsutum* (PI126445; EB resistant). LOD = log-likelihood; PVE = percent phenotypic variation explained; \( E = L. esculentum \) allele; \( H = L. hirsutum \) allele; Phenotypic effect = difference between \( E/H \) and \( E/E \) in AUDPC. Shadowed rows correspond to non-significant LOD scores. Data taken from Foolad et al. 2002.

<table>
<thead>
<tr>
<th>QTLs</th>
<th>Interval</th>
<th>Chromosome</th>
<th>LOD</th>
<th>PVE</th>
<th>Phenotypic effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BC₁</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>EBR1.1</em></td>
<td>TG559 - TG208A</td>
<td>1</td>
<td>6.98</td>
<td>21.9</td>
<td>-409.86</td>
</tr>
<tr>
<td><em>EBR2.1</em></td>
<td>TG337 - CT59</td>
<td>2</td>
<td>2.89</td>
<td>15.3</td>
<td>-343.11</td>
</tr>
<tr>
<td><em>EBR5.2</em></td>
<td>CT202 - TG318</td>
<td>5</td>
<td>2.64</td>
<td>8.4</td>
<td>-295.93</td>
</tr>
<tr>
<td><em>EBR6.1</em></td>
<td>TG279 - CT107B</td>
<td>6</td>
<td>2.39</td>
<td>7.6</td>
<td>-261.60</td>
</tr>
<tr>
<td><em>EBR8.1</em></td>
<td>TG176 - CT92</td>
<td>8</td>
<td>2.30</td>
<td>7.3</td>
<td>-245.48</td>
</tr>
<tr>
<td><em>EBR9.1</em></td>
<td>RLRR-130 - CLRR-950</td>
<td>9</td>
<td>3.95</td>
<td>13.6</td>
<td>-321.51</td>
</tr>
<tr>
<td><em>EBR9.2</em></td>
<td>SS14-520.3 - TG429</td>
<td>9</td>
<td>5.13</td>
<td>16.2</td>
<td>-352.68</td>
</tr>
<tr>
<td><em>EBR9.3</em></td>
<td>SS19-530 - CT143</td>
<td>9</td>
<td>4.27</td>
<td>15.9</td>
<td>-348.10</td>
</tr>
<tr>
<td><em>EBR10.1</em></td>
<td>TG241 - TG403</td>
<td>10</td>
<td>6.83</td>
<td>20.2</td>
<td>-397.33</td>
</tr>
<tr>
<td><em>EBR11.1</em></td>
<td>CT168 - TG508</td>
<td>11</td>
<td>3.82</td>
<td>13.3</td>
<td>-332.22</td>
</tr>
<tr>
<td><em>EBR12.2</em></td>
<td>SS14-520.1 - SS1-530.1</td>
<td>12</td>
<td>4.24</td>
<td>13.4</td>
<td>-344.80</td>
</tr>
<tr>
<td><strong>BC₁S₁-1999</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>EBR1.1</em></td>
<td>TG559 - TG208A</td>
<td>1</td>
<td>3.63</td>
<td>11.9</td>
<td>-273.79</td>
</tr>
<tr>
<td><em>EBR2.1</em></td>
<td>TG337 - CT59</td>
<td>2</td>
<td>2.80</td>
<td>15.9</td>
<td>-316.98</td>
</tr>
<tr>
<td><em>EBR3.1</em></td>
<td>TG411 - TG214</td>
<td>3</td>
<td>2.87</td>
<td>9.1</td>
<td>-272.66</td>
</tr>
<tr>
<td><em>EBR5.1</em></td>
<td>TG441 - CT242</td>
<td>5</td>
<td>2.55</td>
<td>7.9</td>
<td>-234.39</td>
</tr>
<tr>
<td><em>EBR5.2</em></td>
<td>XLRR-370 - SAS5-250.3</td>
<td>5</td>
<td>3.68</td>
<td>11.2</td>
<td>-309.38</td>
</tr>
<tr>
<td><em>EBR8.1</em></td>
<td>CD40 - TG176</td>
<td>8</td>
<td>2.98</td>
<td>10.3</td>
<td>-262.64</td>
</tr>
<tr>
<td><em>EBR8.2</em></td>
<td>TG330 - TG294</td>
<td>8</td>
<td>5.44</td>
<td>21.0</td>
<td>-363.48</td>
</tr>
<tr>
<td><em>EBR9.1</em></td>
<td>CLRR-950 - SAS5-250.1</td>
<td>9</td>
<td>8.52</td>
<td>25.0</td>
<td>-395.26</td>
</tr>
<tr>
<td><em>EBR10.1</em></td>
<td>TG241 - TG403</td>
<td>10</td>
<td>5.56</td>
<td>16.3</td>
<td>-322.32</td>
</tr>
<tr>
<td><em>EBR11.1</em></td>
<td>TG508 - TG651</td>
<td>11</td>
<td>3.82</td>
<td>11.5</td>
<td>-279.00</td>
</tr>
<tr>
<td><em>EBR11.2</em></td>
<td>CT55 - CD17</td>
<td>11</td>
<td>2.97</td>
<td>9.9</td>
<td>-270.98</td>
</tr>
<tr>
<td><em>EBR11.3</em></td>
<td>SAS11-760.2 - TG393</td>
<td>11</td>
<td>3.49</td>
<td>11.5</td>
<td>-289.57</td>
</tr>
<tr>
<td><em>EBR12.1</em></td>
<td>TG68 - CT79</td>
<td>12</td>
<td>2.47</td>
<td>8.2</td>
<td>-240.24</td>
</tr>
</tbody>
</table>
Table 4. One-thousand genotype-shuffle permutation test performed for RGA markers associated with QTLs for EB resistance based on the BC$_1$-1998 disease evaluations. The LOD column corresponds to the values obtained from the unshuffled mapping and phenotypic data for each marker crossing the QTL’s significant LOD threshold of 2.4. The LOD 95%ile and LOD 99%ile indicate the markers’ LOD values that were greater than the 95% and 99%, respectively, of the total 1000 shuffles LOD scores. The LOD max column contains the markers’ maximum 1000 shuffles LOD score.

<table>
<thead>
<tr>
<th>Marker</th>
<th>LOD</th>
<th>LOD (95%ile)</th>
<th>LOD (99%ile)</th>
<th>LOD max</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA6-220.1</td>
<td>2.41</td>
<td>4.23</td>
<td>7.06</td>
<td>10.77</td>
</tr>
<tr>
<td>LRK-500.1</td>
<td>2.41</td>
<td>4.03</td>
<td>6.75</td>
<td>17.71</td>
</tr>
<tr>
<td>CLRR-490</td>
<td>2.52</td>
<td>3.69</td>
<td>5.75</td>
<td>11.90</td>
</tr>
<tr>
<td>C3-180</td>
<td>2.64</td>
<td>3.51</td>
<td>6.47</td>
<td>17.01</td>
</tr>
<tr>
<td>S1-160.2</td>
<td>3.03</td>
<td>4.16</td>
<td>6.74</td>
<td>12.24</td>
</tr>
<tr>
<td>S5-240.1</td>
<td>3.14</td>
<td>4.31</td>
<td>7.09</td>
<td>10.49</td>
</tr>
<tr>
<td>SAS5-250.1</td>
<td>2.44</td>
<td>4.44</td>
<td>7.78</td>
<td>13.73</td>
</tr>
<tr>
<td>C2-170</td>
<td>2.70</td>
<td>4.22</td>
<td>6.65</td>
<td>12.71</td>
</tr>
<tr>
<td>S11-210</td>
<td>3.10</td>
<td>3.92</td>
<td>7.05</td>
<td>18.08</td>
</tr>
<tr>
<td>XLRR-360</td>
<td>3.52</td>
<td>3.63</td>
<td>5.70</td>
<td>9.30</td>
</tr>
<tr>
<td>SS19-530</td>
<td>3.70</td>
<td>3.79</td>
<td>6.13</td>
<td>11.63</td>
</tr>
<tr>
<td>CLRR-950</td>
<td>4.37</td>
<td>3.97</td>
<td>6.54</td>
<td>14.07</td>
</tr>
<tr>
<td>RLRR-130</td>
<td>4.52</td>
<td>3.89</td>
<td>7.52</td>
<td>12.32</td>
</tr>
<tr>
<td>RLRR-140</td>
<td>5.00</td>
<td>3.85</td>
<td>7.13</td>
<td>10.33</td>
</tr>
<tr>
<td>SS14-520.1</td>
<td>2.42</td>
<td>4.27</td>
<td>6.71</td>
<td>13.16</td>
</tr>
<tr>
<td>PK12-320</td>
<td>2.58</td>
<td>3.89</td>
<td>6.02</td>
<td>11.66</td>
</tr>
<tr>
<td>AN23-390</td>
<td>4.72</td>
<td>4.07</td>
<td>6.39</td>
<td>10.35</td>
</tr>
</tbody>
</table>
Table 5. One-thousand genotype-shuffle permutation test performed for RGA markers associated with QTLs for EB resistance based on the BC₁S₁-1999 disease evaluations. The LOD column corresponds to the values obtained from the unshuffled mapping and phenotypic data for each marker crossing the QTL’s significant LOD threshold of 2.4. The LOD 95%ile and LOD 99%ile indicate the markers’ LOD values that were greater than the 95% and 99%, respectively, of the total 1000 shuffles LOD scores. The LOD max column contains the markers’ maximum 1000 shuffles LOD score.

<table>
<thead>
<tr>
<th>Marker</th>
<th>LOD</th>
<th>LOD (95%ile)</th>
<th>LOD (99%ile)</th>
<th>LOD max</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLRR-490</td>
<td>2.45</td>
<td>3.72</td>
<td>5.73</td>
<td>10.90</td>
</tr>
<tr>
<td>PK12-700</td>
<td>2.68</td>
<td>4.04</td>
<td>7.50</td>
<td>13.76</td>
</tr>
<tr>
<td>S1-160.2</td>
<td>3.80</td>
<td>3.76</td>
<td>6.86</td>
<td>16.81</td>
</tr>
<tr>
<td>S5-240.1</td>
<td>4.47</td>
<td>4.15</td>
<td>7.49</td>
<td>14.48</td>
</tr>
<tr>
<td>XLRR-370</td>
<td>2.91</td>
<td>4.00</td>
<td>6.48</td>
<td>14.25</td>
</tr>
<tr>
<td>SAS13-510</td>
<td>3.81</td>
<td>4.06</td>
<td>6.71</td>
<td>11.42</td>
</tr>
<tr>
<td>SS11-270.2</td>
<td>4.41</td>
<td>4.09</td>
<td>7.22</td>
<td>12.39</td>
</tr>
<tr>
<td>P2-220</td>
<td>4.46</td>
<td>4.23</td>
<td>6.40</td>
<td>10.29</td>
</tr>
<tr>
<td>SAS5-250.1</td>
<td>4.50</td>
<td>3.96</td>
<td>7.05</td>
<td>19.33</td>
</tr>
<tr>
<td>SS11-270.4</td>
<td>4.56</td>
<td>3.79</td>
<td>6.67</td>
<td>13.91</td>
</tr>
<tr>
<td>NBS8-300.1</td>
<td>4.56</td>
<td>4.31</td>
<td>6.40</td>
<td>11.86</td>
</tr>
<tr>
<td>SAS18-130</td>
<td>4.84</td>
<td>4.15</td>
<td>6.52</td>
<td>12.57</td>
</tr>
<tr>
<td>NBS4-300.3</td>
<td>5.37</td>
<td>4.31</td>
<td>7.62</td>
<td>11.35</td>
</tr>
<tr>
<td>C2-170</td>
<td>5.41</td>
<td>4.22</td>
<td>7.11</td>
<td>9.94</td>
</tr>
<tr>
<td>S11-210</td>
<td>5.99</td>
<td>3.58</td>
<td>5.85</td>
<td>11.50</td>
</tr>
<tr>
<td>RLRR-130</td>
<td>6.39</td>
<td>4.13</td>
<td>6.61</td>
<td>10.13</td>
</tr>
<tr>
<td>SS19-530</td>
<td>6.52</td>
<td>3.90</td>
<td>6.57</td>
<td>19.65</td>
</tr>
<tr>
<td>RLRR-140</td>
<td>6.97</td>
<td>3.51</td>
<td>5.78</td>
<td>10.03</td>
</tr>
<tr>
<td>XLRR-360</td>
<td>7.54</td>
<td>3.92</td>
<td>7.40</td>
<td>14.82</td>
</tr>
<tr>
<td>CLRR-950</td>
<td>8.24</td>
<td>3.92</td>
<td>7.35</td>
<td>12.88</td>
</tr>
<tr>
<td>SAS11-760.2</td>
<td>2.89</td>
<td>3.69</td>
<td>7.78</td>
<td>11.53</td>
</tr>
</tbody>
</table>
Table 6. One-thousand genotype-shuffle permutation test performed for RGA markers associated with QTLs for EB resistance based on the BC1S1-2000 disease evaluations. The LOD column corresponds to the values obtained from the unshuffled mapping and phenotypic data for each marker crossing the QTL’s significant LOD threshold of 2.4. The LOD 95%ile and LOD 99%ile indicate the markers’ LOD values that were greater than the 95% and 99%, respectively, of the total 1000 shuffles LOD scores. The LOD max column contains the markers’ maximum 1000 shuffles LOD score.

<table>
<thead>
<tr>
<th>Marker</th>
<th>LOD</th>
<th>LOD (95%ile)</th>
<th>LOD (99%ile)</th>
<th>LOD max</th>
</tr>
</thead>
<tbody>
<tr>
<td>S5-240.1</td>
<td>3.16</td>
<td>4.58</td>
<td>8.93</td>
<td>15.77</td>
</tr>
<tr>
<td>C2-170</td>
<td>3.08</td>
<td>3.98</td>
<td>8.01</td>
<td>12.53</td>
</tr>
<tr>
<td>SAS18-130</td>
<td>3.10</td>
<td>3.87</td>
<td>7.78</td>
<td>13.39</td>
</tr>
<tr>
<td>SS19-530</td>
<td>3.28</td>
<td>4.40</td>
<td>6.69</td>
<td>10.99</td>
</tr>
<tr>
<td>NBS8-300.1</td>
<td>3.47</td>
<td>4.35</td>
<td>6.91</td>
<td>11.32</td>
</tr>
<tr>
<td>SAS5-250.1</td>
<td>3.49</td>
<td>4.30</td>
<td>8.06</td>
<td>11.72</td>
</tr>
<tr>
<td>SS11-270.2</td>
<td>3.50</td>
<td>3.92</td>
<td>6.51</td>
<td>10.91</td>
</tr>
<tr>
<td>P2-220</td>
<td>3.54</td>
<td>4.06</td>
<td>8.72</td>
<td>17.79</td>
</tr>
<tr>
<td>NBS4-300.3</td>
<td>3.54</td>
<td>3.93</td>
<td>7.18</td>
<td>12.10</td>
</tr>
<tr>
<td>SAS13-510</td>
<td>3.60</td>
<td>4.20</td>
<td>6.70</td>
<td>10.17</td>
</tr>
<tr>
<td>SS11-270.4</td>
<td>4.02</td>
<td>3.81</td>
<td>5.84</td>
<td>12.29</td>
</tr>
<tr>
<td>RLRR-130</td>
<td>4.77</td>
<td>3.51</td>
<td>5.93</td>
<td>11.13</td>
</tr>
<tr>
<td>RLRR-140</td>
<td>4.98</td>
<td>4.47</td>
<td>8.20</td>
<td>10.17</td>
</tr>
<tr>
<td>XLRR-360</td>
<td>5.48</td>
<td>4.16</td>
<td>6.61</td>
<td>10.37</td>
</tr>
<tr>
<td>CLRR-950</td>
<td>6.06</td>
<td>4.37</td>
<td>6.61</td>
<td>12.89</td>
</tr>
<tr>
<td>S11-210</td>
<td>6.12</td>
<td>3.95</td>
<td>8.10</td>
<td>14.54</td>
</tr>
<tr>
<td>PK12-320</td>
<td>2.43</td>
<td>3.81</td>
<td>5.85</td>
<td>10.11</td>
</tr>
<tr>
<td>AN23-390</td>
<td>3.28</td>
<td>3.85</td>
<td>6.5</td>
<td>10.39</td>
</tr>
</tbody>
</table>
Figure 1. A linkage map of tomato based on a BC$_1$ population of a cross between *Lycopersicon esculentum* (breeding line NC84173; pistillate and recurrent parent) and *L. hirsutum* (accession PI126445) and 241 DNA markers (142 RFLPs and 99 RGAs). Markers are listed at the right of chromosomes and centimorgan (cM) distances (based on Kosambi function) between adjacent markers at the left. RFLP markers included 127 random tomato genomic (TG) or cDNA (CD or CT) clones from Cornell University and 15 informative cDNAs clones (A, C, CC, CEL, KJB) from University of California at Davis. PAGE-polymorphic and PAGE-monomorphic RGA markers are shown in red and blue font respectively. The name of each RGA marker was derived from the names of the corresponding primers (see Table 2 for the list of primers) and the fragment size. The approximate locations of disease resistance genes (*R*-genes) and quantitative resistance loci (QRLs), as inferred from other published research, are shown to the right of chromosomes. The descriptions of the *R*-genes and QRLs are as follow: *Asc*, resistance to *Alternaria* stem canker (van-der-Biezen et al. 1995); *Bw* (*Bw*-1, *Bw*-3, *Bw*-4, *Bw*-5), QRLs for resistance to bacterial wilt (*Ralstonia solanacearum*) (Mangin et al. 1999); *Cf* (*Cf*-1, *Cf*-2, *Cf*-4, *Cf*-5, *Cf*-9, *Cf*-ECP2), resistance to leaf mould (*Cladosporium fulvum*) (Thomas et al. 1998; Haanstra et al. 1999); *Cm* (*Cm*1.1, *Cm*6.1, *Cm*7.1, *Cm*8.1, *Cm*9.1, *Cm*10.1), QRLs for resistance to bacterial canker (*Clavibacter michiganensis*) (Sandbrink et al. 1995); *Fen*, sensitivity to fenthion (Martin et al. 1994); *Fr*1, resistance to Fusarium wilt (*Fusarium oxysporum* f.sp. *radicis-lycopersici*) (Vakalounakis et al. 1997); *Hero*, resistance to potato cyst nematode (*Globodera rostochiensis*) (Ganal et al. 1995); *I* (*I*-1, *I*-2, *I*-3), resistance to different races of Fusarium wilt (*Fusarium oxysporum* f.sp. *lycopersici*) (Ori et al. 1997); *LB*-1 and *LB*-2, QRLs for resistance to tomato late blight (*Phytophthora infestans*) (Frary et al. 1998); *Lv*, resistance to powdery mildew (*Leveillula taurica*) (Chunwongse et al. 1994); *Mi* and *Mi*-3, resistance to root knot nematodes (*Meloidogyne* spp.) (Yaghoobi et al. 1995; Veremis et al. 1999); *N18*, resistance to tobacco mosaic virus (Whitham et al. 1994); *Ol*-1, resistance to powdery mildew (*Oidium lycopersicum*) (van-der-Beek et al. 1994); *Ph* (*Ph*-1, *Ph*-2, *Ph*-3), resistance to late blight (*Phytophthora infestans*) in tomato (Chunwongse et al. 1998); *Pto* and *Prf*, resistance to bacterial speck (*Pseudomonas syringae* pv. *tomato*) (Martin et al. 1993a; Salmeron et al. 1996); *Py*-1, resistance to corky root rot (*Pyrenochaeta*
lycopersici) (Doganlar et al. 1998); \textit{rx} (\textit{rx-1}, \textit{rx-2}, \textit{rx-3}), resistance to bacterial spot (\textit{Xanthomonas campestris}) (Yu et al. 1995); \textit{Sm}, resistance to \textit{Stemphilium} (Behare et al. 1991); \textit{Sw-5}, resistance to tomato spotted wilt virus (Brommenschchenkel and Tanksley 1997); \textit{Tm-1} and \textit{Tm-2a}, resistance to tobacco mosaic virus (Young and Tanksley 1989); \textit{Ty-1} and \textit{Ty-2}, resistance to yellow leaf curl virus (Zamir et al. 1994; Hanson et al. 2000); \textit{Ve}, resistance to \textit{Verticillium dahliae} (Diwan et al. 1999).
Figure 1
Figure 2. A linkage map of tomato displaying chromosomal locations of resistance gene analogs (RGAs) and QTLs for Early Blight (EB) resistance. The map includes 141 RFLP markers (shown in black font) and 73 RGAs (PAGE-polymorphic and PAGE-monomorphic RGAs shown in red and blue font respectively). The names of the markers and map distances between them are shown at the right of the chromosomes. The LOD (log 10 of the odd ratio) plots at the left of the chromosomes indicate the most likely positions of QTLs for EB resistance identified in the BC₁ (black curves), BC₁S₁-1999 (red curves) and BC₁S₁-2000 (blue curves) populations. The LOD plots were derived based on single interval mapping (SIM) using the QGENE computer program. The height of each LOD curve indicates the strength of the evidence for the presence of a QTL at each location. The dotted vertical lines indicate a LOD value of 2.4, a threshold value that the LOD score must cross to allow the presence of a QTL to be inferred. The maximum-likelihood position of the QTL is the highest point on the curve, which is shown at the left of the chromosomes together with the name(s) of the QTL(s). The highest LOD score obtained for each chromosome is also shown on the Y axis.
Figure 3. Permutation analysis for the BC$_1$S$_1$-1999 disease evaluation. Genotypic data was shuffled one thousand times for the Early Blight resistance trait at each marker locus and the LOD scores were recalculated for every iteration. The chromosomes are numbered and displayed in a horizontal linear order at the top of the plot. The black curves and peaks in the plot represent the original calculated LOD scored based on the BC$_1$S$_1$-1999 disease evaluation (unshuffled data). The colored dots represent the recalculated 95$^{\text{th}}$ (blue), 99$^{\text{th}}$ (green) and 100$^{\text{th}}$ (red) percentile for each LOD set of shuffles.
Chapter 3

Sequence analysis of *R*-gene-derived markers

Abstract

In the search for structural or sequential similarities, newly identified sequences are continuously compared with *R*-gene and *R*-protein accessions that are stored in public databases. Potential similarities would help determine the identity, taxonomic and evolutionary relationships and putative roles which any new gene sequence may have in disease resistance.

In this study, DNA markers amplified by *R*-gene-derived primers were previously mapped in an RFLP-anchored map of tomato constructed based on a BC₁ population of a cross between a *Lycopersicon esculentum* and a *L. hirsutum* lines. The BC₁ and BC₁S₁ (self-progeny of BC₁) populations were previously evaluated for early blight (EB) resistance and QTLs were identified. In the present study, to determine their identity, RGA markers (particularly those with interesting co-localization with *R*-genes and EB resistance QTLs) were cloned, sequenced and compared with accessions in the GenBank. Sequence comparisons and homology searches revealed that an unexpected low percentage of the mapped markers were identical to the previously identified tomato *R*-genes and RGAs. Very few markers were genuine RGAs, unlike what was initially expected based on the map coincidence with *R*-genes and EB-resistance QTLs.
Furthermore, most of the markers that exhibited clustering in particular chromosomal regions were not significantly homologous to any annotated \( R \)-gene sequences, although some significant homologies to non-\( R \)-genes were observed.

The overall results demonstrate that most of the markers produced by \( R \)-gene-derived degenerate primers are not \( R \)-gene analogs, unlike what has been generally assumed and reported in previous investigations. However, the \( R \)-gene-derived primer approach appears to be a potentially valuable molecular breeding tool for the production of markers that might be used to saturate chromosomal regions where QTLs for disease resistance reside.

**Introduction**

The most important, and rather unexpected discovery in the “gene-for-gene” mechanism was the similar structural features shared by the majority of cloned resistance genes (\( R \)-genes) in spite of the diversity of pathogens against which they act (Bent 1996; Hammond-Kosack and Jones 1997). Thus, based on these structural similarities, all of the characterized \( R \)-genes involved in the interaction with the pathogen \( avr \)-gene products have been assigned to six different gene classes, as described in the previous chapter.

To assist in the discovery of novel \( R \)-gene specificities, \( R \)-gene analogs (RGAs) from different plant species have been thoroughly investigated by PCR-amplification of genomic DNA using degenerate primers corresponding to conserved domains of the original \( R \)-proteins (Hammond-Kosack and Jones 1997; Hulbert et al. 2001). The sequence of the \( R \)-genes is the ultimate determining factor to elucidate their map location, architecture, origin and evolution. For example, a study conducted by Pan et al. (2000) demonstrated that only the sequences of NBS domains across monocots and dicots were capable of detecting the independent generation and diversification of two smaller domains (coiled-coil and leucine-zipper, Chapter 2), involved in the specificity of pathogen recognition. Furthermore, despite the generation of these two divergent domains in the evolution of the plant kingdom, the NBS domain can faithfully predict the
overall structure of an \( R \)-gene.

An RGA is any amplified fragment whose predicted peptide contains the conserved motifs corresponding to a primer’s annealing sites as well as additional intermediate regions needed for pathogen ligands recognition and for defense signaling detection and initiation (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996). For example, within the NBS of the *Arabidopsis* genes *RPS2* and *RPM1* there are three domains (kinase 1a P-loop, kinase 2 and kinase 3a) embedded in approximately 250 amino acids of their amino termini (Hammond-Kosack and Jones 1997). Those conserved domains aid in the alignment of sequences for phylogenetic analyses, in the classification of newly isolated sequences based on their NBS region, and in predicting whether they belong to TIR or non-TIR classes, as described in Chapter 2 (Pan et al. 2000). Thus, the sequencing of RGAs serves to assign them a putative role by comparing and predicting their homology to characterized \( R \)-genes. Because structurally related \( R \)-genes, containing for example the NBS-LRR domains, control race-specific resistance to bacteria, fungi and viruses, it is impossible to predict the role of a new RGA to a particular pathogen only by deducing its nucleotide sequence.

PCR amplification of plant genomic DNA with degenerate primers is expected to result in the amplification of new \( R \)-genes. In plant breeding studies, however, the amplification may result in two types of PCR products for genetic mapping, RFLP probes and PCR markers. RFLP-based RGAs have been the most commonly used markers in the identification and mapping of \( R \)-gene and \( R \)-gene families in more than 15 plant species (Pan et al. 2000; Hulbert et al. 2001). Because these markers are cloned and sequenced beforehand, one can be assured that only actual \( R \)-genes or new analog sequences are used to search for polymorphisms among several parental lines and progenies prior to genetic mapping. Sequencing of RGAs has been extremely important in revealing the nature of the amplification products, to select for particular clones for future mapping, and to elucidate their complex gene family organization (Yu et al. 1996; Leister et al. 1998; Seah et al. 1998; Mago et al. 1999; Timmerman-Vaughan et al. 2000).

RGAs displayed as dominant markers on polyacrylamide gels have been shown to co-
localize with QTLs conferring resistance to blast, bacterial blight, sheath blight and brown plant-hopper in rice (Ramalingam et al. 2003), stripe rust in wheat (Chen et al. 1998) and early blight in tomato (Foolad et al. 2002). If the mapped RGA markers were genuine $R$-genes or $R$-gene analogs, then the separation of the total amplification product of degenerate primers on polyacrylamide gels would be a significant shortcut to $R$-gene discovery. In this scenario, RFLP hybridization would no longer be necessary, making RGA mapping a more efficient and rapid operation in plant genetics and breeding research. The above-mentioned reports have assumed that the mapped markers were RGAs based on their coincidence with QTLs for disease resistance and their relative proximity to other $R$-genes. However, confirmation that the mapped markers were genuine RGAs by comparing their nucleotide sequences or their deduced peptides has not been performed.

The sequencing of RGAs reported in this chapter complements the previous chapter on mapping RGAs and comparing their chromosomal locations with $R$-genes and QTLs, which were previously mapped on tomato chromosomes. Thus, the purpose of this study was to reveal the nature of the segregating DNA-fragments on polyacrylamide gel electrophoresis (PAGE-polymorphic), and to determine if their genomic locations clustering on QTLs for EB resistance were correlated with their possible homology to $R$-genes. Furthermore, in order to exhaust the search for potential $R$-gene markers, DNA-fragments that appeared monomorphic on polyacrylamide gels (PAGE-monomorphic) and that were converted to RFLP probes for genetic mapping (Chapter 2) were also sequenced and compared, in terms of their utility, with PAGE-polymorphic markers.

**Material and Methods**

*Isolation, cloning and sequencing of RGA fragments.*

RGA fragments were directly isolated from dried polyacrylamide gels by fine-needle scratching of the surface and re-amplified by PCR (Stumm 1997). Amplified products were separated on 1.5% agarose gels and stained with ethidium bromide. Fragment
lengths were determined using standard DNA step-ladders, and their lengths were compared with the original lengths in polyacrylamide gels. Cloning was performed by excising the re-amplified fragments from the agarose gels, cleaning them with the QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA) and inserting them into pSTBlue-1 vectors (Novagen, Madison, WI). Five randomly selected colonies were chosen for each cloning event to verify by sequencing either the presence of a single band or several bands of approximately the same length. RGA fragments were sequenced using either an Applied Biosystems ABI PRISM 3700 DNA Analyzer (Carlson’s Lab) or an Applied Biosystems 3100 Genetic Analyzer (Nucleic Acid Facility), both at the Huck Institutes for the Life Sciences, The Pennsylvania State University. All sequences were submitted to the GenBank through the BankIt web submission tool (http://www.ncbi.nlm.nih.gov/BankIt/).

Sequence analysis and homology searches

Computer-aided homology search of the sequenced RGA fragments was performed using the current version of BLAST (Basic Local Alignment Search Tool, (Altschul et al. 1997)) algorithm with the GenBank database (http://www.ncbi.nlm.nih.gov/BLAST). In order to minimize the number of non-specific ‘hit’ results, the homology search was limited to the Viridiplantae subset of the GenBank, which includes only accessions from plant and algae species. The Jellyfish software, Version 3.0 (Labvelocity, San Francisco, CA) was used for RGA sequence storage, handling, translation, alignment and BLAST searches.

Sequence alignment and phylogenetic analysis

Marker sequences were aligned using the “multiple alignment generation” option of the CLUSTALW program (Thompson et al. 1994). UPGMA phylogenetic trees for the selected clusters in QTLs were generated using the latest version of the Molecular Evolutionary Genetics Analysis (MEGA2) software analysis (Kumar et al. 2001).
Results and Discussion

Cloning and sequencing of RGA markers

The PCR products of thirteen primer combinations were considered for direct marker scoring and genetic mapping as well as fragment isolation, cloning and sequencing. Some combinations of primers produced both PAGE-polymorphic and PAGE-monomorphic markers, whereas others produced only one type (Table 7). For example, primer pairs S1-AS3 and S2-AS1 produced respectively 10 and 13 prominent PAGE-monomorphic bands and no visible polymorphic bands.

A total of 29 PAGE-polymorphic RGAs, which were mapped in the tomato BC₁ population (Chapter 2, Figure 1, Table 8), were isolated from dried polyacrylamide gels, purified and cloned. After sequencing, all PAGE-polymorphic RGAs were aligned in multiple combinations to determine the presence of redundant or multiple-copy sequences, especially for those markers that were mapped in close proximity to each other. Only two markers, AN23-410 and AN23-380, produced by the same primer combination, were 86% identical to each other, however, they were mapped to different chromosomes, 1 and 10, respectively (Figure 1).

Overall, there were more PAGE-monomorphic than polymorphic bands. Ninety-five fragments that were visualized as monomorphic on PAGE were selected for isolation, cloning and sequencing. The selection was based primarily on the intensity of the bands and their reproducibility across different PCR and gel runs. After sequencing and performing multiple alignments, several copies of the same sequences were identified, and thus the 95 PAGE-monomorphic clones were re-arranged in 75 classes with less than 80% of nucleotide sequence identity (Table 9). Only one clone per class (the longest in length) was chosen for further RFLP hybridization analysis. Of the 75 clones, 37 exhibited polymorphism between *L. esculentum* and *L. hirsutum* and were used for genetic mapping in the BC₁ population (Chapter 2).

The nucleotide sequences of the PAGE-polymorphic and monomorphic markers were submitted to GenBank and their accession numbers are shown in Tables 8 and 9.
RGA sequence analysis

The search for nucleotide and protein homologies between the RGA markers (query sequences) and GenBank accessions (subject sequences) was performed by two of the BLAST programs, ‘blastx’ and ‘blastn’. ‘Blastx’ translates the query and uses it to search in the GenBank’s protein database, whereas ‘blastn’ performs a standard nucleotide-nucleotide search.

Depending on the uniformity of the query sequence data, different directions can be followed to interpret the BLAST search output. Most studies utilize the expected value (E-value) provided with the search results for each positive “hit”, which is a subject sequence with homology to the query. The E-value indicates the probability that the homology found between the query and the subject happened by chance. Because E is a statistical value specific for a particular ‘query-subject’ test, it cannot be used equally for all query sequences. For a heterogeneous collection of sequences, the E-value will fluctuate mainly due to three factors: the total number of subject sequences in the GenBank, the length of the query sequence, and the presence of repetitive segments in the query (Altschul et al. 1997). Since the number of subjects in the GenBank does not change dramatically over short periods of time and because no repetitive segments were found in the cloned RGAs, the only factor that could affect the E-value was the variable length of the queries, which in the present study ranged between 140 and 860 bp. Therefore, in order to secure an accurate estimation of the similarities between the queries and the GenBank’s subjects, the percentage of identity of each Blast search event was used to report how homologous the two sequences were. The percentage of identity states the number of nucleotides or peptides in one sequential length matrix that are identical for a ‘query-subject’ test, using the length of the query as the basal length. A very high homology between a subject and a query is reflected in a very low E-value, and usually (and depending on the length of both) in a very high percentage of identity (Altschul et al. 1997). In the present study, therefore, when both the percentage of identity and the E-value were highly significant, we could safely speculate that the query and the subject were part of the same gene or members of the same gene family.
PAGE-polymorphic markers

The primary technical advantages of PAGE-polymorphic markers are their reproducibility and the ability to directly score them as dominant segregating markers. In the present study, the interesting locations of some PAGE-polymorphic markers, coinciding with the locations of several tomato R-genes or resistance QTLs (Chapter 2), naturally lead one to hypothesize that they might actually be results of amplification of R-genes or defense response genes containing any of the NBS, LRR or STK domains. However, after BLAST searches, none of the 29 PAGE-polymorphic markers were truly homologous to R-genes or defense-response genes, and only 4 exhibited significant homology to non-R-gene sequences in the GenBank, ranging from 90 to 97% of nucleotide identity and from 71 to 94% of amino acid identity (Table 10). Nonetheless, some tentative indirect associations with disease resistance could be postulated based on the known functions of the GeneBank annotated homologous sequences and the map locations of the RGA markers. For example, with the exception of the markers AN23-380 and AN23-410, which were identical to potato mitochondrial genes (Table 10), three PAGE-polymorphic RGAs could be evaluated as resistance candidate gene markers, as described below.

First is the marker XLRR-140, which exhibited a significant homology to a tomato expansin precursor. Expansins are cell wall proteins that induce a pH-dependent wall extension and stress relaxation. Some expansin-like domains have been detected in virulence factors of plant pathogenic bacteria, which are believed to participate in the breakdown of plant cell walls (Cosgrove et al. 2002). Expansins are known to be differentially expressed during tomato fruit ripening. Although cell wall modification is an important early event during some plant defense responses (Hammond-Kosack and Jones 1996), no direct role of expansins has been demonstrated in defense response to pathogen attack. XLRR-140 was mapped to chromosome 3, where no QTL for EB resistance and no known tomato resistance genes have been located (Foolad et al. 2002).

Second is XLRR-360, which might be a stronger gene-marker candidate in plant defense responses because of its homology to a geraniol 10-hydroxylase, an enzyme
member of the diverse family of cytochrome P450s monooxygenases. Some P450s are enzymes involved in cell detoxification processes and others are important in the biosynthesis of the plant antimicrobial phytoalexins (Schuler and Werck-Reichhart 2003).

The third is XLRR-390, a marker with similarity to a small region of a bacterial artificial chromosome (BAC) from melon, which contains several NBS-LRR type of R-genes as well as reverse transcriptases, RNase H and an integrase (van Leeuwen et al. 2003).

Because of its homology to a member of the P450 family and its co-localization with a QTL, XLRR-360 can be considered the only positional and functional candidate among the PAGE-polymorphic markers obtained. Of the three PAGE-polymorphic markers homologous to candidate genes for disease resistance, only XLRR-360 mapped to a QTL (EBR9.1) for EB resistance. The potential role and effect of this putative P450 in EB resistance cannot be determined based on only the data presented here. However, regardless of its participation in disease resistance, this P450 and the marker cluster identified in EBR9.1 are linked to gene(s) with a considerable quantitative effect against Alternaria solani pathogenesis. Fine mapping of the EBR9.1 region is necessary for determination of the number and characterization of the gene(s) responsible for the large phenotypic effect of this chromosomal region (Table 3). The other PAGE-polymorphic and monomorphic markers in the cluster of XLRR-360 with no obvious significant homology might represent amplification products of non-annotated genes, or remains of non-functional NBS-LRR-type of genes. In either case, the results suggest that unknown and non-functional genes can still be detected and mapped with degenerate primers to their original locations and in proximity of the existing functional genes, some of which might be the ones producing the quantitative resistance phenotype.

As indicated before, none of the mapped PAGE-polymorphic markers were homologous to R-genes or RGAs in any plant species. This result was rather unexpected, since many PAGE-polymorphic RGAs were clustered around R-genes and QTLs for disease resistance, as reported by Chen et al. (1998), Zhang et al. (2002) and Foolad et al. (2002). Furthermore, Ramalingam et al. (2003) determined the locations of clusters of
PAGE-polymorphic markers with cDNA probes, whose sequences had been previously determined as *R*-genes within QTLs for resistance to three different rice diseases. In the present study, it was anticipated that PAGE-polymorphic RGAs were in fact fragments of *R*-gene alleles or QTLs conditioning EB resistance that were originated in the donor parent *L. hirsutum* and were absent in the recurrent parent *L. esculentum*. Nevertheless, it is important to note that in the studies mentioned above, PAGE-polymorphic markers were solely used as genetic markers and their sequences were not analyzed, and therefore the identities of the amplified bands were never determined.

Because of the lack of DNA and protein homology demonstrated in this study, and despite their co-localization with QTLs and *R*-genes, it can be concluded that the 29 PAGE-polymorphic markers isolated were not genuine RGAs. Furthermore, by comparing genetic mapping techniques, it is also possible that the majority of the markers similarly mapped by previous investigations in wheat and rice were not genuine RGAs either. These results demonstrate that it is imperative to determine by sequencing the identity and putative function of any mapped PAGE-polymorphic markers produced by degenerate *R*-gene primers, since map locations are not sufficient evidence to assume immediately that they are real RGAs.

**PAGE-monomorphic markers**

PAGE-monomorphic markers were not directly useful for genetic mapping, and, as described in chapter 2, it was necessary to convert them to RFLP probes after cloning to determine any positional information that they might contain. Of the 75 sequenced clones (Table 9) only 11 showed significant homologies to GenBank sequences, 9 that corresponded to actual *R*-genes and RGAs; the other 2 corresponded to non-*R*-gene sequences (Table 11).

The two PAGE-monomorphic markers homologous to non-*R*-gene sequences were NBS2-300 and A6-160. NBS2-300 mapped to tomato chromosome 7 and was significantly similar to a polygalacturonase-like protein from *Arabidopsis* (Table 11). Polygalacturonases are polysaccharide-degrading enzymes involved in the ripening of tomato fruit by degrading the plant cell wall. However, while no direct involvement on
resistance to pathogens has been reported, inhibitors of fungal polygalacturonases (polygalacturonase-inhibiting proteins, PGIPs) are part of plant defense mechanisms and possess LRR domains similar to those of the tomato Cf-2 R-gene (Leckie et al. 1999). The map position of NBS2-300 also coincided with other genes involved in tomato fruit ripening in chromosome 7 (Alpert et al. 1990). As no significant QTLs were found in this region, no positional or functional role in EB resistance could be assigned to this marker.

The second PAGE-monomorphic marker homologous to a non-R-gene sequence, A6-160, showed a very high similarity to a drought-responsive lipid transfer protein from L. pennellii (Trevino and Connell 1998) (Table 11). Some lipid transfer proteins are induced by pathogen infection or elicitor molecules, which make them good candidates for defense response gene mapping. However the probe A6-160 produced no polymorphism after Southern hybridization with digested genomic DNAs of the parental lines and thus it could not be mapped.

The most interesting subset of PAGE-monomorphic RGAs included SAS11-760, SAS21-520, S4-530, S3-440, SS1-530, SS14-520, SS19-530, SS20-860, NBS6-330, which were highly homologous to the NBS-LRR type of R-genes and RGA loci reported in tomato, potato, pepper and Arabidopsis (Table 11).

It is noteworthy that of all primer combinations used, only the ones designed based on the NBS domains of R-genes were efficient in the identification of these 9 genuine RGAs. Sequences containing either of the less abundant domains (i.e. LRR or STK) were not recovered by any of the other primer combinations. The apparent absence of amplified STK-like sequences contradicts the report by Van der Hoeven et al. (2002) that in the tomato genome, based on the analysis of over 120,000 ESTs and six BAC libraries (compared and correlated with the Arabidopsis genome), protein kinases represent the highest copy number gene family. Similarly, in Arabidopsis approximately 1% (~200 genes) of the total number of genes has been predicted to encode NBS-LRR-containing proteins, although not all are necessarily R-proteins (Meyers et al. 2003). Thus, since the number of primer combinations and the experimental conditions were the same for both domains, the number of STK-like sequences was expected to be as high as the number of
NBS-RGAs. This lack of PCR-amplified STK-like sequences with the primers used may indicate that most STKs in tomato cannot be identified by the same conserved Pto-gene domains that were used for the primers’ design (Leister et al. 1996; Feuillet et al. 1997; Chen et al. 1998). Probably, the domains present in protein kinase R-genes (e.g. the tomato Pto) used for the primer design are specific for pathogen recognition and are different from other STKs that are not involved in gene-for-gene interactions. Nevertheless, the active motifs of protein kinase R-genes are evolutionarily conserved and are required in the triggering of the signal transduction pathways (via phosphorylation) that activates plant cell defense systems upon recognition of the pathogen. These results, however, do not necessarily indicate the abundance in tomato of R-genes and RGAs containing NBS domains over the ones containing only LRR or STK.

Current R-gene discovery combines the use of degenerate primers with gene prediction from genome sequencing initiatives (Chapter 2). Therefore, in further studies of architecture and evolution of R-gene families, it may be imperative to determine whether the conserved protein domains that are used for primer design are the same as the ones that determine similarity between NBS or STK for phylogenetically grouping them into families.

The 9 genuine RGAs isolated in this research belong to the NBS-LRR class (as were identified with primers for the NBS domains only), although similarities to known R-genes and RGAs against bacteria, fungi and viruses are detected (Table 11). For example, translated nucleotide sequences indicated that the RGA SS20-860 encodes a typical NBS-LRR domain homologous to the tomato R-proteins Sw5-a and Sw5-b for resistance to the spotted wilt virus (Folkertsma et al. 1999) and to the R-protein Prf for resistance against bacterial speck (Salmeron et al. 1996). The RGA NBS6-330, on the other hand, was homologous to the Arabidopsis RPP13 gene controlling resistance to downy mildew (Bittner-Eddy et al. 2000). However, despite their significant homologies to R-genes, when used as hybridization RFLP probes, unfortunately neither SS20-860 nor NBS6-330 detected polymorphism between the parental lines of the BC1 population and thus could not be mapped. Nevertheless, in a similar candidate gene mapping project conducted in our laboratory with a different interspecific population of tomato, the hybridization of RGA NBS6-330 produced two segregating loci that mapped approximately 14 cM from
each other on the tomato chromosome 4. When comparing the relative positions of markers on chromosome 4 of both interspecific maps, the two NBS6-330 loci coincided with the five-marker cluster detected in this study (Figure 1). Furthermore, the two NBS6-330 loci mapped in close proximity to two R-genes against bacterial wilt (Bw-4) and potato cyst nematode (Hero), demonstrating once again the clustering of PAGE-polymorphic and monomorphic markers in the vicinity of R-gene loci.

The remaining RGAs, SAS11-760, SAS21-520, S4-530, S3-440, SS1-530, SS14-520 and SS19-530 from the subset above, detected loci that either mapped to QTLs for EB resistance or coincided with tomato R-genes (Table 11). Some of them are discussed below.

The RGA SAS11-760 detected two segregating loci, one of them was highly skewed and mapped to the SI region of chromosome 1 and the second mapped to the precise location of the R-gene I2C-2 complex in chromosome 11 and coincided with a QTL detected between the anchors TG36 and TG393 (Figure 1). The high homology (Table 11) and the map location indicate that the RGA SAS11-760 is most likely a member of the I2C gene family.

The RGA SS14-520 is identical to at least three R-genes that confer resistance to a virus and a nematode (Table 11). The SS14-520 probe detected a total of four loci, two of them mapped to chromosome 1 and the other two mapped to QTLs EBR9.2 and EBR12.2 and in proximity to R-genes Sw-5 and Lv in chromosomes 9 and 12, respectively. The sequence of SS14-520 revealed its highest similarities to tospovirus resistance proteins Sw5. One of the Sw5 family members was identified on the far end of tomato chromosome 9 along with QTL EBR9.2 (Figure 1, Figure 2). Interestingly, SS14-520 is also homologous to the potato rx gene against bacterial spot, which in tomato has been located on chromosome 1 proximal to the S locus. It is most likely that one of the loci detected by this probe that mapped to chromosome 1 (locus SS14-520.4, Figure 1) in fact hybridized to a member of the rx gene cluster. However, as explained in the previous chapter, the mapping results and any potential association with QTLs on chromosome 1 are affected by the skewed segregation caused by the phenotypic selections against SI.
The RGA SS19-530 was mapped to QTL EBR9.3 in chromosome 9 and is significantly homologous to some NBS-type of RGAs identified in an interspecific cross of *Solanum* species (Figure 1, Table 11). Interestingly, SS19-530 also mapped in proximity to the *Ve* locus that controls resistance to *Verticillium dahliae* (vascular wilt) and the *Cm9.1* quantitative resistance locus to *Clavibacter michiganensis* (bacterial canker). However, since *Ve* is a cytoplasmic protein receptor that has been reported to carry LRRs and a LZ but that lacks a NBS (Kawchuk et al. 2001), it is likely that SS19-530 is an amplified fragment of a NBS-type of *R*-gene in the *Cm9.1* locus or of another NBS-type of *R*-gene locus not yet identified in the nearby chromosomal region.

The RGAs S3-440 and SS1-530 are 64% identical to each other and homologous to the first *R*-gene families identified in potato by Leister et al. (1996). Despite their identity, S3-440 and SS1-530 detected loci in different chromosomes, mapping to the short arm of chromosome 10 and to QTL EBR12.2 in chromosome 12, respectively. The locations of RGAs S3-440 and SS1-530 also coincided with a location in a different tomato map recently developed in our laboratory of an interspecific cross between *L. esculentum* and *L. pimpinellifolium* (unpublished data).

In the examples mentioned above, there is a correlation between the homology of the markers and their chromosomal positions with respect to QTLs for EB resistance and previously identified *R*-genes or NBS-LRR gene analogs. In some cases the percentage of identity is so high that one can easily conclude identification and mapping of an *R*-gene family member; for example, SAS11-760 with 99% and 94% of sequence identity at the nucleotide and peptide level. However, in other cases the percentage of identity between the marker and its homolog is much lower and thus, it is difficult to predict whether the marker is in fact a member of the *R*-gene family in whose vicinity it was located. A few scenarios can be explored based on the information the genetic map and the sequences provided. RGA SS14-520, for example, is approximately 86% identical to the Sw5 *R*-gene, maps in its vicinity and co-localizes with QTL EBR9.2 on chromosome 9. It is possible that SS14-520 is a novel *L. hirsutum* allele of the Sw5 gene and only 86% identical to the original Sw5 isolated from *L. esculentum* by Folkertsma et al. (1999). However, it should also be considered that SS14-520 might be one member of a *L.*
*hirsutum* Sw5 gene family, or product of a duplication event sharing a common ancestor sequence with Sw5 in a nearby chromosomal location that diverged and accumulated nucleotide variation over the years. Since the genetic mapping of SS14-520 relied only on DNA hybridization, it is not possible at this point to establish whether this RGA probe corresponds to a fully functional *R*-gene.

Because of the absence of a gene-for-gene interaction in the tomato EB resistance system, and without an *R*-gene functional analysis, the quantitative resistance to EB cannot be attributed to the effect of functional *R*-genes or RGAs. The genetic locations of either PAGE-polymorphic or monomorphic markers at this point only elucidate the markers’ linkages to minor genes with quantitative effects on resistance (e.g. genes involved in defense response mechanisms). Functional genomics studies, for example the detection of any change in the resistance phenotype after the specific inactivation or silencing of an *R*-gene, have the potential to confirm the hypothesis that *R*-genes explain a significant portion of the quantitative variation for resistance, even to non-specific virulent pathogens (Leonards-Schippers et al. 1994; Michelmore 1996). This study supports this hypothesis only by the genetic locations of markers and QTLs, but does not indicate whether the QTLs detected are affected by the roles of the linked *R*-genes themselves.

*Associations between PAGE-monomorphic and PAGE-polymorphic markers*

QTLs *EBR5.2* and *EBR9.1* are two examples of the coincidence of PAGE-monomorphic and PAGE-polymorphic markers with QTLs for EB resistance and relative proximities to previously mapped tomato *R*-genes.

The QTL *EBR5.2* covers most of the lower half of chromosome 5, which includes the region where the two structurally different *Pto* and *Prf* genes (STK and NBS-LRR respectively) for resistance to bacterial speck (*Pseudomonas syringae* pv. *tomato*) have been located (Martin et al. 1993b; Salmeron et al. 1996). Furthermore, recent QTL mapping data produced in our laboratory have indicated that *EBR5.2* might be the only QTL for EB resistance shared between two *L. hirsutum* and *L. pimpinellifolium* accessions. This shared QTL suggests that the two tomato species conserve in that
particular QTL region the most beneficial alleles for the expression of the resistance phenotype, and that it may be possible to use the 7 linked RGA marker loci for its detection (Figure 2).

The QTL $EBR9.1$, on the other hand, harbors 4 PAGE-polymorphic and 7 PAGE-monomorphic markers in an 18.7-cM region and is in relative close proximity to the $Tm-2a$ and $Fr1$ R-genes that control resistance to tobacco mosaic virus and fusarium wilt, respectively (Figure 1, Figure 2). $EBR9.1$ can explain up to 25% of the total phenotypic variation of the resistance to EB.

In order to investigate any possible phylogenetic relationships between the PAGE-polymorphic and monomorphic marker sequences in the QTLs $EBR5.2$ and $EBR9.1$, a standard Jukes-Cantor UPGMA (Unweighted Pair Group Method with Arithmetic Mean) (Sneath and Snokal 1973) was implemented using the aligned nucleotide sequences of the markers (Figure 3). UPGMA is the simplest and most straightforward method for the construction of trees (taxonomic phenograms) that reflect, for example, the genetic similarities between operational taxonomic units (OTUs, i.e. RGA marker sequences compared). Briefly, UPGMA employs a sequential algorithm that clusters OTUs in order of similarity based on their identified local homology, and constructs a tree in a stepwise manner. In the construction of the tree, two highly similar OTUs are first determined and then treated as a new single 'composite' OTU. Subsequently from the new rearranged group of OTUs (composite and simple), the pair with the second highest similarity is identified and clustered. The grouping and pairing steps continue until only two OTUs are left. The UPGMA’s algorithm assumes that the two most closely related OTUs are more similar to each other than they are to any other (Sneath and Snokal 1973). The Jukes-Cantor correction, on the other hand, assumes that during the construction of the tree the rate of nucleotide substitution (which determines the differences between the sequences being compared) is the same for all pairs of the four nucleotide sites (Jukes and Cantor 1969).

The UPGMA tree for the PAGE-polymorphic and monomorphic markers of QTL $EBR5.2$ (Figure 3A) demonstrates the lack of sequence similarities and the absence of a
very immediate common ancestor. The pairing and branching of the markers in the tree, along with the Blast homology search, are weak evidence to speculate that the cluster in the QTL EBR5.2 may consist of a family of RGAs.

The UPGMA tree for the QTL EBR9.1 appeared very similar to the one for EBR5.2, except that it grouped in one branch the sequences of the PAGE-polymorphic markers RLRR-130, RLRR-140 and the PAGE-monomorphic marker SAS18-130 (Figure 3B). This branch closely grouping the three markers suggests that they might have originated very recently from a common ancestor; which would not be surprising since these three markers are only 70% identical when aligned and compared (RLRR-130 and RLRR-140 alone are 96.6% identical). The genetic map indicates that the distance between RLRR-130 and RLRR-140 is very short, probably too short to detect a recombination event between both markers for the population size used. The distance from RLRR-130 and RLRR-140 to SAS18-130, however, is approximately 8 cM. By observing the map distances and the UPGMA tree it can be proposed that the primer pairs RLRRforward–RLRRreverse and S1–AS3 amplify at least three repetitive and linked chromosomes regions that might have arisen very recently by gene duplication events.

Without the complete sequence of tomato chromosomes 5 and 9, the involvement of these markers in the phenotypic effects of the QTLs EBR5.2 and EBR9.1 remains undetermined. The presence of duplicated marker sequences that are associated with disease resistance but that are not homologous to any currently known plant gene, might imply the existence of ancient clusters of R-genes (or any identical NBS-, LRR-containing genes) that once controlled resistance to pathogens now extinct. This could be explained, for example, by assuming that if a host plant, carrying an active and duplicating R-gene family, is no longer challenged by a specific biotic stress (i.e. the pathogen carrying a detectable avr signal) then the lack of selection forces acting to retain the functional genes results in the accumulation of deleterious mutations and the final silencing of the genes. The non-functional gene sequences will remain clustered, with minimum genetic recombination events between them and yet linked to functional genes responsive to common pathogenesis or stress signaling pathways responsible for the phenotypic effects of the QTLs.
The similar clusters of PAGE-polymorphic and monomorphic markers observed in chromosomes 1, 2, 4 and 10 do not co-localize with any QTLs for EB resistance. Because their sequences do not provide any evidence about their possible R-gene origin, it can be speculated that they originated by similar duplication and subsequent inactivation of NBS-LRR genes with no known involvement in disease resistance. However, it is important to consider other explanations for the presence of markers with neither genetic associations to QTLs and R-genomes nor noticeable similarities to current NBS-LRR type of genes. As our knowledge of plant genomics increases, such explanations may include the amplification of mere pseudogenes, silenced redundant genes or random DNA fragments linked by chance to other markers sharing the same origin.

**Evaluation of the degenerate primer approach**

In this study, the major difference between the PAGE-polymorphic and PAGE-monomorphic markers was the information contained in their nucleotide sequences, which ultimately determines whether what is being amplified by the degenerate primers is an actual R-gene or an analog. PAGE-polymorphic markers were confirmed as straightforward, reproducible and reliable for use in genetic mapping; in particular, they might be valuable in the saturation of key chromosomal regions where QTLs are located. However, in this study PAGE-polymorphic markers did not provide any R-gene information in their sequences, and although they showed linkages to R-gene and R-loci, they were not actual R-genes or analogs. Additional primer combinations are needed in order to provide a better estimation of the abundance and arrangement of PAGE-polymorphic markers. Thus, it is necessary to constantly monitor the emergence of novel R-gene domains and to use them in the construction of new primers that target all analog sequences and exhaust the identification of gene families. By producing a sufficient amount of segregating markers perhaps one could distinguish a clustering pattern or a preference for a chromosomal region that may or may not contain an R-gene. Alternatively, the larger the number of markers that are mapped and cloned, the higher the probability of finding interesting candidate genes that can be amplified by these primers, such as the geraniol 10-hydroxylase (XLRR-360), that are not involved in pathogen recognition by a gene-for-gene interaction, but that possess the primer
annealing sites and perhaps a similar functional protein domain. For this purpose, it would also be interesting in mapping projects to complement the use of degenerate primers that are more likely to provide QTL-linked markers and candidate gene sequences with defense-response gene mapping. For example combining the primer pair XLRR with other known members of the P450 family as DNA markers might confirm the role of members of this gene family on the expression of EB resistance.

In this study, PAGE-monomorphic markers were as likely as the PAGE-polymorphic to map and cluster around QTLs and R-genes. However, in order to reach this mapping step it was necessary to isolate, clone and sequence the PAGE-monomorphic fragments for DNA hybridizations. Despite the additional time and effort invested in mapping, it seems that PAGE-monomorphic markers are more likely to be homologous to real R-genes and to hybridize to other analogs or to the target R-gene itself. Because molecular marker polymorphisms are cross-specific, assessing the degenerate primer approach in different mapping populations will be necessary to conclude whether there is a correlation between PAGE-monomorphic markers an a greater homology to R-genes and R-gene analogs.

As suggested in Chapter 2, the use of both types of markers as specific sequence characterized amplified regions (SCAR, Paran and Michelmore (1993)) in marker-assisted selection (MAS) is feasible, since the full length of the marker is already available and the fragments have appropriate sizes for specific primer hybridization windows. The main requisite for the selection of a SCAR is the tight linkage between a marker and a QTL, regardless of the information its sequence may contain. The present study has provided a total of 10 PAGE-polymorphic markers and 13 PAGE-monomorphic clones that are capable to detect the QTLs EBR5.2, EBR9.1, EBR9.2, EBR9.3, EBR11.3 and EBR12.2, and whose sequence can be used for developing SCAR markers (Figure 2).

Many R-gene and RGA sequences are currently available on public electronic databases, and the chromosomal locations of most of them can be found in the literature. In a breeding program, including those functional marker sequences during the
construction of backbone maps will contribute significantly to identify potential sources of QTLs. Subsequently, the production of PAGE-polymorphic markers would quickly saturate chromosomal regions where R-genes are likely to map. Based on the results of this study, it is important to suggest that, from an applied molecular breeding perspective, further analyses of PAGE-monomorphic bands are not necessary. Working with PAGE-monomorphic markers requires additional time and effort and the final outcome would be the isolation of the same RGAs whose sequences and map locations could be more easily obtained elsewhere. The future use of degenerate primers in breeding programs may rely upon the necessity to saturate chromosomal regions for fine mapping and precise locations of QTL.

Alternative approaches must be considered before implementing the degenerate primer and PAGE method for R-gene mapping. These approaches will complement R-genes and RGAs already positioned in a genetic map by any other less demanding means. For example, R-genes and RGAs can be rapidly mapped as expressed sequence tags, RFLP probes, or as cleaved amplified polymorphic sequences (CAPS, (Quint et al. 2001). Degenerate primers in combination with the PAGE approach should be considered as a supplementary choice if additional map saturation is needed, but not as a first choice for the identification and mapping of new R-genes.

References


like kinase gene encoded at the Lr10 disease resistance locus of wheat. Plant J 11: 45-52.

Evolutionary aspects of ‘chloroplast-like' trnN and trnH expression in higher-plant 

296.

Folkertsma, R.T., Spassova, M.I., Prins, M., Stevens, M.R., Hille, J., and Goldbach, R.W. 
1999. Construction of a bacterial artificial chromosome (BAC) library of 
Lycopersicon esculentum cv. Stevens and its application to physically map the 

reaction (PCR) in plant genome analysis. In Fundamental Methods of Plant Cell, 
Tissue and Organ Culture and Laboratory Operations. (eds. O.L. Gamborg, and 


field, greenhouse, and detached-leaflet evaluations of tomato germplasm for early 

QTLs for early blight (Alternaria solani) resistance in tomato using backcross 


Kumar, S., Tamura, K., Jakobsen, I., and Nei, M. 2001. MEGA2: Molecular Evolutionary Genetics Analysis, 2.1 ed. Arizona State University, Tempe, AZ.


Table 7. Final count of amplification products obtained with degenerate RGA-primers. Only amplification products (bands) of clear signal on polyacrylamide gel electrophoresis are considered.

*Monomorphic bands (75) converted into probes for RFLP hybridization.

**Monomorphic bands converted into RFLP probes showing useful polymorphism for genetic mapping (37)

<table>
<thead>
<tr>
<th>Primer pair</th>
<th>PAGE-polymorphic</th>
<th>PAGE-monomorphic*</th>
<th>PAGE-monomorphic for genetic mapping**</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLRRfor-CLRRrev</td>
<td>3</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>RLRRfor-RLRRrev</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>XLRRfor-XLRRrev</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NBS-F1 - NBS-R1</td>
<td>0</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>ANo. 2 - ANo. 3</td>
<td>6</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>S1-AS1</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>S1-AS3</td>
<td>0</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>S2-AS1</td>
<td>0</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>S2-AS3</td>
<td>1</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>PtoKin1-PtoKin2</td>
<td>6</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>PtoKin3-PtoKin4</td>
<td>2</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>PtoFen-S - PtoFen-AS</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>LRKfor - LRKrev</td>
<td>0</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29</strong></td>
<td><strong>75</strong></td>
<td><strong>37</strong></td>
</tr>
</tbody>
</table>
Table 8. Twenty-nine PAGE-polymorphic markers genetically mapped in the tomato BC$_1$ population map. Each fragment was isolated from dried gels, sequenced and posted on the GenBank. The accession numbers are indicated next to the markers in parentheses.

*CLRR-950, PK12-600 and PK12-560 failed the isolation and cloning steps, no sequence is available for them

<table>
<thead>
<tr>
<th>Primer Pair Combination</th>
<th>PAGE-polymorphic Markers*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLRRfor – CLRRrev</td>
<td>CLRR-950</td>
</tr>
<tr>
<td></td>
<td>CLRR-490 (AF534344)</td>
</tr>
<tr>
<td></td>
<td>CLRR-450 (AF534343)</td>
</tr>
<tr>
<td>RLRRfor – RLRRrev</td>
<td>RLRR-220 (AF534353)</td>
</tr>
<tr>
<td></td>
<td>RLRR-140 (AF534352)</td>
</tr>
<tr>
<td></td>
<td>RLRR-130 (AF534351)</td>
</tr>
<tr>
<td>XLRRfor – XLRRrev</td>
<td>XLRR-390 (AF534362)</td>
</tr>
<tr>
<td></td>
<td>XLRR-380 (AF534361)</td>
</tr>
<tr>
<td></td>
<td>XLRR-370 (AF534360)</td>
</tr>
<tr>
<td></td>
<td>XLRR-360 (AF534359)</td>
</tr>
<tr>
<td></td>
<td>XLRR-140 (AF534358)</td>
</tr>
<tr>
<td>ANo.2 – ANo.3</td>
<td>AN23-410 (AF534342)</td>
</tr>
<tr>
<td></td>
<td>AN23-390 (AF534341)</td>
</tr>
<tr>
<td></td>
<td>AN23-380 (AF534340)</td>
</tr>
<tr>
<td></td>
<td>AN23-230 (AF534339)</td>
</tr>
<tr>
<td></td>
<td>AN23-220 (AF534338)</td>
</tr>
<tr>
<td></td>
<td>AN23-210 (AF534337)</td>
</tr>
<tr>
<td>S1 – AS1</td>
<td>S11-210 (AF534356)</td>
</tr>
<tr>
<td></td>
<td>S11-200 (AF534355)</td>
</tr>
<tr>
<td></td>
<td>S11-150 (AF534354)</td>
</tr>
<tr>
<td>S2 – AS3</td>
<td>S23-200 (AF534357)</td>
</tr>
<tr>
<td>PtoKin1 – PtoKin2</td>
<td>PK12-700 (AF534348)</td>
</tr>
<tr>
<td></td>
<td>PK12-600</td>
</tr>
<tr>
<td></td>
<td>PK12-560</td>
</tr>
<tr>
<td></td>
<td>PK12-340 (AF534347)</td>
</tr>
<tr>
<td></td>
<td>PK12-320 (AF534346)</td>
</tr>
<tr>
<td></td>
<td>PK12-300 (AF534345)</td>
</tr>
<tr>
<td>PtoKin3 – PtoKin4</td>
<td>PK34-300 (AF534350)</td>
</tr>
<tr>
<td></td>
<td>PK34-210 (AF534349)</td>
</tr>
</tbody>
</table>
Table 9. Seventy-five PAGE-monomorphic markers isolated from dried polyacrylamide gels. All fragments were sequenced and posted on the GenBank. The accession numbers are indicated next to the markers in parentheses.

*Markers in bold font (37) displayed useful polymorphism when hybridized to blots containing restriction-digested DNA of the BC\textsubscript{1} population parents.

<table>
<thead>
<tr>
<th>Primer Pair Combination</th>
<th>PAGE-monomorphic Markers*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLRRfor – CLRRrev</td>
<td>C1-290 (AF534264)</td>
</tr>
<tr>
<td></td>
<td>C2-170 (AY428540)</td>
</tr>
<tr>
<td></td>
<td>C2-560 (AF534265)</td>
</tr>
<tr>
<td></td>
<td>C3-180 (AY428541)</td>
</tr>
<tr>
<td></td>
<td>C4-490 (AF534266)</td>
</tr>
<tr>
<td></td>
<td>C5-230 (AF534267)</td>
</tr>
<tr>
<td></td>
<td>C6-150 (AF534268)</td>
</tr>
<tr>
<td>NBS-F1 – NBS-R1</td>
<td>NBS1-450 (AF534266)</td>
</tr>
<tr>
<td></td>
<td>NBS2-300 (AF534270)</td>
</tr>
<tr>
<td></td>
<td>NBS3-210 (AF534271)</td>
</tr>
<tr>
<td></td>
<td>NBS4-300 (AF534272)</td>
</tr>
<tr>
<td></td>
<td>NBS6-330 (AF534273)</td>
</tr>
<tr>
<td></td>
<td>NBS7-170 (AF534274)</td>
</tr>
<tr>
<td></td>
<td>NBS8-300 (AF534275)</td>
</tr>
<tr>
<td></td>
<td>A1-150 (AF534276)</td>
</tr>
<tr>
<td></td>
<td>A2-190 (AF534277)</td>
</tr>
<tr>
<td>ANo.2 – ANo.3</td>
<td>A5-160 (AF534278)</td>
</tr>
<tr>
<td></td>
<td>A5-330 (AF534279)</td>
</tr>
<tr>
<td></td>
<td>A6-160 (AF534280)</td>
</tr>
<tr>
<td></td>
<td>A7-460 (AF534281)</td>
</tr>
<tr>
<td></td>
<td>SA2-300 (AF534282)</td>
</tr>
<tr>
<td></td>
<td>SA3-370 (AF534283)</td>
</tr>
<tr>
<td>S1 – AS1</td>
<td>SA4-100 (AF534284)</td>
</tr>
<tr>
<td></td>
<td>SA5-500 (AF534285)</td>
</tr>
<tr>
<td></td>
<td>SA6-220 (AF534286)</td>
</tr>
<tr>
<td></td>
<td>S1-160 (AF534297)</td>
</tr>
<tr>
<td></td>
<td>S2-410 (AF534298)</td>
</tr>
<tr>
<td>S2 – AS3</td>
<td>S3-440 (AF534299)</td>
</tr>
<tr>
<td></td>
<td>S4-530 (AF534300)</td>
</tr>
<tr>
<td></td>
<td>S5-240 (AF534301)</td>
</tr>
<tr>
<td></td>
<td>S6-260 (AF534302)</td>
</tr>
</tbody>
</table>
Table 9. (Continuation)

<table>
<thead>
<tr>
<th>S1 – AS3</th>
<th>S2 – AS1</th>
<th>LRKfor-LRKrev</th>
<th>PtoFen-S – PtoFen-AS</th>
<th>PtoKin1 – PtoKin2</th>
<th>PtoKin3 – PtoKin4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAS5-250 (AF534295)</td>
<td>SS1-530 (AF534310)</td>
<td>LRK-500 (AF534319)</td>
<td>PF3-220 (AF534324)</td>
<td>P2-220 (AF534331)</td>
<td></td>
</tr>
<tr>
<td>SAS6-540 (AF534296)</td>
<td>SS2-510 (AF534311)</td>
<td>LRK-450 (AF534322)</td>
<td>PF6-220 (AF534326)</td>
<td>K1-260 (AF534327)</td>
<td></td>
</tr>
<tr>
<td>SAS11-760 (AF534287)</td>
<td>SS5-380 (AF534314)</td>
<td>LRK-430 (AF53421)</td>
<td>P3-310 (AF534332)</td>
<td>K2-300 (AF534328)</td>
<td></td>
</tr>
<tr>
<td>SAS12-240 (AF534288)</td>
<td>SS7-290 (AF534315)</td>
<td>LRK-400 (AF534317)</td>
<td>P4-500 (AF534333)</td>
<td>K5-310 (AF534329)</td>
<td></td>
</tr>
<tr>
<td>SAS13-510 (AF534289)</td>
<td>SS1-270 (AF534303)</td>
<td>LRK-350 (AF534316)</td>
<td>P4-600 (AF534334)</td>
<td>K6-160 (AF534330)</td>
<td></td>
</tr>
<tr>
<td>SAS18-130 (AF534291)</td>
<td>SS11-270 (AF534304)</td>
<td>LRK-250 (AF534320)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAS21-520 (AF534292)</td>
<td>SS14-520 (AF534305)</td>
<td>LRK-210 (AF534318)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAS23-310 (AF534293)</td>
<td>SS15-220 (AF534306)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAS26-320 (AF534294)</td>
<td>SS16-550 (AF534307)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SS18-410 (AF534308)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SS19-530 (AF534309)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SS20-860 (AF534312)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SS21-800 (AF534313)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10. PAGE-polymorphic markers and their most significant homologies.

*NS, no significant homologies found using that Blast program.

<table>
<thead>
<tr>
<th>PAGE-polymorphic marker</th>
<th>Most Significant Homology</th>
<th>Highest Percentage of Identity*</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN23-380</td>
<td>Solanum tuberosum mitochondrial trnC, trnN1, trnY, nad2 genes</td>
<td>94%</td>
<td>(Fey et al. 1997)</td>
</tr>
<tr>
<td>AN23-410</td>
<td>Solanum tuberosum mitochondrial trnC, trnN1, trnY, nad2 genes</td>
<td>90%</td>
<td>(Fey et al. 1997)</td>
</tr>
<tr>
<td>XLRR-140</td>
<td>L. esculentum expansin precursor (Exp3)</td>
<td>97%</td>
<td>(Brummell et al. 1999)</td>
</tr>
<tr>
<td>XLRR-360</td>
<td>Geraniol 10-hydroxylase (a cytochrome P450 enzyme)</td>
<td>NS</td>
<td>(Collu et al. 2001)</td>
</tr>
<tr>
<td>XLRR-390</td>
<td>Reverse transcriptase, RNase H and integrase</td>
<td>NS</td>
<td>(van Leeuwen et al. 2003)</td>
</tr>
</tbody>
</table>
Table 11. PAGE-monomorphic markers with their most significant homologies to R-genes, RGAs and other plant genes.

*NS, no significant homologies found using that Blast program.

**Markers in bold (8) were mapped in the BC1 population

<table>
<thead>
<tr>
<th>PAGE-monomorphic Marker**</th>
<th>Most Significant Homology</th>
<th>Highest Percentage of Identity*</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A6-160</td>
<td>Lycopersicon pennellii lipid transfer protein 1</td>
<td>85% 88%</td>
<td>Trevino and O’Connell (1996, unpublished) New Mexico State University</td>
</tr>
<tr>
<td>NBS6-330</td>
<td>A. thaliana resistance protein RPP13</td>
<td>NS 71%</td>
<td>Bittner-Eddy et al. 2000</td>
</tr>
<tr>
<td>SAS11-760</td>
<td>L. esculentum resistance complex protein I2C-2, I2C-5 and I2C-1</td>
<td>99% 94%</td>
<td>(Ori et al. 1997; Sela-Buurlage et al. 2001)</td>
</tr>
<tr>
<td>SAS21-520</td>
<td>RGAs and NBS of resistance-like genes in several species</td>
<td>98% 64%</td>
<td>(Pan et al. 2000)</td>
</tr>
<tr>
<td>S4-530</td>
<td>Potato disease resistance protein homologs 332 and 334; Capsicum annuum clone PR-protein gene</td>
<td>95% 85%</td>
<td>(Leister et al. 1996) Yi et al. (2002 Unpublished) Pepper Breeding Group, Chinese Academy of Agricultural Science</td>
</tr>
<tr>
<td>S3-440</td>
<td>R-gene homolog family St124, St125 and St13</td>
<td>85% 62%</td>
<td>(Leister et al. 1996)</td>
</tr>
<tr>
<td>SS1-530</td>
<td>Potato disease resistance protein homologs 332, 334</td>
<td>98% 82%</td>
<td>(Leister et al. 1996)</td>
</tr>
<tr>
<td>SS14-520</td>
<td>L. esculentum tospovirus resistance proteins (Sw5-a, Sw5-b, Sw5-c, Sw5-d and Sw5-e); Nematode resistance protein Mi-1.1; Potato Rx protein</td>
<td>87% 85%</td>
<td>(Rossi et al. 1998; Bendahmane et al. 1999; Folkertsma et al. 1999)</td>
</tr>
<tr>
<td>SS20-860</td>
<td>L. esculentum tospovirus resistance protein B (Sw5-b) and disease resistance protein Prf</td>
<td>82% 86%</td>
<td>(Salmeron et al. 1996; Folkertsma et al. 1999)</td>
</tr>
<tr>
<td>SS19-530</td>
<td>Solanum phureja x S. stenotomum NBS-type of RGAs</td>
<td>99% 97%</td>
<td>Simko and Jones (2001, unpublished) USDA-ARS, Beltsville, MD 20705, USA</td>
</tr>
</tbody>
</table>
Figure 4. Jukes-Cantor UPGMA trees displaying phylogenetic relationships between PAGE-polymorphic and PAGE-monomorphic markers amplified with $R$-genes degenerate primers. **A**, tree corresponding to the cluster of markers located within the QTL $EBR5.2$ in the tomato chromosome 5. **B**, cluster of markers within the QTL $EBR9.1$ in chromosome 9.
Figure 4

A

XLRR-370
SAS5-250
PK12-340
NBS8-300
PF6-220
RLRR-220

B

RLRR-130
SAS18-130
RLRR-140
XLRR-360
P2-220
SAS5-250
SAS13-510
NBS4-300
NBS8-300
C2-170
Chapter 4

Conclusions

This study confirmed the use of degenerate primers derived from \( R \)-genes as a very useful approach in plant genetics and breeding programs for the development of genetic markers, construction of linkage maps and QTL mapping.

Two types of DNA markers were produced by PCR-amplification of the degenerate primers and visualized by polyacrylamide gel electrophoresis (PAGE). They were PAGE-polymorphic, useful for direct genetic mapping, and PAGE-monomorphic markers, which had to be converted to RFLP probes for DNA hybridization and genetic mapping. Both markers exhibited simple dominant Mendelian segregations in a tomato \( BC_1 \) population, and were affected (as any other RFLP-anchor marker) by phenotypic selections imposed on the population. All of the markers were mapped throughout the tomato chromosomes in a non-random distribution, usually with a tendency to cluster and resembling the complex organization of \( R \)-gene families observed in other plant genetic maps.

Because of their clustering, proximity to previously mapped \( R \)-genes, associations with QTLs for early blight (EB) resistance, and the functional \( R \)-gene origins of the degenerate primers, the PAGE-polymorphic and monomorphic markers were originally assumed to be \( R \)-gene analogs (RGAs) tightly linked in a family cluster. Similar mapping
and breeding studies in tomato, rice, wheat and barley that considered the total amplified markers as RGAs also supported these assumptions.

This study, however, demonstrated that not all amplification products of degenerate primers are actual RGAs. In fact only a small proportion of the total amplified fragments, regardless of their utility in genetic mapping, are genuine RGAs, as determined by nucleotide and peptide sequence homologies using the Blast search tools. It appears that most of the other sequences, collectively called “RGAs”, may simply be pseudogenes or random and polymorphic amplification products. However, because of their confirmed linkages to QTL and R-genes, some markers might also represent amplified regions of non-annotated genes, or ancient non-functional R-gene-like sequences that once encoded for proteins with similar domains.

The results support the generation of DNA markers with R-gene-derived primers for genetic mapping and marker-assisted selection (MAS) to improve resistance responses to EB in tomato. Nevertheless, it is necessary to encourage the production of the undemanding and highly reproducible PAGE-polymorphic markers over the PAGE-monomorphic ones, since their application is exclusive for genetic mapping and the information contained in their sequences is irrelevant. Although PAGE-polymorphic markers resemble RAPDs, they may appear more attractive to breeders because of their intriguing chromosomal distribution and associations with QTLs for disease resistance.

The isolation and sequencing of polymorphic markers and monomorphic bands from PAGE is justified only for gene discovery and cloning projects, although PAGE might not be the most appropriate approach for such ventures because only the most prominent bands on the gels are considered and many others are lost. This study supports the combination of PAGE-polymorphic markers (of unknown sequences) with genuine RGA markers (in the form of RFLP probes) whose sequences have been confirmed by other investigators and their clones are publicly available at several research centers’ on-line databases. Thus, the cloning and sequencing of PAGE-monomorphic bands for the generation of RFLP probes is not recommended due to the relatively low rate of success in finding an RGA.
In this study, the terms “PAGE-monomorphic” and “PAGE-polymorphic”, assigned to the developed markers, seemed the most appropriate because they implied that only the markers that could be observed (and therefore scored and mapped) from polyacrylamide gels were considered for further investigation. These arbitrarily assigned terms (“PAGE-monomorphic” and “PAGE-polymorphic”) also intended to avoid the use of “RGAs” for all the identified markers, mainly because their sequences’ identities (that finally define a marker as an RGA) were not available until after the genetic mapping step (Chapter 3). To prevent confusion, markers amplified with R-gene-derived primers should not be immediately regarded as “RGAs”. Furthermore, in future studies, and depending on the method used in the visualization of the PCR-products of the R-gene primers, a different nomenclature might be needed for the description of the markers. It will be imperative to specify the primer combinations used (preferably indicating which conserved domain of an R-protein was used as the template for their design), how the polymorphism was obtained (e.g. RFLP hybridization), and especially, whether the marker met the criteria to be an authentic RGA based on the information of its nucleotide and deduced peptide sequences.
Vita
David Orlando Niño-Liu


Research Experience


Publications


Awards and Scholarships


Scientific Societies