Comparative genomics enabled the isolation of the $R3a$ late blight resistance gene in potato

Sanwen Huang1,†, Edwin A.G. van der Vossen1,2, Hanhui Kuang3, Vivianne G.A.A. Vleeshouwers1, Ningwen Zhang1, Theo J.A. Borm1, Herman J. van Eck1, Barbara Baker3, Evert Jacobsen1 and Richard G.F. Visser1,*

1Laboratory of Plant Breeding, Department of Plant Sciences, Graduate School Experimental Plant Sciences, Wageningen University, Binnenhaven 5, 6709 PD Wageningen, The Netherlands,
2Plant Research International BV, PO Box 16, 6700 AA Wageningen, The Netherlands, and
3Department of Plant and Microbial Biology, Plant Gene Expression Center, University of California, Berkeley, 800 Buchanan Street, Albany, CA 94710, USA

Received 15 October 2004; revised 20 December 2004; accepted 11 January 2005.
†For correspondence (fax +31 317 483457; e-mail richard.visser@wur.nl).
*Present address: Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, No 12, Zhong Guan Chun Nan Da Jie, Beijing 100081, P.R. China.

Summary

Comparative genomics provides a tool to utilize the exponentially increasing sequence information from model plants to clone agronomically important genes from less studied crop species. Plant disease resistance ($R$) loci frequently lack synteny between related species of cereals and crucifers but appear to be positionally well conserved in the Solanaceae. In this report, we adopted a local RGA approach using genomic information from the model Solanaceous plant tomato to isolate $R3a$, a potato gene that confers race-specific resistance to the late blight pathogen Phytophthora infestans. $R3a$ is a member of the $R3$ complex locus on chromosome 11. Comparative analyses of the $R3$ complex locus with the corresponding $I2$ complex locus in tomato suggest that this is an ancient locus involved in plant innate immunity against oomycete and fungal pathogens. However, the $R3$ complex locus has evolved after divergence from tomato and the locus has experienced a significant expansion in potato without disruption of the flanking colinearity. This expansion has resulted in an increase in the number of $R$ genes and in functional diversification, which has probably been driven by the co-evolutionary history between $P$. infestans and its host potato. Constitutive expression was observed for the $R3a$ gene, as well as some of its paralogues whose functions remain unknown.

Keywords: $R3a$, comparative genomics, potato, tomato, disease resistance, Phytophthora infestans.

Introduction

Comparative genomics investigates the similarity and differences in structure and function of genomes across taxa. Full genome sequences of Arabidopsis (Arabidopsis Genome Initiative, 2000) and rice (Goff et al., 2002; Yu et al., 2002) and several other ongoing sequencing projects will offer unprecedented resources to study the evolution of sequence and function of orthologous genes and to understand diversification and adaptation. A fundamental and practical question in comparative genomics is whether this vast amount of sequence information from model plant species will facilitate the cloning of genes with agronomic importance from crop species with larger genomes. Resistance to plant pathogens, often defined by single dominant disease resistance ($R$) genes (Dangl and Jones, 2001), is an important crop trait that could benefit from the sequencing of model species. This benefit, however, will largely depend on the given plant family. Extensive loss of colinearity has been reported within crucifers and grasses (Gale and Devos, 1998; Paterson et al., 2000). Comparative analysis based on DNA sequences has revealed that disease resistance ($R$) loci may be evolving faster than the rest of the crucifer and grass genomes (Gale and Devos, 1998; Leister et al., 1998; Paterson et al., 2000). In contrast, the remarkable conservation of gene order (Bonierbale et al., 1988) makes the Solanaceae, a botanical family that includes many important crops such as tomato and potato, an attractive subject for comparative genomics. Low-resolution comparative mapping indicated that $R$ loci may be positionally conserved within Solanaceae (Grube et al., 2000; Leister
et al., 1996; Pan et al., 2000). To date, this knowledge has never led to the isolation of a new R gene with known function.

The distal end of the long arm of chromosome 11 of tomato is a particularly interesting genomic region to investigate the evolution of R loci in a comparative genomics perspective. First, the overall structural colinearity is excellent between tomato and potato (Tanksley et al., 1992) and relatively good between tomato/potato and pepper (Livingstone et al., 1999). Secondly, this genomic region is a hotspot for R genes, harboring major genes encoding resistance to the fungi *Fusarium oxysporum* (I2, Ori et al., 1997; Simons et al., 1998) and *Stemphylium* spp. (Sm, Behare et al., 1991), and to yellow leaf curl virus (Hanson et al., 2000) in tomato, to the oomycete *Phytophthora infestans* (R3, R6, and R7, El Kharbotly et al., 1994, 1996) in potato, and to tobacco mosaic virus (L, Lefebvre et al., 1995) in pepper. The region also contains several quantitative trait loci conferring resistance to the cyst nematode *Globodera rostochiensis* (Gro1.3, Kreike et al., 1993), to *P. infestans* (phy7, Oberhagemann et al., 1999) in potato, and to cucumber mosaic virus (cmv4, Grube et al., 2000) and to *P. capsici* (phyt3, Lefebvre and Palloix, 1996) in pepper. Last, the molecularly well characterized I2 complex locus in tomato (Ori et al., 1997; Simons et al., 1998) provides an excellent template for a comparative study. The complex locus consists of two clusters, SL8D and SL8E. The SL8D cluster contains seven coiled-coil nucleotide binding site and leucine-rich repeat (CC-NBS-LRR) type R gene sequences, including the I2 gene conferring complete resistance to race 2 of *F. oxysporum* f. sp. *lycopersici* (Simons et al., 1998) and the I2C-1 (Ori et al., 1997) and I2C-5 (Sela-Buurlage et al., 2001) genes conferring partial resistance to the same pathogen.

Potato is the most important non-cereal crop plant. One of the major constraints to potato production in the world is the late blight disease caused by the oomycete *P. infestans* (Fry and Goodwin, 1997). There is considerable interest in identifying late blight R genes and in understanding their evolution and mode of action (Ballvora et al., 2002; Song et al., 2003; van der Vossen et al., 2003). In a previous study (Huang et al., 2004), we discovered that the *R3* complex locus at chromosome 11 of potato comprises two functionally distinct late blight R genes, *R3a* and *R3b*. Here, we use genomic information from the model Solanaceous plant tomato to clone *R3a*, a gene that confers race-specific resistance to *P. infestans*.

**Results**

The potato *R3* and tomato I2 regions are highly colinear

We previously showed that the genomic regions harboring the *R3* late blight resistance locus in potato and the I2 *Fusarium* wilt resistance locus in tomato are colinear (Huang et al., 2004). To determine the extent of colinearity between these regions, we mapped seven tomato markers on the high-resolution genetic map of the R3 complex locus (Figure 1). All seven markers retained their order in tomato and potato. We identified a cluster of I2 gene analogues (I2GAs) in potato that was mapped 0.1 cM centromeric to the TG105 marker. This potato I2GA cluster positionally corresponds to the SL8D cluster of the I2 complex locus in tomato and was therefore named the St-I2 cluster. The perfect micro-colinearity within the TG105-cLET24J2A interval indicates that the *R3a* cluster in potato is syntenic to the SL8E cluster in tomato. Despite the fact that the counterpart of the *R3b* cluster was not detected in tomato, the orthologous relationships of SL8D versus St-I2 and SL8E versus R3a point to an ancient R locus prior to the tomato–potato divergence.

**R3a candidates were identified using a local resistance gene analogue (RGA) approach**

Although the potato R3a-tomato SL8E synteny was established (Figure 1), we could not directly use it for *R3a* cloning as there was no sequence available from the SL8E cluster.

---

**Figure 1.** Comparative genetic maps of the I2 complex locus in tomato and the R3 complex locus in potato. The left and right solid bars represent the tomato and potato chromosomes, respectively. Orientation is indicated by CEN (centromere) and TEL (telomere). Mapped markers are connected by black lines. The syntonic relationships of R gene clusters are highlighted using gray rectangles. On the tomato map, the positions of TG105, TG26, and CT120 are according to Tanksley et al. (1992); cLET5E4, T1660, and cLET24J2A according to the Solanaceae Genomics Network (http://www.sgn.cornell.edu), and SL8D (containing the I2 gene), TA01-1, and SL8E according to Ori et al. (1997). On the potato map, the positions of cLET5E4, TG105, R3a, TG26, R3b, and CT120 were determined by a previous work (Huang et al., 2004), TA01-1, T1660, and cLET24J2A by aligning the marker sequences to the sequences of BACs mapped on the high-resolution map of R3 (data not shown). The position of St-I2 was determined as described in the text.
However, the SL8E cluster was defined by cross-hybridization with the LRR part of the I2C-1 gene from the SL8D cluster (Ori et al., 1997). Using this information, a local RGA approach was applied. Instead of targeting conserved motifs within the NBS applied by global RGA approaches (Aarts et al., 1998; Kanazin et al., 1996; Leister et al., 1996; van der Linden et al., 2004; Yu et al., 1996), the conserved sequences within the LRR of the SL8D cluster were used to design the pSL8D primers (Table 1C). A BLAST search (Altschul et al., 1990) for short, nearly exact matches proved the specificity of the pSL8D primers, which hit no sequence other than the SL8D cluster with a reasonable low E-value. On DNA from the parental clones SH83-92-488 (SH) and RH89-039-16 (RH), the primers amplified a major smearing band of approximately 1 kb (data not shown), indicating that many copies of I2GAs with similar sizes were amplified. The size fitted the prediction from the sequences of the SL8D cluster (884–1049 bp). The above findings confirmed the suitability of the pSL8D primers for identifying I2GAs in potato.

To identify I2GAs physically close to R3a, an association analysis on bacterial artificial chromosome (BAC) pools was conducted. Similar to other physical mapping methods like radiation hybrid mapping (Cox et al., 1990) or HAPPY mapping (Dear and Cook, 1993), our method is based on the frequency that a given marker and I2GAs co-appear in BAC pools to estimate physical distance between them. The mapping panel is represented by BAC pools that contain 384 BACs, equivalent to 0.05 genome.

To determine whether any R3a marker (Figure 2a) was physically close to an I2GA, we screened 255 BAC pools with

<table>
<thead>
<tr>
<th>Marker</th>
<th>Primer (5’-3’)</th>
<th>Tm (°C)</th>
<th>Restriction enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A cLET5E4</td>
<td>CCA GGC ATG CTC AAT TTG GAG T</td>
<td>55</td>
<td>Hhal</td>
</tr>
<tr>
<td></td>
<td>TTC CCT GTG ACT AGT TGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT120</td>
<td>CGA GGG GCC AAA CTG GAA ACC AGT</td>
<td>52</td>
<td>Tsp509I</td>
</tr>
<tr>
<td></td>
<td>CCA TGA ATG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B 241S</td>
<td>GGC ACT GAT AAG TTT TGG TTT TG</td>
<td>55</td>
<td>a.s.</td>
</tr>
<tr>
<td></td>
<td>GTG GTT TAT GAA TGG TAC TCA TGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>194T</td>
<td>AAG CTT GAA TCC GTG GAC GG</td>
<td>55</td>
<td>a.s.</td>
</tr>
<tr>
<td></td>
<td>CTT TAT ACC AAC AGG TTG TCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4T</td>
<td>AAG CTT TCA AAC CAA ACT GC</td>
<td>52</td>
<td>a.s.</td>
</tr>
<tr>
<td></td>
<td>AAA ATG ACT TTA CGT GGT CT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAT CCG ACC ACC ACC AAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP250</td>
<td>ACC AGT AGG ACC ACC ACC AAT</td>
<td>60–52b</td>
<td>VspI</td>
</tr>
<tr>
<td></td>
<td>GAT CGT GAC GCC TCT ACT TTT TTA TGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP250-2</td>
<td>CGT GCA TAA GTA GTA ATG ATT TTT TGA CT</td>
<td>60–52b</td>
<td>VspI</td>
</tr>
<tr>
<td></td>
<td>GAT CCT GAC TCG TCT ACT TTA CGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C pSL8D</td>
<td>AAT TTG AGA GTT CCC TAC ACT TGA GAG GAG GGG AGC GCT GAT TCT</td>
<td>50</td>
<td>AflI</td>
</tr>
<tr>
<td>D RACE4</td>
<td>CGA AAG GAG TGG CAT TAA CAG AGA CGA</td>
<td>68–60b</td>
<td></td>
</tr>
<tr>
<td>RACE5</td>
<td>CTT CCT CTA TCC AAA GGC TCA TCC TAA TAG T</td>
<td>68–60</td>
<td></td>
</tr>
<tr>
<td>RACE6</td>
<td>CTC ACC TCT TTT TAA TTA TTT ATT AAG</td>
<td>68–60</td>
<td></td>
</tr>
<tr>
<td>RACE0</td>
<td>CAT TGG AAA CGG CTT TAG CAA GTG TGG TC</td>
<td>68–60</td>
<td></td>
</tr>
<tr>
<td>RACE4</td>
<td>CAT TCC AAT AGG AAG TAC AGT CGG C</td>
<td>68–60</td>
<td></td>
</tr>
<tr>
<td>RACE5</td>
<td>CAT TCT GCC TCT CAA AGA TAC CAG AG</td>
<td>68–60</td>
<td></td>
</tr>
<tr>
<td>RACE6</td>
<td>CAT CAC TTA GCA CAA TCT GAA GAC CG</td>
<td>68–60</td>
<td></td>
</tr>
<tr>
<td>E RT-SH23-1</td>
<td>GCT TCC GAT GTG TAT TGA TCC CAC G</td>
<td>68–55b</td>
<td></td>
</tr>
<tr>
<td>RT-R3a</td>
<td>TGT GGC ATT CCT CAA ACA ATG T</td>
<td>68–55</td>
<td></td>
</tr>
<tr>
<td>RT-SH23-3</td>
<td>ATC GTG GTC ATG CTA TGA GTG TGT T</td>
<td>68–55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTT CAA GGT ATG GGG CAG TAT GCT T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT-SH194-2</td>
<td>CAG CAT GAA ATT ATT TCA TGC TGT C</td>
<td>68–55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGA GGA ATT TCA AAC AAG GGA GTT C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actinb</td>
<td>CAG CAA CTG GAG TGA TAT GG</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATT TCG TTC TCA GCA GTG GT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A, flanking markers; B, physical mapping; C, RGA marker; D, RACE; E, RT-PCR.
bPrimers kindly provided by Chengwei Li of the Laboratory of Plant Breeding, Wageningen University.
bTouchdown PCR, −1°C/cycle.
bAllele-specific.

the nine R3a markers and the PSL8D primers. A total of 30 BAC pools were determined to be positive for one or more of the R3a markers, placing these markers in five groups on the basis of recombination events and co-occurrence in BAC pools (Figure 2a,b). The PSL8D primers identified 92 I2GA-containing BAC pools, 19 of which coincided with those positive for the R3a markers. Chi-square tests showed that I2GAs were significantly associated with the markers in groups III and IV that cosegregated with R3a (Figure 2b). This suggested a close physical relationship between R3a and specific I2GAs, which therefore were regarded as R3a candidates.

From the BAC pools identified by the R3a markers and by chromosome walking, we isolated 11 BAC clones that span at least 700 kb (Figure 2c). Subsequently, BAC clone SH194C17 and SH23G23 were chosen for sequencing, as they were respectively identified by group III and IV markers (Figure 2c) and Southern hybridization (data not shown) indicated they contained five I2GAs. Annotation of the BAC sequences indeed predicted five I2GAs that were designated I2GA-SH194-1, -2, I2GA-SH23-1, -2, and -3 (Figure 2d). Single continuous open reading frames (ORF) were present in I2GA-SH194-2 and the three I2GAs in BAC SH23G23. I2GA-SH194-1 is a pseudogene due to insertion of a Ty1/Copia-type retroelement. In the BAC sequences, the GP250 marker and the amplified fragment length polymorphism (AFLP) markers could also be recognized. The physical proximity between the I2GAs and the group III/IV markers (Figure 2d) validated the predicted association at the level of BAC pools (Figure 2b). The four I2GAs with full ORFs were considered as R3a candidates.

R3a is an I2GA

The genes I2GA-SH194-1 and -SH23-1, -2, -3 together with 2-3 kb up- and down-stream sequences were introduced into the susceptible clone 1029-31 and cv. Desiree via Agrobacterium-mediated transformation. At least 10 in vitro plantlets of each primary transformant were tested for resistance to each of the three P. infestans isolates 89148-9, IPO-0, and H30P04 in duplicate experiments. Cv. Desiree, 1029-31, RH, SH, and SW850-025 (R3a recombinant) were included as controls. Only a 10.4-kb subclone containing I2GA-SH23-2 was able to specifically complement the susceptible phenotype. All 15 primary transformants of this subclone exhibited R3a-specific resistance, that is, were resistant to isolate 89148-9 and IPO-0 and susceptible to H30P04 (Table 2 and Figure 3). The results were confirmed using a detached-leaf assay in duplicate tests. As the subclone contained only one ORF, we designated I2GA-SH23-2 as R3a.
The transcript structure of the R3a gene was determined by comparing the genomic sequence with cDNA fragments generated by random amplification of cDNA ends (RACE). The R3a transcript is 4176 nt long and encodes a predicted polypeptide of 1282 amino acids (a.a.) with a relative molecular mass of 145.9 kDa. The R3a gene and the I2 gene (Simons et al., 1998) have a similar intron–exon pattern and both have no intron in the coding region, but the intron close to the stop codon is much longer in R3a than in I2 (Figure 4a).

The R3a gene encodes a putative CC-NBS-LRR protein and shares 88% DNA identity and 83% a.a. similarity to I2 (Figure 4b). Considerable dissimilarity occurs at the CC domain, including a seven-a.a. indel. R3a and I2 proteins are quite conserved in the NBS domain (86% a.a. identity), especially at the motifs that define the domain (94% a.a. identity). The proteins mainly diverge at the LRR domain where two major differences were observed. First, although both proteins carry 29 LRR units, the R3a protein lacks a complete LRR unit of 28 a.a. at LRR 14/15 and contains an extra copy of a unique LRR unit of 23 a.a. at the LRR 26/27 (Figure 4b). Secondly, 45% (63/140) of the putative solvent-exposed residues (x in the frame xxLxxx) differ between the I2 and R3a proteins.

The R3a and I2 proteins are more related to each other than to other known R proteins, as indicated by the zero E-value using BLASTP (Altschul et al., 1990). Thus the I2 and R3a genes belong to the same R gene family. Outside of this family, the most related known R gene is RpO1-b (E = e⁻¹⁷⁸, 35% a.a. identity) from soybean mediating recognition of the Type III effector protein AvrB from Pseudomonas syringae (Ashfield et al., 2004). The R3a protein bears only limited similarity (15 and 30% a.a. identity) to the other two known late blight R proteins, R1 (Bailvora et al., 2002) and RB/Rpi-blb1 (Song et al., 2003; van der Vossen et al., 2003), respectively.

**Dynamic evolution occurred in the R3a cluster**

Despite the excellent overall colinearity in the I2 and R3 genomic regions (Figure 1), the R3 region is physically larger in potato. In tomato the TG105-TG26 interval spans about 500 kb and contains one or few I2GAs, as demonstrated by long-range physical mapping (Simons et al., 1998) and Southern analysis (Ori et al., 1997). In contrast, the TG105-TG26 interval in SH spans more than 1 MB and contains at least nine I2GAs. The three contigs (11 BACs in total) in the R3a region (Figure 2c) span at least 700 kb. In addition, gaps between the depicted contigs and those between the depicted contigs harboring the markers TG105 and TG26 are predicted to be larger than 100 kb, the average

---

**Table 2** Disease test of primary transformants of R3a and its paralogues. Numbers of plants showing resistance to the three Phytophthora infestans isolate IPO-0, H30P04, and 89148-09 are indicated.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Clone</th>
<th>No. plants tested</th>
<th>89148-09</th>
<th>IPO-0</th>
<th>H30P04</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2GA-SH23-1</td>
<td>1029-31</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Desiree</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I2GA-SH23-2 (R3a)</td>
<td>1029-31</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Desiree</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I2GA-SH23-3</td>
<td>1029-31</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Desiree</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>I2GA-SH194-2</td>
<td>1029-31</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Desiree</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>1029-31</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Desiree</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>RH</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SW8540-025</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

*For constructs – number of independent transformants; for control – number of in vitro plantlets of the same genotype inoculated with each isolate.

*89148-09 and H30P04 kindly provided by Dr F. Govers of Laboratory of Phytopathology, Wageningen University and IPO-0 by Dr W. Flier of Plant Research International, Wageningen, The Netherlands.

---

Figure 3. In vitro inoculation of the primary transformants of R3a. Massive sporulation (S) and localized hypersensitive reactions (HR) are observed on compatible and incompatible interactions, respectively. RH and 1029-31 are susceptible to both isolates. SH is resistant to both isolates. SW8540-025 (an SH x RH progeny carrying only R3a) and 1029-31-T005 (a 1029-31 transformant with I2GA-SH23-2) are resistant to IPO-0 (with Avr3a) and susceptible to H30P04 (without Avr3a).
Figure 4. Comparison between R3a and I2.

(a) Schematic diagram of the transcript structures of I2 and R3a. Horizontal lines indicate exons. Open rectangles represent coding sequences. Lines angled downwards indicate the position of introns, whose sizes are indicated below.

(b) Primary structure and alignment of the I2 and R3a resistance proteins. Amino acid residues of I2 and R3a that are identical are shown in normal script. I2- and R3a-specific residues are shown in bold at top and bottom lines, respectively. Dashes indicate deletion. The predicted coiled coils (Lupas et al., 1991) in the CC domains are underlined. The conserved motifs of the NBS domain are overlined. The 30 LRR units in the alignment are numbered. The B-sheet (consensus xxxLxxxx) of each LRR is highlighted with a gray background. The R3a residues under significant diversifying selection are highlighted with a black background. The calculation is based on the alignment of R3a, I2GA-SH23-1, -3, and I2GA-SH194-2 using PAML (Yang et al., 2000). Between vertical lines are the unique LRR units (consensus NLSQSLAESALPSMHLLEIDCP) that display distinct copy numbers between I2, R3a, and their paralogues.

insert size of the SH BAC library. In addition to the five I2GAs in BAC clone SH194C17 and SH23G23, we identified a new I2GA in each of the four BACs (SH170015, SH24E01, SH193P23, and SH32O08, Figure 2c) by Southern blotting (data not shown). Altogether, these data indicate that the R3 complex locus in SH has expanded significantly compared to that of tomato (cv. Motelle and Mogeor, Ori et al., 1997; breeding line E22, Simons et al., 1998), resulting in an increase in the number of R gene sequences. It remains unknown whether such variation in size and R gene copy number reflects de facto tomato and potato genomes or whether such variation also exists within each species.

Diversifying selection has been detected in many R gene complex loci (Hulbert et al., 2001). Sites under diversifying selection were investigated using the program paml (Yang, 1997; Yang et al., 2000). Models M7 and M8 in the program ‘codeml’ of paml were run for the four I2GAs (R3a, I2GA-SH23-1, -3, and I2GA-SH194-2) at the R3a cluster in potato. Model M7 is a special case of model M8 and assumes no selection, whereas model M8 allows for positively selected sites (Yang et al., 2000). Diversifying selection can be confirmed using a likelihood-ratio test by comparing the likelihood calculated using models M8 and M7 (Yang et al., 2000). Comparison of the results from model 7 and model 8 showed that the I2GAs at the R3a cluster have undergone diversifying selection ($\chi^2 = 92.8$, df = 2; $P < 0.001$). Selection at each site of the R3a homologues was calculated using model M8. Thirteen sites were found to be under significant diversifying selection (Figure 4b and Figure S1). Twelve of them are in the LRR domain and seven are putative solvent-exposed residues. This observation is compatible with the idea that the LRR region of an R protein mainly defines recognition specificities (Parniske et al., 1997; Van der Hoorn et al., 2001; Wulff et al., 2001).

Putative sequence exchange between paralogues was previously observed at several R gene complex loci (Noel et al., 1999; Parniske et al., 1997) and is likely a mechanism for creating new specificities (reviewed in Hulbert et al., 2001). Sequence relationships within the R3a cluster were analyzed by determining informative polymorphic sites (IPS, Parniske et al., 1997). In total, 129 IPS were detected (Figure S2). The R3a and I2GA-SH23-1 genes have 94% DNA identity and they share 91 IPS, suggesting that they are derived from the same ancestral gene. Interestingly, in the central part of the alignment, the R3a cluster exhibits a complex patchwork. The R3a and I2GA-SH194-2 genes share sequence affiliations at three patches, comprising 21 IPS, which result in 10 non-synonymous substitutions between R3a and I2GA-SH23-1. An almost continuous sequence affiliation (nine IPS) between the R3a and I2GA-SH23-2 genes was also found, comprising three non-synonymous substitutions. It remains unclear whether the putative sequence exchanges between the paralogues have lead to new resistance specificities in the R3a cluster.

R3a is constitutively expressed

To examine the expression patterns of the genes R3a, I2GA-SH23-1, -3, and I2GA-SH194-2, gene-specific semiquantitative RT-PCR was performed on mRNA isolated from P. infestans challenged (isolate IPO-0 and 90128) or mock-inoculated leaves of SH at 0, 8, 16, 24, 32, 48, and 72 h post-inoculation. All the four I2GAs are constitutively expressed in all treatments, but the expression level of I2GA-SH23-3 is significantly lower than the other paralogues (Figure 5).

Discussion

In this study, we used genomic information from tomato to isolate the potato late blight resistance gene R3a from an ancient locus involved in plant innate immunity in the Solanaceae. Despite technological advances in genomics in recent years, positional cloning of R genes from large-genome crop species is still far from a routine procedure mainly due to low recombination frequencies and the high repetitive nature that are characteristic of most complex R loci (Hulbert et al., 2001). In potato, the heterozygous genome is an additional complicating factor (Kanyuka et al., 1999). To partially circumvent these obstacles, our local RGA approach analyzes genetic markers on BAC pools or super-pools, offering enough template complexity in PCR to allow most markers to be allelic- and locus-specific when compared with the hybridization method on single BAC filters. Although the RGA-specific primers (FSLAD) were designed for the exclusive identification of the I2 gene family, they were also designed to be allele- and locus-insensitive in order to include all candidate RGAs at the target region. Through association on BAC pools, we identified a subset of RGAs that were in coupling phase with and physically close to genetic markers linked to R3a and thus ‘jump’ rather than ‘walk’ into BACs carrying such RGAs. The success of the method depends on the marker saturation level at the target region and the capacity of the RGA-specific primers to
amplify all the candidate RGAs at the target region. In this case, we benefited from the marker saturation provided by the ultra-high dense (UHD) AFLP potato map (Isidore et al., 2003), as the same mapping population (SH × RH) was used to generate the UHD map and for the initial mapping of $R3a$ (Huang et al., 2004). In general, a bulked segregant analysis (Michelmore et al., 1991) in combination with AFLP marker technology will generate adequate marker saturation of the $R$ gene regions as the ‘wild’ $R$ haplotype often differs a lot from the ‘domestic’ $r$ haplotype, which is also the reason that they hardly recombine. The synteny at the target region between tomato and potato was beneficial to the RGA approach, despite extensive expansion of the $R3$ complex locus compared to the syntenic $I2$ complex locus.

Most of the molecularly characterized $R$ genes belong to families of tightly linked genes and at many such complex loci, most paralogues appear to encode proteins similar to the functional $R$ genes (Hulbert et al., 2001). We demonstrated that all four paralogues of the $R3a$ cluster are constitutively expressed (Figure 5), even in unchallenged plants. It is unclear whether these paralogues are capable of interacting with unknown elicitors or whether they are just relics of a recent ‘birth-and-death’ process (Michelmore and Meyers, 1998). Given the fitness cost of $R$ gene expression (Tian et al., 2003), unnecessary $R$ genes should undergo a ‘death’ process. We also found many truncated paralogues, mutated via frame-shift (data not shown), or inserted by a retroelement (Figure 2d). The weaker expression of $I2GA-SH23-3$ (Figure 5) might suggest that mutations accumulated in the promoter region can lead to a lower expression and thus eventually to the ‘death’ of the gene. Alternatively, the expressed paralogous proteins may play a role through heteroduplex formation with $R3a$, as inter- and intra-molecular interactions between $R$ protein domains may function as activation switches upon recognition of Avr elicitors (Moffett et al., 2002).

Plants cannot move to escape their microbial environment. To combat disease, plants develop a sophisticated innate immunity system, where $R$ genes play a central role (Dangl and Jones, 2001). Comparative genomics may provide insight into how diseases have resulted in differential evolution of $R$ loci between closely related plant species. The potato $R3$ region has undergone a significant physical expansion compared to the syntenic tomato $I2$ region. This expansion may have resulted in functional diversification. For instance, at least 10 $I2GAs$ have been found at the $R3b$ region (S. Huang, H. Kuang, E. van der Vossen, V. Vleeshouwers, E. Jacobsen, B. Baker and R. Visser, unpublished data), suggesting $R3b$ could also be an $I2GA$. The contrasting evolutionary fates of the ancient $I2-R3$ complex locus in the closely related tomato and potato genomes are consistent with the opposite evolutionary potentials of the interacting pathogens (McDonald and Linde, 2002). *Fusarium oxysporum* f. sp. *lycopersici* is a soilborne fungus with low genotype diversity, whereas the late blight pathogen *P. infestans* is notorious for its ability to move and mutate (McDonald and Linde, 2002). The great evolutionary potential of *P. infestans* may have stimulated the interacting $R3$ complex locus to expand its $R$ gene repertoire, as supported by the observed physical expansion and the fact that two functionally distinct genes locate in the locus (Huang et al., 2004). *Phytophthora infestans* also infects tomato, but genomic regions controlling late blight resistance show very limited overlap between tomato and potato (Brouwer et al., 2004; Gebhardt and Valkonen, 2001; Grube et al., 2000). Late blight resistance in the two closely related species is likely conferred by different loci that have evolved independently.

Plant disease resistance genes display two distinct evolutionary patterns contrasting at the rate of evolution (Kuang et al., 2004). The late blight $R$ gene $RB/Rpi-b1b1$ from *Solanum bulbocastanum* belongs to the type II (slow evolving) $R$ gene class (Song et al., 2003). However, the occurrence of sequence exchanges between paralogues (Figure S2), the multi-allelic nature of the locus (see below), high nucleotide identities between homologues (90–94%) and obscure allelic/orthologous relationships between the SH $R3$ haplotype and other *S. demissum* haplotypes (H. Kuang, S. Huang, X. Wang, R. Visser and B. Baker, unpublished data) lead to the putative classification of $R3a$ into the type I (fast evolving) class. It remains unknown why these two late blight $R$ genes differ by the rate of evolution, but there are some hints. $RB/Rpi-b1b1$ is resistant to all tested races of *P. infestans* and provides resistance by reducing the infection rate (Song et al., 2003; van der Vossen et al., 2003), while $R3a$ is a race-specific gene and displays a typical hypersensitive necrosis response (Huang et al., 2004). *Solanum bulbocastanum* often grows under quite dry conditions (Hawkes, 1990), suggesting the absence of co-evolution with *P. infestans*, whereas *S. demissum* has co-evolved with the pathogen on the cool mountain forests in Toluca Valley of Mexico (Rivera-Pena, 1990), which was recognized as a center of diversity for *P. infestans*.

An intriguing question is why *S. demissum* appears to contain only easily broken $R$ genes (Wastie, 1991) but displays durable resistance at the population level (Rivera-Pena, 1990). The polymorphism of parasite recognition capacity in a host population will restrict most isolates of the parasite to grow on most hosts (Hamilton et al., 1990). Allelism is an efficient way of creating recognition polymorphism (Bergelson et al., 2001), and we are currently investigating the multiple allelism of the $R3$ complex locus. (S. Huang, V. Vleeshouwers, E. Jacobsen and R. Visser, unpublished data), which might be a natural mechanism of *S. demissum* to suppress late blight epidemics, similar to the concept of $R$-gene polycultures or multilines (Jones, 2001; Mundt, 2002). The isolation of $R3a$ and characterization of the $R3$ complex locus will facilitate the cloning of other
alleles and thus provide a platform to test the R-gene polyculture concept in late blight disease management in commercial potato production.

Experimental procedures

Plant material
The F1 population of SH83-92-488 (SH) × RH89-039-16 (RH) that segregates for both R3a and R3b was used for genetic mapping (Huang et al., 2004). SW8540-025 is a recombinant that only carries R3a. The susceptible potato clones 1029-31 and cv. Desiree were used for genetic transformation.

Phytophthora infestans isolates and inoculation
Phytophthora infestans isolates 89148-09 (Avr3a, Avr3b), IPO-0 (Avr3a, avr3b), H30P04 (avr3a, Avr3b), and 90128 (avr3a, avr3b) were inoculated on in vitro plantlets (S. Huang and V. Vleeshouwers, unpublished data) or detached leaves (Huang et al., 2004).

Genetic mapping
A total of 3000 SH × RH progeny were screened for recombinations in the R3a region using the flanking markers cLET5E4 and CT120 (Table 1A). The recombinants were used to determine the position of the R3a markers (Figures 1 and 2a). DNA isolation and analysis of molecular markers were described previously (Huang et al., 2004).

Physical mapping
A 10-genome equivalent BAC library of the resistant parent SH, stored in 255 384-well microtiter plates, was screened with the R3a markers stepwise: first, positive superpools (1 superpool = 8 BAC pools) were identified; secondly, positive pools were identified from those positive superpools. Plasmid DNA was isolated using the standard alkaline-lysis protocol from pooled bacteria to produce 28/3 BAC pools. Screening of the BAC pools and identification of single positive BACs were performed as described previously (Rouppe van der Voort et al., 1999). Plasmids of single BACs were purified using MIDI-PREP columns (Qiagen, Hamburg, Germany). Sequences generated through BAC-end sequencing with SP6 or T7 primers were employed to design PCR primers to develop BAC end markers. The BAC end markers were named according to the BAC pool number followed by S (SP6) or T (T7) (Table 1B). In view of the repetitive nature of the R3 complex locus, the overlap of two BACs was verified by: (1) HindIII restriction mapping, (2) non-selective AFLP fingerprinting using HindIII and Msel/TaqI (Simons et al., 1998), and (3) analysis of BAC end markers through Tm-gradient PCR, restriction with 24 frequent cutters, and sequencing of the PCR products. The length of overlapping between two BACs was determined by adding the sizes of co-migrating HindIII restriction fragments.

DNA sequencing and analysis
The DNA sequences of the BAC clones SH194C17 and SH23G23 (Figure 2d) were determined by using a shotgun sequencing strategy (van der Vossen et al., 2000). Positions of putative genes were predicted using GENSCAN (Burge and Karlin, 1997) and GENEFLASH (Lukashin and Borodovsky, 1998). Multiple sequence alignments were conducted by using CLUSTAL X 1.81 (Thompson et al., 1997) and IGS (Parniske et al., 1997) were determined. Diversifying selection was investigated using PAML (Yang, 1997; Yang et al., 2000). Models M7 and M8 in program ‘codem’ of PAML were run for all 12GAs at the R3a cluster. Diversifying selection was confirmed using a likelihood-ratio test by comparing the likelihood calculated using models M8 and M7 (Yang et al., 2000). Selection in each site of the R3a homologues was calculated using model M8.

Subcloning and transformation of 12GAs
Candidate 12GAs were subcloned into the binary vector pBINPLUS (van Engelen et al., 1995) as described previously (van der Vossen et al., 2003). Binary plasmids harboring the candidate genes were transformed to Agrobacterium tumefaciens strain AGL0 (Lazo et al., 1991) and introduced into 1029-31 and cv. Desiree according to standard protocols (Visser et al., 1991).

Transcript analysis of the R3a cluster
Total RNA was extracted from detached leaves of SH inoculated with P. infestans IPO-0, 90128, and water using TRIZOL (Invitrogen, Carlsbad, CA, USA). The mRNA was isolated from the total RNA samples using the OLIOTEC kit (Qiagen). The 5’- and 3’-ends of the R3atranscript were determined by rapid amplification of cDNA ends (RACE) using the MARATHON kit (BD Bioscience, USA) with nested gene-specific primers (3RACE-4 and 3RACE-6, Table 1D). The cDNA for 5’ RACE was synthesized with the primer SRACE0. PCR products were cloned into pGEM-T vector (Promega, Madison, WI, USA). For gene-specific RT-PCR (Table 1E), an aliquot of 0.1 μg mRNA from each sample was used to generate single-stranded cDNA using SUPERSCHRIFT II (Invitrogen). Actin primers were used to monitor mRNA concentration (Table 1E). Genomic DNA of SH and RH and BAC subclones were used as control. Gene-specific amplification was confirmed by sequencing.

Acknowledgements
We thank Bart Brugmans for help with physical mapping, Marjon Arens-de Reuver for mRNA extraction, Jing Zhang and Isolda Pereira for plant transformation, Andy Pereira and Chun-Ming Liu for critical reading, and Sophien Kamoun for inspiring discussions and valuable advices.

Supplementary Material
The following material is available from http://www.blackwellpublishing.com/products/journal/suppmat/TPJ/TPJ2365/TPJ2365s.htm
Figure S1. The distribution of codons under diversifying selection in R3a.
Figure S2. Sequence relationships between members of the R3a cluster.

References

DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Mol. Plant Microbe Interact. 11, 251–258.


Cloning the potato late blight R3a gene by synteny 261

Sequence accession numbers: AY849382, AY849383, AY849384, AY849385