The Plant Journal (2005) **44**, 37–51 doi: 10.1111/j.1365-313X.2005.02506.x

The *R1* resistance gene cluster contains three groups of independently evolving, type I *R1* homologues and shows substantial structural variation among haplotypes of *Solanum demissum*

Hanhui Kuang[1,†], Fusheng Wei[1,†], María Rosa Marano[1,2,†], Uwe Wirtz[1,†], Xiaoxue Wang[1,†], Jia Liu[3], Wai Pun Shum[1], Jennifer Zaborsky[3], Luke J. Tallon[3], Willem Rensink[3], Stacey Lobst[3], Peifen Zhang[1], Carl-Erik Tornqvist[1], Ahmet Tek[4], John Bamberg[4,5], John Helgeson[6], William Fry[7], Frank You[8], Ming-Cheng Luo[8], Jiming Jiang[4], C. Robin Buell[3] and Barbara Baker[1,*]

[1]*Plant Gene Expression Center, USDA-ARS and Department of Plant and Microbial Biology, University of California, Berkeley, 800 Buchanan Street, Albany, CA 94710, USA,*

[2]*Instituto de Biología Molecular y Celular de Rosario, Facultad de Ciencias Bioquimicas y Farmaceuticas, Universidad Nacional de Rosario, Suipacha 531, S2002LRK Rosario, Argentina,*

[3]*The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA,*

[4]*Department of Horticulture, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706, USA,*

[5]*Potato Introduction Station, 4312 Hwy 42 Sturgeon Bay, WI 54235, USA,*

[6]*Department of Plant Pathology, University of Wisconsin and USDA-ARS, 1630 Linden Drive, Madison, WI 53706, USA,*

[7]*Department of Plant Pathology, Cornell University, Ithaca, NY 14853, USA, and*

[8]*Department of Plant Sciences, University of California, Davis, One Shields Ave, Davis, CA 95616-8780, USA*

Received 29 April 2005; revised 23 June 2005; accepted 28 June 2005.
[*]For correspondence (fax +1 510 559 6340; e-mail bbaker@socrates.berkeley.edu).
[†]These authors contributed equally to this work.

Summary

Cultivated and wild potatoes contain a major disease-resistance cluster on the short arm of chromosome V, including the *R1* resistance (*R*) gene against potato late blight. To explore the functional and evolutionary significance of clustering in the generation of novel disease-resistance genes, we constructed three approximately 1 Mb physical maps in the *R1* gene region, one for each of the three genomes (haplotypes) of allohexaploid *Solanum demissum*, the wild potato progenitor of the *R1* locus. Totals of 691, 919 and 559 kb were sequenced for each haplotype, and three distinct resistance-gene families were identified, one homologous to the potato *R1* gene and two others homologous to either the *Prf* or the *Bs4* *R*-gene of tomato. The regions with *R1* homologues are highly divergent among the three haplotypes, in contrast to the conserved flanking non-resistance gene regions. The *R1* locus shows dramatic variation in overall length and *R1* homologue number among the three haplotypes. Sequence comparisons of the *R1* homologues show that they form three distinct clades in a distance tree. Frequent sequence exchanges were detected among *R1* homologues within each clade, but not among those in different clades. These frequent sequence exchanges homogenized the intron sequences of homologues within each clade, but did not homogenize the coding sequences. Our results suggest that the *R1* homologues represent three independent groups of fast-evolving type I resistance genes, characterized by chimeric structures resulting from frequent sequence exchanges among group members. Such genes were first identified among clustered *RGC2* genes in lettuce, where they were distinguished from slow-evolving type II *R*-genes. Our findings at the *R1* locus in *S. demissum* may indicate that a common or similar mechanism underlies the previously reported differentiation of type I and type II *R*-genes and the differentiation of type I *R*-genes into distinct groups, identified here.

Keywords: *R1*, resistance gene (*R*-gene), *R*-gene clusters, type I and type II *R*-gene, sequence exchange, *R*-gene evolution.

Introduction

Potato (*Solanum tuberosum*) production suffers tremendous economic losses worldwide due to a number of persistent virus-, bacteria-, fungus-, nematode- and oomycete-borne diseases. Current disease control in the field relies almost exclusively on pesticides, adding significantly to agricultural costs and posing potential hazards to humans and surrounding ecosystems. Some wild species demonstrate durable resistance against certain pathogens. For instance, hexaploid wild potato, *Solanum demissum* displays resistance against all tested strains of *Phytophthora infestans*. However, when introduced into cultivated potato as a single resistance locus, these resistances have been shown to be rapidly overcome by new strains of *P. infestans* (Wastie, 1991). Thus it has been postulated that durable resistance to a certain pathogen may require simultaneous introduction of several resistance (*R*) genes or alleles, each conferring resistance to different strains.

Chromosome V of wild and cultivated potato has been shown to contain a variable set of *R*-genes encoding resistance to a diverse group of pathogens, including viruses, oomycetes, nematodes and insects (reviewed by Grube *et al.*, 2000). The region of the short arm between RFLP markers GP21 and GP179 harbors genes encoding resistance to *P. infestans* (*R1*; Leonards-Schippers *et al.*, 1992; Meksem *et al.*, 1995); extreme and hypersensitive resistance to Potato virus X (*Rx2* and *Nb*; De Jong *et al.*, 1997; Ritter *et al.*, 1991); quantitative resistance loci (QRLs) against nematode species *Globodera pallida* (*Gpa* and *Gpa5*; Bryan *et al.*, 2002; Kreike *et al.*, 1994; Rouppe van der Voort *et al.*, 1998), *G. pallida* and *Globodera rostochiensis* (*Gpr1*; Rouppe van der Voort *et al.*, 1998); and major QRLs against *P. infestans* (*phyt 3*, *Pi01* and *Pi0*; Leonards-Schippers *et al.*, 1994; Oberhagemann *et al.*, 1999; Visker *et al.*, 2003). A QRL involved in trichome-mediated insect resistance was also mapped in this region (Bonierbale *et al.*, 1994). The *Rx2* and *R1 R*-genes have been cloned, and both encode coiled-coil–nucleotide binding site–leucine-rich repeat (CC–NBS–LRR) proteins (Ballvora *et al.*, 2002; Bendahmane *et al.*, 2000). Recently, the *Bs4* gene encoding a putative Toll-Interleukin 1-Receptor (TIR–NBS–LRR) protein against bacterial spot caused by *Xanthomonas campestris* pv. *vesicatoria* in tomato was cloned from the syntenic region of tomato chromosome V, corresponding to the potato chromosome V region between markers GP21 and GP179 (Schornack *et al.*, 2004).

The presence of several *R*-genes within the potato chromosome V disease-resistance region is not surprising, given that clustering of resistance loci has been observed in many plant genomes (Hulbert *et al.*, 2001; Meyers *et al.*, 2003; Michelmore and Meyers, 1998). However, the functional and evolutionary significance of the clustered arrangement of *R*-genes remains unclear. Many *R*-gene clusters comprise tandem arrays of homologous *R*-gene sequences belonging to the same NBS–LRR-encoding gene family. *R*-genes found within the same cluster can confer resistance to multiple pathogens as well as multiple variants of a single pathogen (Botella *et al.*, 1998; Ellis *et al.*, 1999; Kesseli *et al.*, 1994; van der Vossen *et al.*, 2000). Clustered *R*-gene sequences may be tightly organized with little intervening sequence, such as the *RPP5* cluster in *Arabidopsis* (Noël *et al.*, 1999), or they may be spread over several megabases, as in the *RGC2* locus in lettuce (Meyers *et al.*, 1998a). Most studies on *R*-gene clusters have examined only a single haplotype, and limited comparative analyses have been performed to examine sequence variation of *R*-gene clusters across different haplotypes (Noël *et al.*, 1999; Seah *et al.*, 2004; Song *et al.*, 2003). A study of the *RPP5* cluster in *Arabidopsis*, one of the few clusters subjected to comparative analysis, revealed low co-linearity between different ecotypes. In this cluster, the locations of the most closely related homologues are inconsistent, and intergenic regions show low sequence similarity (Noël *et al.*, 1999), suggesting that *R*-gene clusters evolve differently from other parts of the genome.

It has been suggested that the clustering of *R*-genes may facilitate sequence exchanges between homologues, and thus rapidly generate novel chimeras (Hulbert *et al.*, 2001). Sequence exchanges between paralogues have been detected in many *R*-gene clusters (Caicedo *et al.*, 1999; Cooley *et al.*, 2000; Dodds *et al.*, 2001; Ellis *et al.*, 1999; McDowell *et al.*, 1998; Meyers *et al.*, 1998a; Noël *et al.*, 1999; Parniske *et al.*, 1997; Song *et al.*, 1997; Sun *et al.*, 2001; Van der Hoorn *et al.*, 2001). A previous study of *RGC2 R*-genes (*Dm3* homologues) in lettuce provided evidence that *R*-genes can be divided into two types, based on the frequency of sequence exchanges (type I and type II; Kuang *et al.*, 2004).

Type I resistance genes are characterized by frequent sequence exchanges between paralogues that obscure their allelic/orthologous relationships with homologues from different genotypes/species. Sequence exchanges among type I *RGC2*s in lettuce homogenized the intron sequences embedded in the LRR-encoding regions, but did not homogenize the LRR-coding sequences flanking the introns (Kuang *et al.*, 2004). This phenomenon can probably be explained by diversifying selection on the solvent-exposed LRR residues, as evidenced by an excess of non-synonymous over synonymous nucleotide substitutions. Diversifying selection is not uncommon; in fact it has been detected in almost all resistance-gene families (McDowell *et al.*, 1998; Meyers *et al.*, 1998b; Mondragon-Palomino *et al.*, 2002; Parniske *et al.*, 1997). The frequent sequence exchanges among type I *RGC2* genes in lettuce resulted in

the generation of numerous gene chimeras in natural populations (Kuang *et al.*, 2004).

In marked contrast with fast-evolving type I genes, type II *RGC2* resistance genes appear to evolve slowly, show no evidence of sequence exchange between paralogues, and reveal high conservation among different genotypes of a species and related species (Kuang *et al.*, 2004). While the underlying cause for differentiation of type I and II *RGC2* genes in lettuce remains unknown, Kuang *et al.* (2004) explored two hypotheses, one based on structure and the other based on function, to account for their formation and maintenance.

The structural hypothesis postulates that the isolation of type II *RGC2* genes was caused by DNA structures of *R*-gene homologues or their flanking regions that prohibited pairing with other *R*-gene homologues. Such a structure might arise from the duplication or translocation of a type I gene into a region that prevents pairing, or the insertion of a long sequence (such as a transposable element) near or within the gene. Current type II genes may be unable to undergo sequence exchange with paralogues due to accumulation of recurrent mutations.

The second hypothesis, based on function, speculates that type I genes gave rise to type II genes, and that the indispensable role assumed by type II genes (resistance specificity) ensured their conservation. Under this hypothesis, gene conversion of type II genes by their paralogues should not be observed, although gene conversion in the reverse direction (gene conversion of type I by type II) would be observed, as such conversion would not alter the sequence of the type II genes themselves. The functional hypothesis was found to be inconsistent with the results of the lettuce *RGC2* analysis. There was little evidence for gene conversions in either direction. In addition, some type II genes were found to be deleted or absent in some genotypes, challenging the indispensability of their function (Kuang *et al.*, 2004).

While type I and II *RGC2* genes in lettuce can be distinguished based on their characteristic high or low frequency exchange between paralogues, it remains to be seen whether *R*-gene clusters in other species adhere to such a subdivision. Analysis in other species may help discern the possible role of local chromosome sequence or structure in differentiation of type I and type II genes.

We chose to study the organization and evolution of an *R*-gene cluster in *Solanum* using wild selfing allohexaploid potato *S. demissum* ($2n = 6x = 72$). It is believed that two of the three genomes in *S. demissum* are related more to each other than to the third genome (reviewed by Spooner *et al.*, 1995). At least one of the three genomes in *S. demissum* is similar to the genomes in autotetraploid potato *S. tuberosum*. However, the ancestral diploid or tetraploid of *S. demissum* is still under debate. *Solanum demissum* has been the major source for late blight-resistance genes in cultivated potato, and at least 11 *R*-genes (*R1*–*R11*) conferring race-specific resistance against *P. infestans* were introgressed from *S. demissum* into cultivated potato (Malcolmson and Black, 1966; Shaw, 1991; Umaerus and Umaerus, 1994). The current study focuses on the *R1* cluster on the short arm of chromosome V. Physical maps were constructed for each of the three haplotypes (one for each genome) of an *R1*-containing *S. demissum* accession with the incorporation of a genetic map constructed in an *R1*-containing *S. tuberosum* cultivar. We present here a sequence comparison of the three haplotypes of the *R1* resistance-gene cluster, and comprehensive analysis of the organization and evolution of the region's *R1* resistance-gene candidates (RGCs). The *R1* locus shows substantial variation among the three genomes. We determined that the *R1* homologues in the region fall into three distinct groups of fast-evolving type I *R*-genes based on frequent sequence exchanges within groups, but not among groups. Our findings at the *R1* locus in *S. demissum* may indicate that a common or similar mechanism underlies the previously reported differentiation of type I and II *R*-genes and the differentiation of type I *R*-genes into distinct groups, identified here.

Results

Identification of BAC clones spanning the R1 *region*

A genomic bacterial artificial chromosome (BAC) library was constructed for *S. demissum* accession number PI161729. The library is composed of 397 056 BAC clones with an average insert size of 125 kb, which is equivalent to 17-fold coverage of hexaploid *S. demissum* (see Experimental procedures).

BAC clones spanning the *R1* region were identified and assembled into contigs using standard chromosome-walking strategies. A genetic map was generated using a mapping population derived from a cross between the tetraploid potato *S. tuberosum* cultivars Kennebec (*R1*) and Katahdin (*r1*; see Experimental procedures). The genomic BAC library of *S. demissum* was hybridized with probes derived from two markers proximal to resistance gene *R1*: GP179 and AFLP1 (Meksem *et al.*, 1995). Two rounds of chromosome walking were performed using markers 123A20T; 311F09S to build a contig proximal to the *R1* gene between markers 695L19T; and GP179. The BAC library was also screened using markers SPUD237, 46R, 46F, GM637 and 98R (De Jong *et al.*, 1997; Marano *et al.*, 2002; M.R.M., unpublished data) and 5′ R1, developed from *R1* sequence data (Ballvora *et al.*, 2002). Several iterations of chromosome walking were performed to extend the BAC contigs to the distal side of the *R1* gene using markers 132D05S, 160O02T and 780I15S (Figure 1; Table S1).

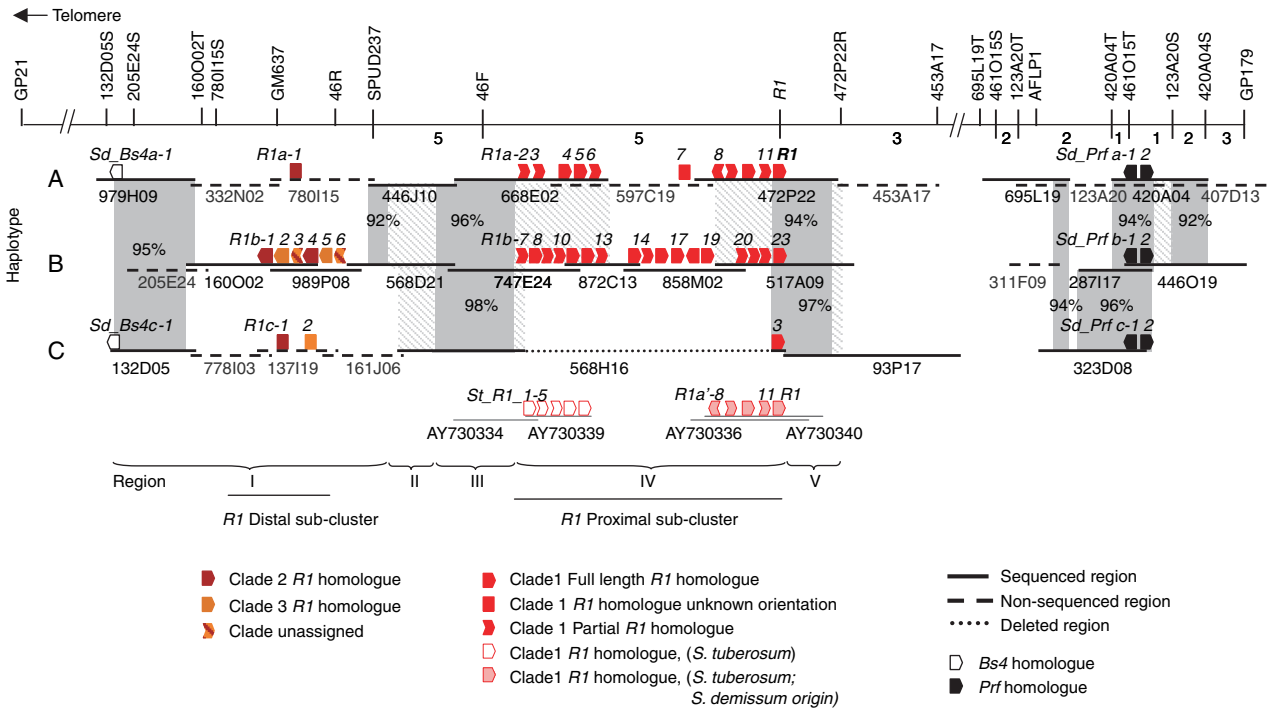

Figure 1. Integrated genetic and physical maps of the *R1* locus on chromosome V of *Solanum demissum*.
Molecular marker locations are indicated on the molecular map (top, thin solid line) and those used for genetic mapping are indicated by marks that span the line. Numbers below the line indicate the number of recombination events between markers among 953 F_1 progeny segregating for *R1* resistance. Physical maps composed of the BAC contigs for haplotypes A–C are drawn as thicker solid lines (sequenced BACs) or thick dashed lines (non-sequenced BACs). The names of BAC clones and their approximate relative positions are indicated. Dark gray-shaded areas between *S. demissum* genotypes/haplotypes designate high sequence conservation; percentage nucleotide identities are indicated. Light gray-shaded areas indicate regions of very low co-linearity. *R*-gene homologues are indicated in red/orange (*R1*), white (*Bs4*), and black (*Prf*) filled boxes with orientation indicated where known. Four BACs comprising two contigs from *Solanum tuberosum* strain P6/210 aligned with the *S. demissum R1* proximal sub-cluster region are shown below the *S. demissum* contigs. *R1* homologues present on two BACs (GenBank AY730334 and AY730339) of one contig are shown as open boxes, and are probably derived from the *S. tuberosum* genome. A second contig composed of two BACs (GenBank AY730336 and AY730340) shows *R1* homologues as filled light-red boxes, and is nearly identical in sequence to that of *S. demissum* haplotype A presented in this study. Locations of regions I–V, and distal and proximal *R1* sub-clusters, are indicated below the maps. Figure not drawn to scale.

Identification of three haplotypes at the R1 *locus of* S. demissum

Southern hybridization, PCR-based marker analysis, standard agarose-based BAC fingerprinting (Marra *et al.*, 1997), and a high-throughput fingerprinting capillary-based approach (Luo *et al.*, 2003) were used to fingerprint 1123 *S. demissum* BAC clones identified in the interval between markers 132D05S and GP179. Three continuous BAC contigs between markers 132D05S and 453A17T were assembled, one for each of the three genomes in *S. demissum* (Figure 1). Three contigs were also assembled between markers 695L19T and GP179. The contigs for the *R1* locus of each homologous chromosome V are referred as 'haplotypes' in this study. BACs bearing marker alleles in coupling with the resistance gene *R1* were designated as haplotype A (Figure 1; Table 1; see Experimental procedures). A gap of unknown length is present between markers 453A17T and 695L19T in haplotype A. The two other assembled contigs between markers 132D05S and 453A17T were designated haplotypes B and C. Note that the contigs shown as B and C between markers 695L19T and GP179 could be assigned to either haplotype B or haplotype C. The representative BACs for each haplotype are shown in Figure 1.

Sequence of BAC clones at the R1 *locus*

To provide a basis for comparative analysis of the structure and evolution of the *R1* locus, and to determine the gene content of the region, minimally overlapping BACs from each haplotype of *S. demissum* were selected for shotgun sequencing. For haplotype A, BACs 979H09, 446J10, 668E02, 472P22, 695L19 and 420A04 were selected, corresponding to 691 kb of non-overlapping sequence. In haplotype B, a total of 919 kb non-overlapping was sequenced, which included a 668 kb continuous region covered by overlapping BACs 160O02, 989P08, 568D21, 747E24, 872C13, 858M02 and 517A09. In haplotype C, four BACs (132D05, 568H16, 93P17 and 323D08) were sequenced, totaling 559 kb non-overlapping sequence. All overlapping sequences are identical, an indication of accurate grouping of the three haplotypes. The relative position of each sequenced BAC is shown in

Table 1 *Solanum demissum* BAC clones sequenced

BAC clone	Size (bp)	GenBank accession	Haplotype	Region[a] (size, kb)
PGEC979H09	110 954	AC151803	A[b]	I (0–110)
PGEC446J10	102 692	AC149288	A	I (0–21), II (21–88), III (88–102)
PGEC668E02	142 575	AC144791	A	III (0–58), IV (58–142)
PGEC472P22	136 150	AC151815	A	IV (52–136)[c], V (7–52)
PGEC695L19 (PGEC407)	73 587	AC149487	A	
PGEC420A04 (PGEC219)	124 895	AC135288	A	
PGEC160O02	125 813	AC150162	B	I (0–125)
PGEC989P08	121 915	AC146506	B	I (0–115), II (115–121)
PGEC568D21	119 525	AC149287	B	II (63–119)[c], III (0–63)
PGEC747E24	92 340	AC149267	B	III (72–92), IV (0–72)[c]
PGEC872C13	100 564	AC149266	B	IV (0–100)
PGEC858M02	114 450	AC149265	B	IV (0–114)
PGEC517A09	133 983	AC149301	B	IV (0–64), V (64–106)
PGEC287I17	78 626	AC142505	ND[d]	
PGEC446O19	122 450	AC145120	ND	
PGEC132D05	118 497	AC154033	C	I (0–118)
PGEC568H16	146 091	AC149291	C	II (113–143)[c], III (38–113), IV (11–38), V (0–11)
PGEC093P17	200 238	AC149290	C	V (0–36)
PGEC323D08	106 746	AC139840	ND	

[a]Regions I–V as described in text (Figure 1).
[b]Haplotype A BACs were identified with marker alleles in coupling with the *R1* resistance gene.
[c]BAC nucleotide numbering in reverse orientation.
[d]ND, Not determined, BACs belong to either haplotype B or C.

Figure 1. The complete sequences of BAC clones have been deposited with GenBank and the corresponding accession numbers are listed in Table 1. A total of 84 genes on genome A (691 kb), 125 on genome B (919 kb) and 77 on genome C (559 kb) were identified through a manual annotation process (see Experimental procedures). A total of 197 genes could be assigned a predicted function; the 89 remaining genes encode a hypothetical protein. The gene density was measured to be one gene per 7.6 kb. Based on alignment with tomato expressed sequence tags (ESTs), 40 of the genes are expressed. The low number of corresponding ESTs is, in part, due to an incomplete number of entries in the databases (189 000) and a very low representation of resistance gene candidates in the EST databases, as resistance genes generally have low expression. Annotation of the 2.17 Mb described above is summarized in Table S2, and can be viewed at BAC level at http://www.tigr.org/tigr-scripts/tdb/potato/BAC_annotation/bac_display.pl.

The R1 *cluster contains both conserved and divergent regions*

The sequenced regions in each haplotype were compared to investigate structural variations among haplotypes. Comparisons were first performed using DOTTER, a dot-matrix program (Sonnhammer and Durbin, 1995), followed by sequence alignment where possible. Surprisingly, some regions within the *R1* cluster show very low co-linearity between haplotypes, while other regions are highly conserved, exhibiting nucleotide identities >92% among the three haplotypes. For convenience, five regions (I–V) have been assigned to the *R1* locus based on sequence conservation among different haplotypes (Figure 1; Table 1).

Although similar in size (approximately 60 kb) in all three haplotypes, region II is highly divergent (Figure 1), and a plot analysis revealed extensive re-arrangement (data not shown). The most intriguing finding in region II is a 3.5 kb fragment that is fully or partially repeated approximately 10 times in haplotypes A (BAC446J10) and B (BAC568D21). The end of each repeat contains a 340 bp inverted repeat encoding a deduced product that shares homology with the F-box protein family. Comparisons of region II sequences of all three haplotypes showed that approximately half the sequences share no significant similarity (are unrelated).

The other divergent region, IV, shows remarkable variation in length, with approximately 200 kb in haplotype A, approximately 320 kb in haplotype B, and only approximately 27 kb in haplotype C (Figure 1; Table 1). A total of 168 kb was sequenced from region IV in haplotype A, and the remaining approximately 35 kb is present in the mapped but unsequenced BAC clone 597C19 (Figure 1). Comparisons between haplotypes A and B revealed very low co-linearity in region IV (Figures 1 and 2). The *R1* homologous genes and transposable elements were the only sequences found to have significant similarity among the three haplotypes (Figure 2).

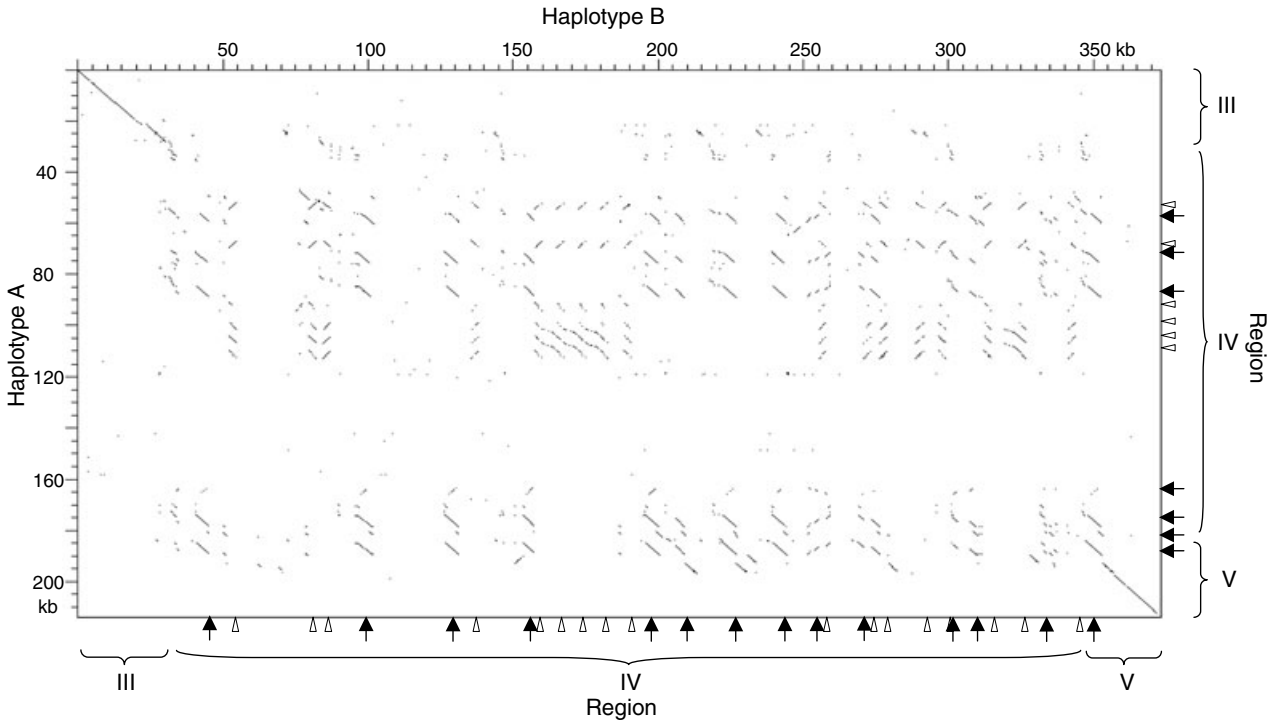

Figure 2. A dot plot revealing sequence variation between haplotypes A and B within the proximal sub-cluster (region IV) of the *R1* locus.
Included for comparison are 20–30 kb of conserved flanking sequence from regions III and V. Arrows indicate positions of some *R1* homologues. Small open triangles indicate members of a transposon family.

Copy number of R1 *homologues varies among haplotypes in dynamic region IV*

In addition to the low level of co-linearity, region IV exhibits a high degree of variation in copy number of *R1* homologues among haplotypes. Haplotype A has one gene with a nucleotide sequence identical to the previously cloned *R1* resistance gene (Ballvora *et al.*, 2002), located at the proximal end of this cluster (Figure 1). In addition to the *R1* gene, region IV of haplotype A has 10 other *R1* homologues, nine of which were sequenced. The tenth *R1* homologue (*R1a-7*) was detected through Southern analysis of the non-sequenced portion in BAC clone 597C19 (Figure 1). In haplotype B, this region was completely sequenced and was found to contain 17 *R1* homologues. In marked contrast, region IV of haplotype C has only a single *R1* homologue, *R1c-3* (Figure 1).

R1 *homologues in the proximal sub-cluster of different haplotypes are highly divergent*

The 29 *R1* homologues from region IV form the proximal sub-cluster of the *R1* locus, and 28 of the 29 homologues were sequenced in this study. Fourteen homologues are partial gene sequences varying from 130 bp (*R1b-9*) to 3.5 kb (*R1b-20*), and the remaining 14 are full-length *R1* homologues. The 14 full-length *R1* homologues encode the same structural domains and exon–intron boundaries as those of the *R1* resistance gene (Ballvora *et al.*, 2002). Of the 14 full-length copies, four are interrupted by transposable element insertions in coding regions; three have frame shifts due to short insertions/deletions; and one has a premature stop codon. Thus at least 22 of the 28 *R1* homologues from the proximal sub-cluster are predicted pseudogenes.

The 21 longest homologues (>1.5 kb) of the proximal sub-cluster were chosen for further sequence analysis. They exhibit nucleotide identities ranging from 87.2 to 100% (92.1 $\pm$ 1.7%). The two most similar pairs, *R1a-4* and *R1a-10*; and *R1b-14* and *R1b-17*, exhibit 100 and 99.4% nucleotide identity, respectively. *R1a-4* and *R1a-10* are highly conserved paralogues, as are *R1b-14* and *R1b-17*: in other words they are homologues from the same haplotype, rather than alleles (genes located at the same position in different haplotypes) (Figure 1). The third most similar pair (*R1b-23* and *R1c-3*) shows 97.1% nucleotide identity. All other *R1* homologues at the proximal sub-cluster are equally divergent from each other, exhibiting nucleotide identities of 87.2–95.7%. The absence of highly similar *R1* homologues from different haplotypes is consistent with the low co-linearity among haplotypes at the proximal sub-cluster (Figure 2). No alleles/orthologues of the *R1* homologues could be identified among the three haplotypes based on their gene sequence or position. Both findings suggest that the proximal *R1* sub-cluster is prone to re-arrangement and/

or that there is strong selection for such genomic re-arrangements in *R*-gene clusters. These *R*-gene re-arrangements are postulated to be associated with generation of novel resistance genes.

R1 *homologues from the proximal sub-cluster show frequent sequence exchanges and nearly identical intron 1 sequences*

To determine whether the similar genetic divergence among the majority of *R1* homologues at the proximal sub-cluster was caused by frequent sequence exchanges, we investigated sequence-exchange events among these *R1* homologues. A total of 14 sequence exchanges were detected using GENECONV ($P < 0.05$), and visual inspection uncovered three additional sequence-exchange events. The highly similar genes (*R1b-14* and *R1b-17*) in haplotype B are the result of a recent 4 kb gene conversion. The 4 kb conversion tract exhibits 99.7% nucleotide identity. The second largest exchange tract was found between the 3′ ends of *R1b-23* in haplotype B and *R1c-3* in haplotype C. Other sequence-exchange tracts were much shorter, varying from 103 to 689 bp. Of the 17 sequence exchanges, seven occurred among homologues of different haplotypes. As a consequence of frequent sequence exchanges, most *R1* homologues from the proximal sub-cluster show similar nucleotide identity with each other (87.2–95.7%). These homologues from the proximal sub-cluster form a tight clade in a distance tree of all *R1* homologues (Figure 3a).

Of the 28 *R1* homologues sequenced from the proximal sub-cluster, 19 have the previously reported intron 1 located within the LRR-coding domain of the *R1* gene (Ballvora *et al.*, 2002; Figure 3c), and the other nine have deletions spanning this intron. The intron 1 sequences from the 19 homologues vary from 77 to 95 bp in length, with the majority (14/19) having a 95-bp-long intron. The intron sequences among these homologues are highly conserved, with nucleotide identities ranging from 89.4 to 100% (95.5 ± 2.5%). This is significantly higher than the nucleotide identities between the coding regions of corresponding homologues

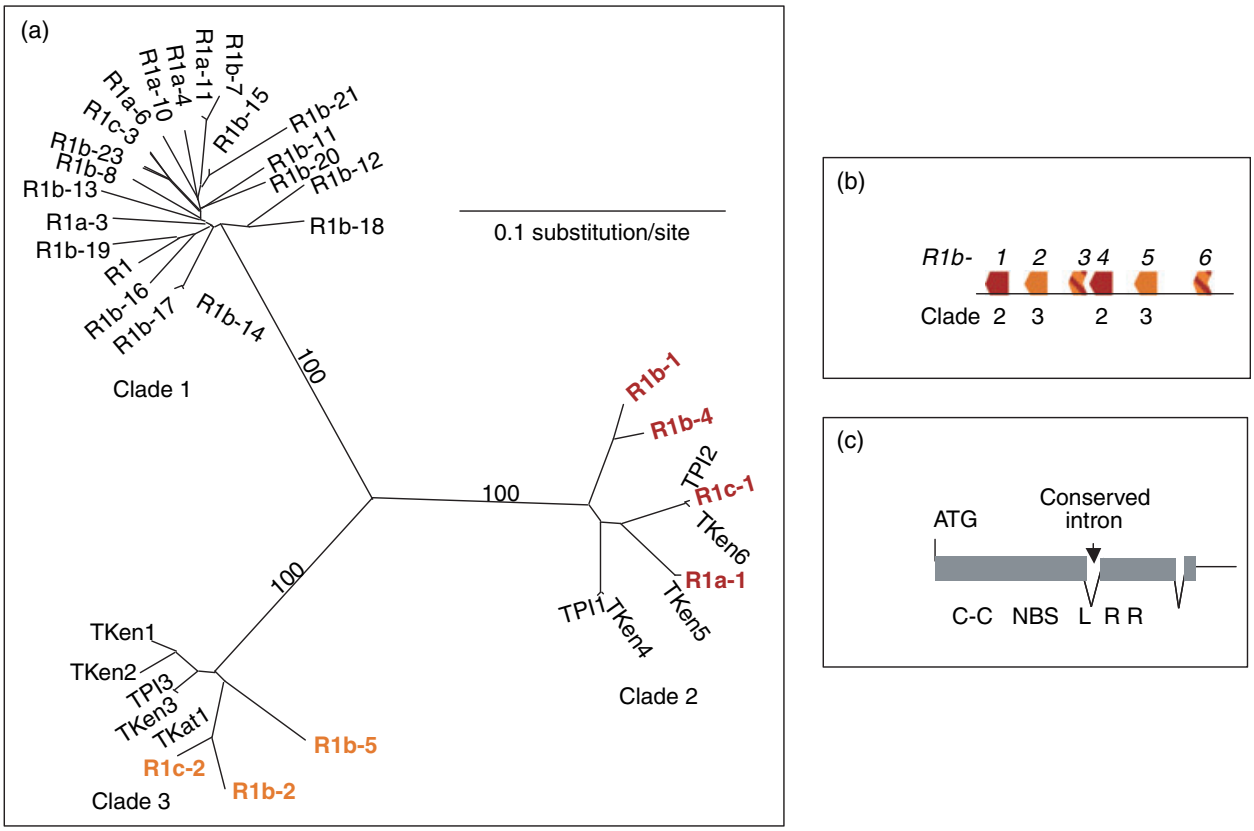

Figure 3. Type I *R1* homologues.
(a) Distance tree showing differentiation of the three groups of type I *R1* homologue. Numbers on the node show bootstrap values. Genes beginning with R1 are from the *R1* locus of *Solanum demissum*; genes beginning with T were PCR amplified from *Solanum tuberosum*: TKen1-6 from cv. Kennebec; TKat1 from cv. Katahdin; TPI1-3 from cv. PI303148. All *R1* homologues of clade 1 were isolated from *S. demissum* and are located within the proximal sub-cluster, while *S. demissum* genes in clades 2 and 3 are located within the distal sub-cluster. The color of gene names in clades 2 and 3 corresponds to that of genes in Figure 1.
(b) *R1* homologues of clades 2 and 3 are interspersed in the *R1* distal sub-cluster of haplotype B.
(c) Structure of resistance gene *R1* showing the location of intron 1 embedded within the coding region of the LRR protein domain. Intron 1 sequences are highly conserved among members of each clade.

(92.1 ± 1.7%) (*t*-test, $P < 0.001$). Similar results were obtained when synonymous mutation rates (K_a) instead of nucleotide identities between coding sequences were used (*t*-test, $P < 0.001$). The relatively high nucleotide identities for the intron sequences were not caused by the frequent sequence exchanges, and homogenized intron sequences among the *R1* homologues at the proximal sub-cluster suggest that they are type I resistance genes (Kuang *et al.*, 2004).

R1 *homologues of the distal sub-cluster form two distinct groups of type I genes*

In addition to the 17 *R1* homologues located at the proximal *R1* sub-cluster of haplotype B, six *R1* homologues, *R1b-1* to *R1b-6*, were found in a 78-kb sequence within region I, and are members of the distal sub-cluster (Figure 1). The distal *R1* sub-cluster is located 190 kb telomeric to the proximal *R1* sub-cluster in haplotype B. Unfortunately, the haplotype A and C BAC clones assembled in this region were not sequenced; however, Southern hybridization revealed the presence of a single *R1* homologue, *R1a-1*, within the distal sub-cluster of haplotype A on BAC 780I15 (Figure 1) and two, *R1c-1* and *R1c-2*, in haplotype C BAC 137I19 (Figure 1). Fragments of these three *R1* homologues were PCR amplified and sequenced (GenBank AY955222, AY955223 and A955224 respectively). Therefore a total of nine *R1* homologues of the distal sub-cluster were obtained. Seven of these homologues (*R1a-1*, *R1b-1*, *R1b-2*, *R1b-4*, *R1b-5*, *R1c-1*, *R1c-2*) are >1.5 kb in length and were chosen for further characterization. The *R1* homologues of the distal sub-cluster appear to be divergent from those at the proximal sub-cluster, with nucleotide identities ranging from 61.8 to 69.5% between homologues from the two sub-clusters.

To better understand the evolution of *R1* homologues of the distal sub-cluster, homologues derived from the distal sub-cluster of cultivated potato were compared with the *R1* homologues from *S. demissum*. A total of 10 *R1* homologues were PCR-amplified and sequenced (GenBank AY935701–AY935710) from three different *S. tuberosum* cultivars: six (*TKen1-6*) from cv. Kennebec; one (*TKat1*) from cv. Katahdin; and three (*TPI1-3*) from cv. PI303148 using primer combinations 1-4-F/R and 2-5-F/R, which specifically amplify the *R1* homologues within the distal sub-cluster. Based on the specificity of the PCR primers, the 10 *S. tuberosum* fragments are assumed to be located within the distal *R1* sub-cluster. The 10 potato fragments, in addition to the seven *S. demissum R1* homologues of the distal sub-cluster, formed two clades (clades 2 and 3) in a distance tree comprised of all *R1* homologues (bootstrap value = 100; Figure 3a).

Clade 2 in the distance tree contains *R1a-1* from haplotype A; *R1b-1* and *R1b-4* from haplotype B; *R1c-1* from haplotype C; and five fragments, *Tken4–6*, *TPI1* and *TPI2* amplified from cultivated potato. Six sequence exchanges were detected among the *R1* homologues in clade 2 (Figure 3a). The nucleotide identities among the nine homologues in clade 2 varied from 91 to 99%. The intron 1 sequences of these homologues are significantly more conserved than their coding sequences (*t*-test, $P < 0.001$). Similarly, clade 3 contains eight fragments; homologues in clade 3 exhibit 91–99% nucleotide identity with each other; and eight sequence exchanges were discovered among them. The homologues in clade 2 exhibited <70% nucleotide identity with those in clade 3, and no sequence exchanges were found to have occurred between homologues from these two clades. These data suggest that clades 2 and 3 each represent a distinct group of type I *R1* homologues; thus there appear to be three groups (the three clades in Figure 3a) of type I genes at the *R1* locus, with frequent sequence exchanges occurring among genes within each group, but no sequence exchanges among genes of different groups.

Diversifying selection on the solvent-exposed residues within the LRR region of R1 *homologues*

The sequences of the *R1* homologues within each of the three type I groups were not homogenized by frequent sequence exchanges, probably due to diversifying selection on the solvent-exposed residues encoded within the LRR region. Sites under diversifying selection were investigated using PAML software (Yang, 1997; Yang *et al.*, 2000). All *R1* homologues obtained from the *R1* locus of *S. demissum* were included for analysis. A likelihood-ratio test was performed by comparison of likelihood models M7 and M8 in program CODEML (Yang *et al.*, 2000) and the results show that the *R1* gene family has been under diversifying selection ($x^2 = 104$; $P < 0.001$).

Prf *and* Bs4 *homologues flanking the* R1 *cluster in the three* S. demissum *haplotypes are type II genes*

Two CC–NBS–LRR-encoding genes, *S. demissum Prf-1 and Prf-2* (*Sd_Prf-1* and *Sd_Prf-2*), homologous to the tomato *Prf* resistance gene (Salmeron *et al.*, 1996), were discovered at the *R1* cluster in each of the three haplotypes in *S. demissum*. *Sd_Prf-1* and *Sd_Prf-2* are positioned in opposite orientation 3 kb apart in haplotypes B and C, and 20 kb apart in haplotype A. Only a partial sequence of *Sd_Prf-2* in haplotype C, *Sd_Prf-c-2*, is present in BAC clone 323D08 (Figure 1). The *Sd_Prf-1* and *Sd_Prf-2* homologues are located in the same relative position in each haplotype, maintain obvious allelic relationships, and are highly conserved. The three *Sd_Prf-1* genes exhibit nucleotide identities of 97%, though several large insertions/deletions were observed in the copy in haplotype A. The three *Sd_Prf-2* genes exhibit nucleotide identities ranging from 97 to 99%. The high degree of

conservation among alleles and lack of sequence exchanges among paralogues indicate that *Sd_Prf-1* and *Sd_Prf-2* homologues at the *R1* locus are type II genes.

The structures of *Sd_Prf-1* and *Sd_Prf-2* differ from each other and from the *Prf* gene in tomato. Tomato *Prf* has a 5475 bp coding region, yet the *Sd_Prf-2* genes in haplotypes A and B are predicted to contain more than 7.7 kb of coding sequence (*Sd_Prf-c-2* is only partially represented). The expansion of *Sd_Prf-a-2* and *Prf-b-2* is mainly due to duplication of their CC-encoding regions, resulting in three copies of the CC domain ($P < 0.1$; Lupas, 1996). One premature stop codon was observed in *Sd_Prf-a-2* (at the 1314th codon) and in *Prf-b-2* (at the 1274th codon), thus both are likely to be pseudogenes. All three *Sd_Prf-1* genes are shorter than the *Sd_Prf-2* genes, and at least 1.6 kb of coding sequence upstream of the CC domain is deleted in these genes. Furthermore, the *Sd_Prf-1* genes have several deletions >200 bp and many frame-shift indels, and are probably pseudogenes. Interestingly, despite the variation in structure between *Sd_Prf-1* and *Sd_Prf-2* and their divergence from tomato *Prf*, the *Sd_Prf-1* and *-2* genes found in all three haplotypes are highly conserved, and show the types of conserved relationship previously described for slowly evolving type II *R*-genes.

Resistance-gene candidates *Sd_Bs4a-1* and *Sd_Bs4c-1* belong to the third *R*-gene family at the *R1* locus and are homologous to *Bs4*, a tomato TIR–NBS–LRR resistance gene against bacterium *X. campestris* pv. *vesicatoria*. A full-length copy of *Sd_Bs4a-1* on BAC 979H09 in haplotype A, and a partial copy of *Sd_Bs4c-1* from the corresponding region in haplotype C on BAC 132D05, were sequenced (Figure 1). *Sd_Bs4a-1* has a 1-bp insertion and a 2-bp deletion in exon 2, while *Sd_Bs4c-1* has a 349-bp insertion in exon 2, therefore both *Bs4* homologues appear to be pseudogenes. Regardless of the frame-shift insertion/deletions, the putative coding regions of *Sd_Bs4a-1* and *Sd_Bs4c-1* exhibit 96% nucleotide identity with each other, and approximately 90% nucleotide identity with the tomato *Bs4* resistance gene. It is likely that the *Bs4* homologues at the *R1* locus also belong to the type II class of *R*-genes.

Discussion

Resistance genes and traits at the R1 *locus*

Three families of NBS–LRR-encoding RGCs were discovered at the *R1* resistance-gene cluster in *S. demissum*, with the majority of homologues belonging to the *R1* gene family. One *R1* homologue identified in this study is identical to the *R1* resistance gene previously cloned from a chromosome segment of *S. tuberosum* P6/210 that was introgressed from *S. demissum* (Ballvora *et al.*, 2002).

Structural comparisons of the *R1* homologues within the three haplotypes of *S. demissum* showed that they can be divided into three groups of independently evolving type I *R*-genes, and that most homologues within each group show >90% nucleotide identity. The finding that the type I *R1*-gene homologues can form independently evolving groups suggests that differentiation of these type I groups parallels the differentiation of type I and type II *R*-genes reported previously for the *RGC2* cluster in lettuce.

Comparison of the *S. demissum R1* region sequences described in this study with publicly available sequences (GenBank accessions AY730336 and AY730340) show that 90 738 bp of *S. tuberosum* strain P6/210 genomic DNA (from two overlapping BAC clones) is highly homologous to approximately 91 kb of region IV DNA of the *S. demissum* 'haplotype A' (Figure 1). The approximately 91-kb introgressed *R1* region from *S. tuberosum* and the corresponding region in haplotype A of *S. demissum* exhibited only 24 nucleotide substitutions and 10 indels (nine 1-bp indels and one 2-bp indel). The high degree of conservation suggests that these sequences are derived from two nearly identical alleles of the *R1* gene region of *S. demissum* (Figure 1). The *R1* gene in *S. tuberosum* P6/10 was probably introgressed from an *S. demissum* accession different from that (PI 161729) described here.

We found that alleles of several markers in the region proximal to the *R1* cluster, that are linked in coupling with the *R1* gene (461O15S, 123A20T, 420A04T and GP179) in the *S. tuberosum* cv. Kennebec *R1* mapping parent, are also present in BACs of the *S. demissum* haplotype A contig. These results suggest that an *S. demissum* chromosome V segment proximal to the *R1* gene cluster, between markers 461O15S and GP179, was introgressed into cv. Kennebec along with the *R1* resistance gene.

Members of two other *R*-gene families, *Prf* and *Bs4*, closely linked with the *R1* homologues, were identified in this study. The tomato resistance gene *Bs4* is a TIR–NBS–LRR-encoding gene, located within the syntenic region of the *Solanum R1* cluster. In this study two *Bs4* homologues were found, one in haplotype A and the other in haplotype C. They are predicted to be pseudogenes as they have a premature stop codon and/or frame-shift deletions. Further mapping and BAC sequence analyses will identify whether additional members of these or other families are located in adjacent chromosomal regions. Our preliminary data suggest that there are several additional *R*-gene families in the region between markers GP21 and SPUD237, immediately distal to the region investigated in this study (Marano *et al.*, 2002; data not shown).

In addition to the *R1*-mediated resistance, other disease-resistance traits, including QRLs for *P. infestans* and nematode resistance, have been mapped near the *R1* locus (see Introduction). The structures and sequences of the RGC families described in this study serve as a resource for assessing the role of related RGCs in mediating disease-resistance traits in Solanaceae species. For example, the

high nucleotide identity of the *R1* homologues within each of the three groups described here could be exploited in gene-silencing approaches to investigate if any or all members of a group encode resistance traits previously mapped to the region. This study also provides information for further comparative genomic analyses in the Solanaceae to investigate the role of local chromosome structure on generation and maintenance of independently evolving groups of *R*-genes.

Similar structures of R1 *locus regions III and IV in* S. tuberosum *and* S. demissum

Comparison of a total of 158 kb of sequence derived from two other overlapping *S. tuberosum* P6/210 BAC clones (GenBank AY730334 and AY730339) with the sequences of *S. demissum R1* locus showed that approximately 61 kb of the *S. tuberosum* sequence is similar (approximately 96% nucleotide identity) to region III in all three *S. demissum* haplotypes (Figure 1). The remaining approximately 97 kb is located proximal to region III and therefore falls into region IV (Figure 1). A total of five *R1* homologues are present in the 97 kb region IV sequence: two full-length; one partially covered by the BAC insert; one with a large deletion; and one with a large insertion. The *S. tuberosum* 97-kb sequence shows very low sequence co-linearity with *S. demissum* and, as was observed for region IV in *S. demissum*, only the *R1* homologous sequences showed significant similarity to the *S. demissum* region IV sequence.

Nucleotide identity, conserved intron sequences and frequency of sequence exchanges suggest that the three largest homologues in the 97-kb *S. tuberosum* sequence belong to the same group of type I genes as those located at the proximal *R1* sub-cluster in *S. demissum* (clade 1, type I genes). No concerted evolution was observed: the three genes from S. *tuberosum* were as similar to each other as to homologues from region IV of *S. demissum*; thus, conserved type I *R*-genes. Thus regions III and IV of *S. demissum* and *S. tuberosum* appear to have similar structures, and it is likely that the differentiation of type I *R1* genes at the proximal sub-cluster occurred prior to speciation of *S. demissum* and *S. tuberosum*.

Sequence exchanges among R-*gene homologues*

Sequence exchange between resistance-gene paralogues is considered to be the dominant mechanism for generating variations of type I resistance genes, and has been found to obscure allelic/orthologous relationships among homologues from different genotypes or related species (Kuang *et al.*, 2004). In this study, no allelic/orthologous relationships among *R1* homologues from different haplotypes could be discerned. Frequent sequence exchanges tend to homogenize members of a gene family; however, such homogenization was not observed either with *R1* homologues or in *R*-genes from other plant genomes, such as the *RGC2* genes in lettuce (Kuang *et al.*, 2004). The high degree of variation among *R1* paralogues may be maintained through diversifying selection.

Interestingly, frequent sequence exchanges did homogenize intron sequences, which are not subject to diversifying selection. The intron sequences of type I *R1* homologues have significantly higher nucleotide identity than their flanking coding sequences. The highly conserved intron sequences may not have a significant effect on the frequency of sequence exchanges, as they are <100 bp in length. Therefore it is likely that intron sequence conservation is the result, rather than the cause, of frequent sequence exchanges. Homogenization of intron sequences has also been observed in the *MHC* and *HLA* gene families in mammals and in type I *RGC2* resistance genes in lettuce (Cereb *et al.*, 1997; Hughes, 2000; Kuang *et al.*, 2004). Like the *R1* homologues, these gene families showed evidence for frequent sequence exchanges and for diversifying selection on coding domains of these genes.

Seven of the 17 sequence exchanges among *R1* homologues of the proximal sub-cluster occurred between different genomes. Sequence exchanges among *R1* homologues from the distal sub-cluster of different haplotypes were also discovered. It is unlikely that cross-over occurs between different haplotypes (genomes) in *S. demissum*. Therefore the observed sequence exchanges among different haplotypes could be generated through gene conversions. Alternatively, they might have occurred through recombination before speciation and have been maintained in *S. demissum*, although little is known about the relationship of the divergence time of the three genomes.

Evolution of R-*gene clusters*

The clustering of *R*-genes is believed to facilitate sequence exchanges (Hulbert *et al.*, 2001). This is a reasonable hypothesis considering that unequal cross-overs can occur among *R*-genes within a cluster, but are unlikely to occur among homologues from different loci. Unequal cross-overs among *Rp1* homologues in maize may be largely responsible for the instability of the *Rp1* cluster (Sun *et al.*, 2001). Haplotype comparisons in this study showed that the *R*-gene cluster is indeed more dynamic than its flanking non-*R*-gene regions. The proximal *R1* sub-clusters of the three *S. demissum* haplotypes vary dramatically in size and number of RGCs. Haplotype C contains only one *R1* homologue. This short haplotype may have been generated by unequal cross-over between two *R1* homologues located at the two ends of the original cluster. Interestingly, the intergenic regions (stretches of non-*R*-gene sequence between RGC repeats) do not show significant similarity with other sequences within the same haplotype, or among haplo-

types. The only significant repeated sequences in this sub-cluster are the *R1* homologues and some transposable elements. If unequal cross-overs occur, they probably occur among the sequences of *R1* homologues and consequently create new *R*-gene chimeras, as has been observed at the *Rp1* locus in maize (Sun *et al.*, 2001). The evolution of the distal *R1* sub-cluster was characterized to a lesser extent. Nevertheless, large variations in *R1* homologue number were found among the three haplotypes at this sub-cluster, with six copies in haplotype B, but only one homologue in haplotype A and two in haplotype C.

Possible common mechanism for differentiation of different groups of type I R*-genes and type I and II genes*

As described in the Introduction, type I *R*-genes are characterized by frequent sequence exchanges among paralogues that obscure allelic/orthologous relationships between homologues from different genotypes/species. Type II *R*-genes, in contrast, do not exchange sequences with paralogues and show high conservation among different genotypes and related species (Kuang *et al.*, 2004). Type I and II *R*-genes can be present at the same cluster (Dodds *et al.*, 2001; Kuang *et al.*, 2004), while an *R*-gene cluster can have only type II *R*-genes (Song *et al.*, 2003) or only type I *R*-genes (such as at the *RPP8* locus in *Arabidopsis*; H.K. and R. Michelmore, unpublished data).

In this study of *Solanum*, all *R1* homologues analyzed were determined to be type I *R*-genes, and no type II *R*-genes were discovered. However, we found that the type I *R1* homologues could be further divided into three independent, evolving groups based on the lack of sequence exchanges among members of different groups. The lack of conserved genes among the members of the different groups argues against selection for conserved function (the functional hypothesis, see Introduction) to account for differentiation of different groups of type I *R1* homologues in *S. demissum*. The lack of sequence exchanges among the three groups of type I *R1* homologues parallels the lack of sequence exchange between type I and II *RGC2* *R*-genes in lettuce. It is possible that the same mechanism responsible for differentiation of type I/type II genes is also responsible for differentiation of different groups of type I genes.

Differentiation of type I and type II R*-gene homologues may be determined by flanking sequence*

As an alternative to the functional hypothesis, we hypothesize that local chromosome structure is the main force preventing sequence exchanges among some homologues, and that local chromosome structure is therefore responsible for subsequent differentiation of different groups of type I genes (as well as differentiation of type I and type II genes). The nucleotide identities between type I and II *RGC2* genes in lettuce, and among different groups of type I *R1* homologues in potato, are <80%. While low nucleotide identity may prevent homologue pairing and subsequent sequence exchange, the original force preventing sequence exchange between newly duplicated homologues can still be explained by structural constraints, for example translocation of the homologue to a region that prevents further pairing, or transposon insertion in a flanking region. Substantial change to an intron sequence, such as a large deletion or insertion, may also prevent homologue pairing and sequence exchange.

The relative physical positions of *R*-gene homologues may have limited effects on the frequency of sequence exchanges and the differentiation of type I/II genes. The distal sub-cluster at the *R1* locus is composed of two different groups of type I genes. The two different groups of genes are interwoven in haplotype B (Figure 3b). Such an interweaving structure was not caused by duplication, as there is no evidence that *R1b-1*/*R1b-2* was duplicated from *R1b-4*/*R1b-5*. Genes *R1b-1* and *R1b-2*, as well as *R1b-4* and *R1b*-5, are separated by <10 kb, but belong to different groups of type I genes and have no sequence exchanges. This evidence suggests that the frequency of sequence exchange is not necessarily determined by physical proximity. The interweaving of type I and II genes was also observed in the *Dm3* gene family in lettuce (Kuang *et al.*, 2004). In *Arabidopsis*, *RPP8* homologues located within the same physical cluster can be either type I or type II genes, and members of the same group of type I *RPP8* homologues can be located at different loci (H.K. and R. Michelmore, unpublished data).

Pseudogenes might act as reservoir of sequences for new chimeras

Many NBS–LRR-encoding genes are characterized as 'pseudogenes' because of premature stop codons, frameshift insertions/deletions or large deletions. Interestingly, most of these pseudogenes are expressed (Meyers *et al.*, 2003; Shen *et al.*, 2002). Pseudogenes are not necessarily junk DNA in the genome, but rather may be involved in gene expression and regulation (reviewed by Balakirev and Ayala, 2003). The majority of *R1* homologues obtained from this study are annotated as pseudogenes. These gene fragments may act as a reservoir for generating new chimeras through gene conversions (Michelmore and Meyers, 1998). Whether or not the *R1* homologues with transposon insertions can still undergo sequence exchanges with other type I *R1* homologues remains unknown. Nevertheless, most pseudogenes, including those with transposon insertions, still retain the characteristics of type I genes. Most pseudogenes have not degenerated through accumulation of recurrent mutations. *R1a-8* (1383 bp) is the only sequence that might be a degenerated type I gene at the *R1* locus. Its nucleotide identity

with other *R1* homologues is <80% and it does not belong to any of the three groups of type I *R1* homologues (data not shown).

Transposable elements at resistance-gene loci

Many transposable elements were found at the *R1* locus, including retrotransposons, transposons, and miniature inverted transposable elements (MITEs; data not shown). Six transposable elements were found inserted in five *R1* homologues. Transposons were also found in *N* resistance genes in flax, *RPP5* homologues in *Arabidopsis*, and *Xa21* homologues in rice (Dodds *et al.*, 2001; Noël *et al.*, 1999; Song *et al.*, 1997). The insertion of a transposable element into the coding region of an *R*-gene apparently abolishes its function. However, subsequent excision of a transposable element may introduce novel polymorphisms (insertions) into *R*-genes (Michelmore and Meyers, 1998). The importance of such a mechanism in the evolution of *R*-genes remains unknown. The six transposable elements inserted into the five *R1* homologues in this study have a target-site duplication of 5, 7 or 8 bp. Therefore the insertion followed by excision of such transposable elements will generate frameshift insertions without contributing to *R*-gene diversity.

Transposable elements may contribute indirectly to the evolution of *R*-genes, for example through initial duplication of a single-gene locus (Michelmore and Meyers, 1998). The presence of several copies of the same transposable element in intergenic regions may provide more opportunity for unequal cross-over, and therefore cause expansion and contraction of the *R*-gene cluster. Another possible impact of transposon insertions on *R*-gene clusters is that insertions of different families of transposable elements can increase diversity in the intergenic region and prevent unequal cross-over.

Several MITE families associated with *R*-gene clusters have been described (Song *et al.*, 1998; Wei *et al.*, 2002). The role of MITEs within *R*-gene regions may vary from that predicted for transposons or retrotransposons, as MITEs are usually short (<700 bp) and thus should have limited effects on recombination. MITEs have been found linked with genes in plant genomes, located mainly in non-coding regions (Wessler *et al.*, 1995), and have been hypothesized to play important roles in regulation of gene expression (Song *et al.*, 1998; Wessler *et al.*, 1995).

Experimental procedures

Mapping population, P. infestans *infection and marker analysis*

Genetic mapping of the *R1* locus in *S. tuberosum* was performed in a cross between the tetraploid potato cultivars Kennebec (*R1*) and Katahdin (*r1*). A total of 953 F_1 plants were tested for late-blight resistance using *P. infestans* isolates US90480, genotype US-8 (mating type A2) and genotype US-1 (mating type A1). Detached leaves and/or whole plants were inoculated with sporangial suspensions from both isolates of *P. infestans* as described previously (Naess *et al.*, 2000; Vleeshouwers *et al.*, 1999). The presence or absence of a hypersensitive response was scored 5–10 days after inoculation.

Markers GP179, SPUD237 and AFLP1 were used for initial genetic mapping and BAC library screening. Twenty-four individual plants were identified with recombination events in the interval between markers SPUD237 and GP179. Resistance gene *R1* was mapped 1 cM (10 recombination events) from SPUD237 and 1.5 cM (14 recombination events) from GP179 (Figure 1). Recombinants between the *R1* flanking markers SPUD237 and GP179 were used for high-resolution genetic mapping using additional probes developed from PCR amplification of BAC-end sequences of clones identified with GP179, AFLP1, SPUD237, GM637 and 98R (see below).

Construction and hybridization of the S. demissum *BAC library*

The hexaploid wild potato, *S. demissum*, accession number PI 161729, was provided by NRSP-6 Potato GenBank (Sturgeon Bay, WI, USA) and carries at least four (*R1–R4*, Black and Gallegly, 1957) race-specific *R*-genes against late blight. It was used to construct a potato BAC library using the pIndigoBAC536 vector (Epicentre, Madison, WI, USA). *Solanum demissum* DNA was isolated from purified cell nuclei, partially digested with *Hin*dIII and cloned into the *Hin*dIII-digested pIndigoBAC536 vector. The *Hin*dIII library is composed of 397 056 BAC clones with an average insert size of 125 kb, equivalent to 17 times genome coverage of the hexaploid *S. demissum* (genome size 2700 Mb). The library contains 0.05% mitochondrial DNA clones and 5.1% chloroplast DNA clones. The *S. demissum* BAC library, SD_PBA, is available from the University of Arizona Genomics Institute (http://www.genome.arizona.edu/orders).

Filter hybridization was performed using a standard protocol (Sambrook *et al.*, 1989). For the initial screening step, high-density colony filters of the *S. demissum* BAC library were probed with potato molecular markers (GP179, AFLP1, SPUD237, GM637 and 98R) mapped on chromosome V (De Jong *et al.*, 1997; Marano *et al.*, 2002; Meksem *et al.*, 1995). Probes for subsequent rounds of screening were developed from PCR amplification of BAC-end sequences of selected clones. Table S1 lists oligonucleotide primers derived from BAC end sequences used for genetic mapping and/or further BAC library screening, as indicated.

BAC fingerprinting and contig development

Positive BAC clones were fingerprinted using several strategies. The first followed the protocols of Marra *et al.* (1997). An average of 30 DNA fragments per BAC clone were measured manually in 1% agarose gels after total digestion with *Hin*dIII. Band calling was performed using IMAGE software (Sulston *et al.*, 1989) with extensive manual editing. Automatic contig assembly of the fingerprinted clones was performed with the program FPC (FingerPrinted Contigs) ver. 4.7 at a cut-off of e^{-12} and a tolerance of 7, as described by Soderlund *et al.* (2000). The empirical tolerance and cut-off values for the FPC automated contig assembly were determined manually (Soderlund *et al.*, 2000).

A snapshot high-throughput fingerprinting method was also used to fingerprint BAC clones and to assemble contigs (Luo *et al.*, 2003). First, BAC DNA was digested with four 6-bp-cutter restriction

endonucleases that generated sticky ends and one 4-bp cutter that generated blunt ends. The digested fragments were labeled using four different fluorescent dyes. The labeled fragments were then sized using a capillary sequencing machine (ABI 3100; Applied BioSystems, Foster City, CA, USA). Contigs were assembled using the FPC program at a cut-off of 1×10^{-35} and a tolerance of 0.4 bp. Contigs obtained from high-throughput fingerprinting were compared with contigs obtained above, and only consensus contigs are presented. Any inconsistencies between these two methods were resolved by screening using additional markers as well as sequencing products amplified from BACs using PCR markers.

Sequencing, annotation and computational analysis

The BAC clones were sequenced using a shotgun approach, as reported previously (Yuan *et al.*, 2002). In brief, two shotgun libraries were constructed for each BAC averaging 2–3 and 8–10 kb. In total, the BAC was sequenced to approximately eight times sequence coverage and assembled with TIGR assembler (Sutton *et al.*, 1995). Gaps were closed using a combination of re-sequencing, primer walking and transposon-mediated sequencing. The final assembly was checked through comparison of an experimentally determined restriction enzyme fingerprint with a computationally generated fingerprint.

Genes at the *R1* locus in each haplotype were identified through manual curation of two data types: first, output from *ab initio* gene finders including GENSCAN (Burge and Karlin, 1997), GENEMARK.HMM (Lukashin and Borodovsky, 1998) and FGENESH (Salamov and Solovyev, 2000); second, sequence similarity as revealed through searches against the TIGR potato gene index (a nucleic acid database composed of a set of non-redundant ESTs; Quackenbush *et al.*, 2001) and a non-redundant amino acid database. Other annotation evidence included similarity with domains from two curated domain databases, including Pfam (Bateman *et al.*, 2002) and TIGRFAMs (Haft *et al.*, 2001). Genes that were identical to known genes in GenBank were assigned that gene name. Genes highly similar to known genes in GenBank were annotated as 'putative XXX'. Genes that aligned only with ESTs were termed 'expressed genes', and genes without significant similarity to known genes and lacking EST evidence were termed 'hypothetical genes'.

Sequences were aligned using CLUSTALX (Thompson *et al.*, 1994) and edited in GENEDOC (http://www.psc.edu/biomed/genedoc). The dot plot shown in Figure 2 was displayed using DOTTER, a dot-matrix program (Sonnhammer and Durbin, 1995). A neighbor-joining distance tree was constructed using Kimura's two-parameter model and bootstrap values were calculated using PAUP* 4.0 (Sinauer Associates, Sunderland, MA, USA). The trees constructed were visualized using the program TREEVIEW (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). Nucleotide identity between two sequences was calculated using PAUP* 4.0. Sequence exchanges were identified using GENECONV (Sawyer, 1989) and visual inspection, as described previously (Kuang *et al.*, 2004).

Models M7 and M8 in codeml of PAML were run for all *R1* fragments obtained from this study. Model M7 is a special case of model 8 that assumes no selection, whereas model 8 allows for positively selected sites (Yang *et al.*, 2000). Diversifying selection was confirmed using a likelihood-ratio test by comparing the likelihood of models M8 and M7 (Yang *et al.*, 2000).

PCR amplification of R1 *homologues*

R1 homologues were amplified using the TaqPlus long PCR system (Strategene, La Jolla, CA, USA) from BACs that were not sequenced. The PCR products were gel purified using the QIAquick gel extraction kit (Qiagen, Valencia, CA, USA), then cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA). Using the above method, *R1* homologues were also amplified from potato cultivars PI303148, Kennebec and Katahdin using primers specific to *R1* homologues at the distal sub-cluster (Table S1).

Acknowledgements

We thank Maureen Richey for critical review and editing of this manuscript. We thank Rod Wing for invaluable help and advice on construction of the *S. demissum* BAC library. This work was supported by National Science Foundation Plant Genome Research Program Grant DBI-9975866 and DBI-0218166. M.R.M. was supported by the International Foundation for Science, Grant No 99-332 and Third World Academy of Science, Grant c/3045-1.

Supplementary Material

The following supplementary material is available for this article online:

Table S1 Oligonucleotide primers used in this study

Table S2 Annotation of BAC sequences from the *S. demissum R1* locus

References

Balakirev, E. and Ayala, F.J. (2003) Pseudogenes: are they 'junk' or functional DNA? *Annu. Rev. Genet.* **37**, 123–151.

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

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L. (2002) The PFAM protein families database. *Nucleic Acids Res.* **30**, 276–280.

Bendahmane, A., Querci, M., Kanyuka, K. and Baulcombe, D.C. (2000) *Agrobacterium* transient expression system as a tool for the isolation of disease resistance genes: application to the *Rx2* locus in potato. *Plant J.* **21**, 73–81.

Black, W. and Gallegly, M.E. (1957) Screening of *Solanum* species for resistance to physiologic races of *Phytophthora infestans*. *Am. Potato J.* **34**, 273–281.

Bonierbale, M.W., Plaisted, R.L., Pineda, O. and Tanksley, S.D. (1994) QTL analysis of trichome-mediated resistance in potato. *Theor. Appl. Genet.* **87**, 973–987.

Botella, M.A., Parker, J.E., Frost, L.N., Bittner-Eddy, P.D., Beynon, J.L., Daniels, M.J., Holub, E.B. and Jones, J.D.G. (1998) Three genes of the Arabidopsis *RPP1* complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. *Plant Cell*, **10**, 1847–1860.

Bryan, G., McLean, K., Bradshaw, J., De Jong, W., Phillips, M., Castelli, L. and Waugh, R. (2002) Mapping QTLs for resistance to the cyst nematode *Globodera pallida* derived from the wild potato species *Solanum vernei*. *Theor. Appl. Genet.* **105**, 68–77.

Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94.

Caicedo, A.L., Schaal, B.A. and Kunkel, B.N. (1999) Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **96**, 302–306.

Cereb, N., Hughes, A.L. and Yang, S.Y. (1997) Locus-specific conservation of the *HLA* class I introns by intralocus recombination. *Immunogenetics*, **47**, 30–36.

Cooley, M.B., Pathirana, S., Wu, H.-J., Kachroo, P. and Klessig, D.F. (2000) Members of the Arabidopsis *HRT*/*RPP8* family of resistance genes confer resistance to both viral and Oomycete pathogens. *Plant Cell*, **12**, 663–676.

De Jong, W., Forsyth, A., Leister, D., Gebhardt, C. and Baulcombe, D.C. (1997) A potato hypersensitive resistance gene against potato virus X maps to a resistance gene cluster on chromosome 5. *Theor. Appl. Genet.* **95**, 246–252.

Dodds, P., Lawrence, G.J. and Ellis, J.G. (2001) Contrasting modes of evolution acting on the complex *N* locus for rust resistance in flax. *Plant J.* **27**, 439–453.

Ellis, J.G., Lawrence, G.J., Luck, J.E. and Dodds, P.N. (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell*, **11**, 495–506.

Grube, R.C., Radwanski, E.R. and Jahn, M. (2000) Comparative genetics of disease resistance within the solanaceae. *Genetics*, **155**, 873–887.

Haft, D.H., Loftus, B.J., Richardson, D.L., Yang, F., Eisen, J.A., Paulsen, I.T. and White, O. (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* **29**, 41–43.

Hughes, A.L. (2000) Evolution of introns and exons of class II *MHC* genes of vertebrates. *Immunogenetics*, **51**, 473–486.

Hulbert, S.H., Webb, C.A., Smith, S.M. and Sun, Q. (2001) Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* **39**, 285–312.

Kesseli, R.V., Paran, I. and Michelmore, R.W. (1994) Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. *Genetics*, **136**, 1435–1446.

Kreike, C.M., de Koning, J.R.A., Vinke, J.H., van Ooijen, J.W. and Stiekema, W.J. (1994) Quantitatively inherited resistance to *Globodera pallida* is dominated by one major locus in *Solanum spegazzinii*. *Theor. Appl. Genet.* **88**, 764–769.

Kuang, H., Woo, S.-S., Meyers, B.C., Nevo, E. and Michelmore, R.W. (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell*, **16**, 2870–2894.

Leonards-Schippers, C., Gieffers, W., Salamini, F. and Gebhardt, C. (1992) The *R1* gene conferring race-specific resistance to *Phytophthora infestans* in potato is located on potato chromosome V. *Mol. Gen. Genet.* **233**, 278–283.

Leonards-Schippers, C., Gieffers, W., Schafer-Pregl, R., Ritter, E., Knapp, S.J., Salamini, F. and Gebhardt, C. (1994) Quantitative resistance to *Phytophthora infestans* in potato: a case study for QTL mapping in an allogamous plant species. *Genetics*, **137**, 67–77.

Lukashin, A.V. and Borodovsky, M. (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**, 1107–1115.

Luo, M.-C., Thomas, C., You, F.M., Hsiao, J., Ouyang, S., Buell, C.R., Malandro, M., McGuire, P.E., Anderson, O.D. and Dvorak, J. (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics*, **82**, 378–389.

Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375–382.

Malcolmson, J.F. and Black, W. (1966) New *R* genes in *Solanum demissum* Lindl. and their complementary races of *Phytophthora infestans* (Mont.) de Bary. *Euphytica*, **15**, 199–203.

Marano, M.R., Malcuit, I., De Jong, W. and Baulcombe, C. (2002) High-resolution genetic map of *Nb*, a gene that confers hypersensitive resistance to potato virus X in *Solanum tuberosum*. *Theor. Appl. Genet.* **105**, 192–200.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D. and Waterston, R.H. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Res.* **7**, 1072–1084.

McDowell, J.M., Dhandaydham, M., Long, T.A., Aarts, M.G.M., Goff, S., Holub, E.B. and Dangl, J.L. (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of Arabidopsis. *Plant Cell*, **10**, 1861–1874.

Meksem, K., Leister, D., Peleman, J., Zabeau, M., Salamini, F. and Gebhardt, C. (1995) A high-resolution map of the vicinity of the *R1* locus on chromosome V of potato based on RFLP and AFLP markers. *Mol. Gen. Genet.* **249**, 74–81.

Meyers, B.C., Chin, D.B., Shen, K.A., Sivaramakrishnan, S., Lavelle, D.O., Zhang, Z. and Michelmore, R.W. (1998a) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell*, **10**, 1817–1832.

Meyers, B.C., Shen, K.A., Rohani, P., Gaut, B.S. and Michelmore, R.W. (1998b) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *Plant Cell*, **10**, 1833–1846.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H. and Michelmore, R.W. (2003) Genome-wide analysis of NBS–LRR-encoding genes in Arabidopsis. *Plant Cell*, **15**, 809–834.

Michelmore, R.W. and Meyers, B.C. (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* **8**, 1113–1130.

Mondragon-Palomino, M., Meyers, B.C., Michelmore, R.W. and Gaut, B.S. (2002) Patterns of positive selection in the complete NBS–LRR gene family of *Arabidopsis thaliana*. *Genome Res.* **12**, 1305–1315.

Naess, S.K., Bradeen, J.M., Wielgus, S.M., Haberlach, G.T., McGrath, J.M. and Helgeson, J.P. (2000) Resistance to late blight in *Solanum bulbocastanum* is mapped to chromosome 8. *Theor. Appl. Genet.* **101**, 697–704.

Noël, L., Moores, T.L., der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E. and Jones, J.D.G. (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of Arabidopsis. *Plant Cell*, **11**, 2099–2112.

Oberhagemann, P., Chatot-Balandras, C, Bonnel, E., Schäfer-Pregl, R., Wegener, D., Palomino, C., Salamini, F. and Gebhardt, C. (1999) A genetic analysis of quantitative resistance to late blight in potato: towards marker-assisted selection. *Mol. Breeding*, **5**, 399–415.

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B. and Jones, J.D. (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4*/*9* locus of tomato. *Cell*, **91**, 821–832.

Quackenbush, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Pertea, G., Sultana, R. and White, J. (2001) The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res.* **29**, 159–164.

Ritter, E., Debener, T., Barone, A., Salamini, F. and Gebhardt, C. (1991) RFLP mapping on potato chromosomes of two genes controlling extreme resistance to potato virus X (PVX). *Mol. Gen. Genet.* **227**, 81–85.

Rouppe van der Voort, J., Lindeman, W., Folkertsma, R., Hutten, R., Overmars, H., Van Der Vossen, E., Jacobsen, E. and Bakker, J. (1998) A QTL for broad-spectrum resistance to cyst nematode species (*Globodera* spp.) maps to a resistance gene cluster in potato. *Theor. Appl. Genet.* **96**, 654–661.

Salamov, A.A. and Solovyev, V.V. (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522.

Salmeron, J.M., Oldroyd, G.E., Rommens, C.M., Scofield, S.R., Kim, H.S., Lavelle, D.T., Dahlbeck, D. and Staskawicz, B.J. (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. *Cell*, **86**, 123–133.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sawyer, S. (1989) Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**, 526–538.

Schornack, S., Ballvora, A., Gurlebeck, D., Peart, J., Ganal, M., Baker, B., Bonas, U. and Lahaye, T. (2004) The tomato resistance protein Bs4 is a predicted non-nuclear TIR–NB–LRR protein that mediates defense responses to severely truncated derivatives of *AvrBs4* and overexpressed *AvrBs3*. *Plant J.* **37**, 46–60.

Seah, S., Yaghoobi, J., Rossi, M., Gleason, C.A. and Williamson, V.M. (2004) The nematode-resistance gene, *Mi-1*, is associated with an inverted chromosomal segment in susceptible compared to resistant tomato. *Theor. Appl. Genet.* **108**, 1635–1642.

Shaw, D.S. (1991) Genetics. In *Phytophthora infestans, the Cause of Late Blight of Potato* (Ingram, D.S. and Williams, P.H., eds). London: Academic Press, pp. 131–170.

Shen, K.A., Chin, D.B., Arroyo-Garcia, R., Ochoa, O.E., Lavelle, D.O., Wroblewski, T., Meyers, B.C. and Michelmore, R.W. (2002) *Dm3* is one member of a large constitutively expressed family of nucleotide binding site–leucine-rich repeat encoding genes. *Mol. Plant–Microbe Interact.* **15**, 251–261.

Soderlund, C., Humphray, S., Dunham, A. and French, L. (2000) Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res.* **10**, 1772–1787.

Song, W.Y., Pi, L.Y., Wang, G.L., Gardner, J., Holsten, T. and Ronald, P.C. (1997) Evolution of the rice *Xa21* disease resistance gene family. *Plant Cell*, **9**, 1279–1287.

Song, W.Y., Pi, L.Y., Bureau, T.E. and Ronald, P.C. (1998) Identification and characterization of 14 transposon-like elements in the noncoding regions of members of the Xa21 family of disease resistance genes in rice. *Mol. Gen. Genet.* **258**, 449–456.

Song, J., Bradeen, J.M., Naess, S.K. *et al.* (2003) Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. *Proc. Natl Acad. Sci. USA*, **100**, 9128–9133.

Sonnhammer, E.L.L. and Durbin, R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**, GC1–GC10.

Spooner, D.M., van den Berg, R.G. and Bamberg, J.B. (1995) Examination of species boundaries of *Solanum* series *Demissa* and potentially related species in series *Acaulia* and series *Tuberosa* (sect. *Petota*). *Syst. Bot.* **20**, 295–314.

Sulston, J., Mallet, F., Dunham, A. and Horsnell, T. (1989) Image analysis of restriction enzyme fingerprint autoradiograms. *Comput. Appl. Biosci.* **5**, 101–106.

Sun, Q., Collins, N.C., Ayliffe, M., Smith, S.M., Drake, J., Pryor, T. and Hulbert, S.H. (2001) Recombination between paralogues at the *Rp1* rust resistance locus in maize. *Genetics*, **158**, 423–438.

Sutton, G., White, O., Adams, M. and Kerlavage, A. (1995) TIGR assembler: a new tool for assembling large shotgun sequencing projects. *Genome*, **1**, 9–19.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Umaerus, V. and Umaerus, M. (1994) Inheritance of resistance to late blight. In *Potato Genetics* (Bradshaw, J.E. and Mackay, G.R., eds). Wallingford, UK: CAB International, pp. 365–401.

Van der Hoorn, R.A., Kruijt, M., Roth, R., Brandwagt, B.F., Joosten, M.H. and De Wit, P.J. (2001) Intragenic recombination generated two distinct *Cf* genes that mediate *AVR9* recognition in the natural population of *Lycopersicon pimpinellifolium*. *Proc. Natl Acad. Sci. USA*, **98**, 10493–10498.

Visker, M.H., Keizer, L.C., Van Eck, H.J., Jacobsen, E., Colon, L.T. and Struik, P.C. (2003) Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type? *Theor. Appl. Genet.* **106**, 317–325.

Vleeshouwers, V., van Dooijeweert, W., Willem, P., Sijpkes, L., Govers, F. and Colon, L. (1999) A laboratory assay for *Phytophthora infestans* resistance in various *Solanum* species reflects the field situation. *Eur. J. Plant Pathol.* **105**, 241–250.

van der Vossen, E.A., van der Voort, J.N., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D.C., Bakker, J., Stiekema, W.J. and Klein-Lankhorst, R.M. (2000) Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. *Plant J.* **23**, 567–576.

Wastie, R.L. (1991) Breeding for resistance. In *Phytophthora infestans, the Cause of Late Blight of Potato* (Ingram, D.S. and Williams, P.H., eds). London: Academic Press, pp. 193–224.

Wei, F., Wing, R.A. and Wise, R.P. (2002) Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. *Plant Cell*, **14**, 1903–1917.

Wessler, S.R., Bureau, T.E. and White, S.E. (1995) LTR-retrotransposon and MITEs – important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5**, 814–821.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS*, **13**, 555–556.

Yang, Z., Nielsen, R., Goldman, N. and Pedersen, A.M.K. (2000) Codon-substitution models for heterogenous selection pressure at amino acid sites. *Genetics*, **155**, 431–449.

Yuan, Q., Hill, J., Hsiao, J., Moffat, K., Ouyang, S., Cheng, Z., Jiang, J. and Buell, C.R. (2002) Genome sequencing of a 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion. *Mol. Genet. Genomics*, **267**, 713–720.

GenBank accession numbers: BAC clone sequences from this article have been deposited with GenBank data libraries under accession numbers: AC135288, AC139840, AC142505, AC144791, AC145120, AC146506, AC149265, AC149266, AC149267, AC149287, AC149288, AC149290, AC149291, AC149301, AC149487, AC150162, AC151803, AC151815, AC154033. The *R1* homologues amplified from *Solanum demissum* BAC clones or cultivated *Solanum tuberosum* genomic DNA using PCR methods were deposited with GenBank accession numbers: AY955222, AY955223, AY955224 (*R1* homologues amplified from *S. demissum* BAC clones); and AY935701, AY935702, AY935703, AY935704, AY935705, AY935706, AY935707, AY935708, AY935709, AY935710 (*R1* homologues amplified from *S. tuberosum* genomic DNA).