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

Hanhu Kuang1,†, Fusheng Wei1,†, Maria Rosa Marano1,2,†, Uwe Wirtz2,†, Xiaoxue Wang1,†, Jia Liu3, Wai Pun Shum1, Jennifer Zaborsky3, Luke J. Tallon3, Willem Rensink3, Stacey Lobst3, Peifen Zhang1, Carl-Erik Tornqvist1, Ahmet Tek4, John Bamberg4,5, John Helgeson6, William Fry7, Frank You8, Ming-Cheng Luo8, Jiming Jiang4, C. Robin Buell3 and Barbara Baker1,*

1Plant Gene Expression Center, USDA-ARS and Department of Plant and Microbial Biology, University of California, Berkeley, 800 Buchanan Street, Albany, CA 94710, USA, 2Instituto de Biología Molecular y Celular de Rosario, Facultad de Ciencias Bioquímicas y Farmaceuticas, Universidad Nacional de Rosario, Suipacha 531, S2002LRK Rosario, Argentina, 3The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA, 4Department of Horticulture, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706, USA, 5Potato Introduction Station, 4312 Hwy 42 Sturgeon Bay, WI 54235, USA, 6Department of Plant Pathology, University of Wisconsin and USDA-ARS, 1630 Linden Drive, Madison, WI 53706, USA, 7Department of Plant Pathology, Cornell University, Ithaca, NY 14853, USA, and 8Department 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 (Hulbert, 1999). 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 (phyt3, 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 TIR-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., 1999; Parniske et al., 1999; Wang 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 RGC2s 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 substitutions over synonymous 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 \textit{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 \textit{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 \textit{RGC2} genes was caused by DNA structures of \textit{R}-gene homologues or their flanking regions that prohibited pairing with other \textit{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 \textit{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 \textit{RGC2} genes in lettuce can be distinguished based on their characteristic high or low frequency exchange between paralogues, it remains to be seen whether \textit{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 \textit{R}-gene cluster in \textit{Solanum} using wild selfing allohexaploid potato \textit{S. demissum} (2\textit{n} = 6\textit{x} = 72). It is believed that two of the three genomes in \textit{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 \textit{S. demissum} is similar to the genomes in autotetraploid potato \textit{S. tuberosum}. However, the ancestral diploid or tetraploid of \textit{S. demissum} is still under debate. \textit{Solanum demissum} has been the major source for late blight-resistance genes in cultivated potato, and at least 11 \textit{R}-genes (R1–R11) confering race-specific resistance against \textit{P. infestans} were introgressed from \textit{S. demissum} into cultivated potato (Malcolmson and Black, 1966; Shaw, 1991; Umaerus and Umaerus, 1994). The current study focuses on the \textit{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 \textit{R1}-containing \textit{S. demissum} accession with the incorporation of a genetic map constructed in an \textit{R1}-containing \textit{S. tuberosum} cultivar. We present here a sequence comparison of the three haplotypes of the \textit{R1} resistance-gene cluster, and comprehensive analysis of the organization and evolution of the region's \textit{R1} resistance-gene candidates (\textit{RGCs}). The \textit{R1} locus shows substantial variation among the three genomes. We determined that the \textit{R1} homologues in the region fall into three distinct groups of fast-evolving type I \textit{R}-genes based on frequent sequence exchanges within groups, but not among groups. Our findings at the \textit{R1} locus in \textit{S. demissum} may indicate that a common or similar mechanism underlies the previously reported differentiation of type I and II \textit{R}-genes and the differentiation of type I \textit{R}-genes into distinct groups, identified here.

**Results**

**Identification of BAC clones spanning the \textit{R1} region**

A genomic bacterial artificial chromosome (BAC) library was constructed for \textit{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 \textit{S. demissum} (see Experimental procedures).

BAC clones spanning the \textit{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 \textit{S. tuberosum} cultivars Kennebec (\textit{R1}) and Katahdin (\textit{r1}; see Experimental procedures). The genomic BAC library of \textit{S. demissum} was hybridized with probes derived from two markers proximal to resistance gene \textit{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 \textit{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' \textit{R1}, developed from \textit{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 \textit{R1} gene using markers 132D05S, 160O02T and 780I15S (Figure 1; Table S1).
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 to 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 160002, 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...
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.

### Table 1 Solanum demissum BAC clones sequenced

<table>
<thead>
<tr>
<th>BAC clone</th>
<th>Size (bp)</th>
<th>GenBank accession</th>
<th>Haplotype</th>
<th>Region (size, kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGEC97H09</td>
<td>110 954</td>
<td>AC151803</td>
<td>A³</td>
<td>I (0–110)</td>
</tr>
<tr>
<td>PGEC446J10</td>
<td>102 692</td>
<td>AC149288</td>
<td>A</td>
<td>I (0–21), II (21–88), III (88–102)</td>
</tr>
<tr>
<td>PGEC668E02</td>
<td>142 575</td>
<td>AC144791</td>
<td>A</td>
<td>III (0–58), IV (58–142)</td>
</tr>
<tr>
<td>PGEC472P22</td>
<td>136 150</td>
<td>AC151815</td>
<td>A</td>
<td>IV (52–136), V (7–52)</td>
</tr>
<tr>
<td>PGEC695L19(PGEC407)</td>
<td>73 587</td>
<td>AC149487</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>PGEC420A04(PGEC219)</td>
<td>124 895</td>
<td>AC135288</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>PGEC160O02</td>
<td>125 813</td>
<td>AC150162</td>
<td>B</td>
<td>I (0–125)</td>
</tr>
<tr>
<td>PGEC899P08</td>
<td>121 915</td>
<td>AC146506</td>
<td>B</td>
<td>I (0–115), II (115–121)</td>
</tr>
<tr>
<td>PGEC568D21</td>
<td>119 525</td>
<td>AC149287</td>
<td>B</td>
<td>II (63–119), III (0–63)</td>
</tr>
<tr>
<td>PGEC747E24</td>
<td>92 340</td>
<td>AC149267</td>
<td>B</td>
<td>III (72–92), IV (0–72)</td>
</tr>
<tr>
<td>PGEC872C13</td>
<td>100 564</td>
<td>AC149266</td>
<td>B</td>
<td>IV (0–100)</td>
</tr>
<tr>
<td>PGEC658M02</td>
<td>114 450</td>
<td>AC149265</td>
<td>B</td>
<td>IV (0–114)</td>
</tr>
<tr>
<td>PGEC517A09</td>
<td>133 983</td>
<td>AC149301</td>
<td>B</td>
<td>IV (0–64), V (64–106)</td>
</tr>
<tr>
<td>PGEC287I17</td>
<td>78 626</td>
<td>AC142505</td>
<td>ND³</td>
<td></td>
</tr>
<tr>
<td>PGEC446O19</td>
<td>122 450</td>
<td>AC145120</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>PGEC132D05</td>
<td>118 497</td>
<td>AC154033</td>
<td>C</td>
<td>I (0–118)</td>
</tr>
<tr>
<td>PGEC568H16</td>
<td>146 091</td>
<td>AC149291</td>
<td>C</td>
<td>II (113–143), III (38–113), IV (11–38), V (0–11)</td>
</tr>
<tr>
<td>PGEC093P17</td>
<td>200 238</td>
<td>AC149290</td>
<td>C</td>
<td>V (0–36)</td>
</tr>
<tr>
<td>PGEC323D08</td>
<td>106 746</td>
<td>AC139840</td>
<td>ND</td>
<td></td>
</tr>
</tbody>
</table>

³Regions I–V as described in text (Figure 1).  
²Haplotype A BACs were identified with marker alleles in coupling with the R1 resistance gene.  
³BAC nucleotide numbering in reverse orientation.  
⁴ND, Not determined, BACs belong to either haplotype B or C.

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).
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 ± 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; TP11-3 from cv. P1303148. 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 (Ks) 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 TPI1-3, Tken4-6, TKat1) from cv. Katahdin; and three (A955224 respectively). Therefore a total of nine R1 homologues, three fragments are assumed to be located within the distal 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 PCR-amplified and sequenced (GenBank TPI1-3 and Tken4-6, TKat1) from cv. Katahdin; and three (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.

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 (χ² = 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_Prfrf-1 and Sd_Prfrf-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_Prfrf-1 and Sd_Prfrf-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_Prfrf-2 in haplotype C, Sd_Prfrf-2, is present in BAC clone 323D08 (Figure 1). The Sd_Prfrf-1 and Sd_Prfrf-2 homologues are located in the same relative position in each haplotype, maintain obvious allelic relationships, and are highly conserved. The three Sd_Prfrf-1 genes exhibit nucleotide identities of 97%, though several large insertions/deletions were observed in the copy in haplotype A. The three Sd_Prfrf-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 \textit{Sd\_Prf-1} and \textit{Sd\_Prf-2} homologues at the \textit{R1} locus are type II genes.

The structures of \textit{Sd\_Prf-1} and \textit{Sd\_Prf-2} differ from each other and from the \textit{Prf} gene in tomato. Tomato \textit{Prf} has a 5475 bp coding region, yet the \textit{Sd\_Prf-2} genes in haplotypes A and B are predicted to contain more than 7.7 kb of coding sequence (\textit{Sd\_Prf-c-2} is only partially represented). The expansion of \textit{Sd\_Prf-a-2} and \textit{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 \textit{Sd\_Prf-a-2} (at the 1314th codon) and in \textit{Prf-b-2} (at the 1274th codon), thus both are likely to be pseudogenes. All three \textit{Sd\_Prf-1} genes are shorter than the \textit{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 \textit{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 \textit{Sd\_Prf-1} and \textit{Sd\_Prf-2} and their divergence from tomato \textit{Prf}, the \textit{Sd\_Prf-1} and \textit{-2} genes found in all three haplotypes are highly conserved, and show the types of conserved relationship previously described for slowly evolving type II \textit{R}-genes.

Resistance-gene candidates \textit{Sd\_Bs4a-1} and \textit{Sd\_Bs4c-1} belong to the third \textit{R}-gene family at the \textit{R1} locus and are homologous to \textit{Bs4}, a tomato TIR–NBS–LRR resistance gene against bacteria \textit{X. campestris} pv. \textit{vesicatoria}. A full-length copy of \textit{Sd\_Bs4a-1} on BAC 979H09 in haplotype A, and a partial copy of \textit{Sd\_Bs4c-1} from the corresponding region in haplotype C on BAC 132D05, were sequenced (Figure 1). \textit{Sd\_Bs4a-1} has a 1-bp insertion and a 2-bp deletion in exon 2, while \textit{Sd\_Bs4c-1} has a 349-bp insertion in exon 2, therefore both \textit{Bs4} homologues appear to be pseudogenes. Regardless of the frame-shift insertion/deletions, the putative coding regions of \textit{Sd\_Bs4a-1} and \textit{Sd\_Bs4c-1} exhibit 96% nucleotide identity with each other, and approximately 90% nucleotide identity with the tomato \textit{Bs4} resistance gene. It is likely that the \textit{Bs4} homologues at the \textit{R1} locus also belong to the type II class of \textit{R}-genes.

Discussion

Resistance genes and traits at the \textit{R1} locus

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

Structural comparisons of the \textit{R1} homologues within the three haplotypes of \textit{S. demissum} showed that they can be divided into three groups of independently evolving type I \textit{R}-genes, and that most homologues within each group show >90% nucleotide identity. The finding that the type I \textit{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 \textit{R}-genes reported previously for the \textit{RGC2} cluster in lettuce.

Comparison of the \textit{S. demissum} \textit{R1} region sequences described in this study with publicly available sequences (GenBank accessions AY730336 and AY730340) show that 90 738 bp of \textit{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 \textit{S. demissum} ‘haplotype A’ (Figure 1). The approximately 91-kb introgressed \textit{R1} region from \textit{S. tuberosum} and the corresponding region in haplotype A of \textit{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 \textit{R1} gene region of \textit{S. demissum} (Figure 1). The \textit{R1} gene in \textit{S. tuberosum} P6/10 was probably introgressed from an \textit{S. demissum} accession different from that (PI 161729) described here.

We found that alleles of several markers in the region proximal to the \textit{R1} cluster, that are linked in coupling with the \textit{R1} gene (461015S, 123A20T, 420A04T and GP179) in the \textit{S. tuberosum} cv. Kennebec \textit{R1} mapping parent, are also present in BACs of the \textit{S. demissum} haplotype A contig. These results suggest that an \textit{S. demissum} chromosome V segment proximal to the \textit{R1} gene cluster, between markers 461015S and GP179, was introgressed into cv. Kennebec along with the \textit{R1} resistance gene.

Members of two other \textit{R}-gene families, \textit{Prf} and \textit{Bs4}, closely linked with the \textit{R1} homologues, were identified in this study. The tomato resistance gene \textit{Bs4} is a TIR–NBS–LRR-encoding gene, located within the syntenic region of the \textit{Solanum} \textit{R1} cluster. In this study two \textit{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 \textit{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 \textit{R1}-mediated resistance, other disease-resistance traits, including QRLs for \textit{P. infestans} and nematode resistance, have been mapped near the \textit{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

© Blackwell Publishing Ltd, \textit{The Plant Journal}, (2005), \textbf{44}, 37–51
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-genomes.

**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-genomes 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 (Dodd 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 crossover.

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 F1 plants were tested for late-blight resistance using P. infestans isolates US90480, genotype US-8 (mating type A2) and genotype US-1 (mating type A11. 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 HindIII and cloned into the HindIII-digested pIndigoBAC536 vector. The HindIII 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 HindII. 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 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 $1 \times 10^{-10}$ 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 resequencing, 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 $R_1$ 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 (Saitou and Nei, 1987) and 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 (Sinauer Associates, Sunderland, MA, USA). The trees constructed were visualized using the program TREES VIEW (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 GENEDCONV (Sawyer, 1989) and visual inspection, as described previously (Kuang et al., 2004). Models M7 and M8 in codeml of PAML were run for all $R_1$ fragments obtained from this study. Model M7 is a special case of model 8 that assumes no selection, whereas model 8 allows for diversifying selection, 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.

**PCR amplification of $R_1$ homologues**

$R_1$ homologues were amplified using the TaqPlus long PCR system (Stratagene, 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 TOP10 TA cloning kit (Invitrogen, Carlsbad, CA, USA). Using the above method, $R_1$ homologues were also amplified from potato cultivars P1930148, Kennebec and Katahdin using primers specific to $R_1$ 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* $R_1$ locus

**References**


