The Fractionated Orthology of *Bs2* and *Rx/Gpa2* Supports Shared Synteny of Disease Resistance in the Solanaceae

Michael Mazourek,* Elizabeth T. Cirulli,*¹ Sarah M. Collier,*[†] Laurie G. Landry,*
Byoung-Cheorl Kang,*[‡] Edmund A. Quirin,[§] James M. Bradeen,[§] Peter Moffett^{†,2}
and Molly M. Jahn*^{*,3}

*Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853, [†]Boyce Thompson Institute for Plant Research, Ithaca, New York 14853, [‡]Department of Plant Sciences, Seoul National University, Seoul 151-921, Korea, [§]Department of Plant Pathology, University of Minnesota, St. Paul, Minnesota 55108 and **College of Agriculture and Life Sciences, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received January 27, 2009
Accepted for publication April 29, 2009

ABSTRACT

Comparative genomics provides a powerful tool for the identification of genes that encode traits shared between crop plants and model organisms. Pathogen resistance conferred by plant R genes of the nucleotide-binding-leucine-rich-repeat (NB-LRR) class is one such trait with great agricultural importance that occupies a critical position in understanding fundamental processes of pathogen detection and coevolution. The proposed rapid rearrangement of R genes in genome evolution would make comparative approaches tenuous. Here, we test the hypothesis that orthology is predictive of R-gene genomic location in the Solanaceae using the pepper R gene *Bs2*. Homologs of *Bs2* were compared in terms of sequence and gene and protein architecture. Comparative mapping demonstrated that *Bs2* shared macrosynteny with R genes that best fit criteria determined to be its orthologs. Analysis of the genomic sequence encompassing solanaceous R genes revealed the magnitude of transposon insertions and local duplications that resulted in the expansion of the *Bs2* intron to 27 kb and the frequently detected duplications of the 5'-end of R genes. However, these duplications did not impact protein expression or function in transient assays. Taken together, our results support a conservation of synteny for NB-LRR genes and further show that their distribution in the genome has been consistent with global rearrangements.

R genes have a central role in plant disease resistance to mediate pathogen detection and response (MARTIN *et al.* 2003; GLAZEBROOK 2005). Although R genes are only one of the components required for these responses, they are consistently identified as a critical determinant for qualitative and quantitative resistance (FLUHR 2001; WISSER *et al.* 2006). The structure, mechanism of action, and evolution of this gene family are still being elucidated and are critical issues for a more efficient deployment of disease resistances in agricultural crops (McDOWELL and SIMON 2006; TAKKEN *et al.* 2006; FRIEDMAN and BAKER 2007; VAN OOIJEN *et al.* 2007).

Comparative studies of sequence similarity between plant R proteins and proteins of innate immunity in animals have made important contributions toward understanding R-protein structure, the role of individual

protein domains, and the mechanism by which R proteins identify and respond to foreign proteins (NURNBERGER *et al.* 2004; TAKKEN *et al.* 2006; RAIRDAN and MOFFETT 2007). Both share a central nucleotide-binding (NB) site and a region of homology termed the "ARC" domain (collectively referred to as the NB-ARC) (VAN DER BIEZEN and JONES 1998; RAIRDAN and MOFFETT 2007). The plant counterparts have a highly variable leucine-rich-repeat (LRR) domain at the C terminus and, at the N terminus, either a domain with homology to the Toll and interleukin-1 receptors (TIR) or lack this feature, instead possessing a domain that may include a coiled-coil motif. Due to uncertainty regarding the presence of a coiled-coil motif, this class of NB-LRRs is often referred to as non-TIR proteins. The LRR domains are highly variable and tend to be under diversifying selection to adapt to continually changing pathogen proteins (MEYERS *et al.* 1998b; MICHELMORE and MEYERS 1998; MONDRAGON-PALOMINO *et al.* 2002). Other conserved patterns have been identified in the N terminus of non-TIR proteins, most notably, an EDxxD motif that mediates an intramolecular interaction (RAIRDAN *et al.* 2008). The interaction with cellular factors is mediated by the N-terminal domains of NB-LRR proteins although domain-swapping experiments between closely related

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.109.101022/DC1>.

¹Present address: Center for Human Genome Variation, Duke University, Durham, NC 27708.

²Present address: Département de Biologie, Université de Sherbrooke, 2500 Blvd. de l'Université, Sherbrooke QC J1K 2R1, Canada.

³Corresponding author: College of Agricultural and Life Sciences, University of Wisconsin, 140 Agricultural Hall, Madison, WI 53706. E-mail: mjahn@cals.wisc.edu

NB-LRR proteins have shown that recognition specificity is determined by the LRR domains (RAIRDAN and MOFFETT 2007; VAN OOIJEN *et al.* 2007).

The clustering of R genes has provided both insight into their ability to evolve rapidly and challenges to their identification and cloning. R genes often occur in clusters of tandem duplications that can span several megabases and include a multitude of copies of functional R genes, pseudogenes, and other genes within the clusters (MEYERS *et al.* 1998a; KUANG *et al.* 2004; SMITH *et al.* 2004). Of the various modes of evolution ascribed to these clusters, sequence exchange between R genes within the cluster by unequal crossing over or illegitimate recombination is especially noteworthy (MICHELMORE and MEYERS 1998; ELLIS *et al.* 2000; HULBERT *et al.* 2001; McDOWELL and SIMON 2006; FRIEDMAN and BAKER 2007; WICKER *et al.* 2007). Under stress conditions, transposon activation, recombination activation, and chromatin modifications related to small RNAs may be induced (LEVY *et al.* 2004; FRIEDMAN and BAKER 2007; YI and RICHARDS 2007).

Two distinct models for the genomewide arrangement and distribution of NB-LRR genes and these clusters have been proposed. The first predicts rapid rearrangement of R-gene distribution during genome evolution, yielding poor conservation of R-gene locations (LEISTER *et al.* 1998; RICHLY *et al.* 2002; MEYERS *et al.* 2003). Indeed, in monocots, extensive loss of genomewide R-gene colinearity has been attributed to frequent R-gene duplication and ectopic transposition (GALE and DEVOS 1998; PATERSON *et al.* 2003). In contrast, the second model supports genomewide conservation of R-gene distribution maintained during speciation. According to this model, most duplication and recombination of R-gene sequences should occur within restricted chromosomal regions, yielding clusters of closely related R-gene sequences. The resulting orthology relationships (homologs related by speciation, not duplication) are complex due to "fractionation" (repeated cycles of duplication, deletion, and recombination) but can, as we have previously shown, be reconstructed (GRUBE *et al.* 2000b). Analysis of R genes using the complete *Arabidopsis thaliana* genome sequence supports this model and accounts for the consensus of NB-LRR sequences (BAUMGARTEN *et al.* 2003). Resistance to a particular pathogen type is not conserved, and highly similar NB-LRR proteins may confer resistance to very different pathogens (GRUBE *et al.* 2000b).

Bs2 encodes a non-TIR NB-LRR protein identified in *Capsicum chacoense* that confers resistance to the bacterium *Xanthomonas campestris* pv. *vesicatoria*. This R gene has greatest sequence identity to *Rx* and *Gpa2* in potato, which confer resistance to a virus and nematode, respectively (BENDAHDANE *et al.* 1999; TAI *et al.* 1999b; VAN DER VOSSEN *et al.* 2000). Despite the difference in the pathogens recognized by these genes, they are

distinguishable from all other known R genes by marked sequence and structural features. In this study, we demonstrate that these three R genes are derived from syntenic regions in solanaceous genomes as predicted by our model of conservation of synteny. In performing these comparisons, we explore conserved amino acid patterns associated with proteins of the non-TIR family and the local genomic context of R genes of the Solanaceae. Finally, advances in the development of the Solanaceae as a system for comparative genomics highlight a role for chromosomal rearrangements in R-gene distribution throughout plant genomes.

MATERIALS AND METHODS

Plant material: Capsicum genotypes used in this study were *Capsicum annuum* NuMex R Naky (R Naky), Early CalWonder 300 (ECW), Early CalWonder-123R (ECW123) (provided by Robert Stall), Yolo Wonder (YW), Perennial (A. Palloix, INRA, Montfavet, France), *Capsicum chinense* P1159234 and *C. chacoense* P1439414 (U. S. Department of Agriculture Agricultural Research Station Southern Regional PI Station, Griffin, GA), and an F₂ population of 75 individuals derived from the cross R Naky × P1159234 (LIVINGSTONE *et al.* 1999). A tomato mapping population of 88 F₂ individuals originating from a cross between *Solanum pennellii* and *Solanum lycopersicum* was provided by S. Tanksley.

R-gene sequence analysis: NB-LRR sequences were obtained from the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov>) in December 2004 using the Bs2 protein sequence (AAF09256) as a query in BLASTP and are detailed in Table 1. Later searches established that since the original survey no proteins in the Bs2/Gpa2/Rx clade have been described with a characterized role in disease resistance.

Dendrogram construction: Input sequences for dendrogram construction consisted of 452 amino acids of the NB-ARC and flanking regions of R proteins aligned using DIALIGN (MORGENSTERN *et al.* 1998; KUMAR *et al.* 2001). The aligned sequences commenced seven amino acids before the GMG motif and extended 10 amino acids past the MHD motif of this region. The high divergence at the nucleotide level did not permit recombination detection. A neighbor-joining dendrogram was constructed using MEGA 2.1 (KUMAR *et al.* 2001). The p-distance model was employed with pairwise deletion gap handling. Ten thousand bootstrap replications were generated to examine the robustness of data trends.

Coiled-coil domain prediction: To predict coiled-coils, deduced R-protein sequences were analyzed using the COILS (LUPAS *et al.* 1991) and Marcoil (DELORENZI and SPEED 2002) programs. When analyzing the data set with COILS, the 14- and 21-amino-acid window sizes were used with the most encompassing matrix, MTIDK. For Marcoil, three matrices were used: 9FAM, MTK, and MTIDK. The outputs were graphed as the coils score along the length of the protein, and results were divided into three categories based on descriptive criteria. Regions that were predicted by both algorithms to contain coiled-coils with likelihood ≥40% were classified as "strong." Regions that were predicted by both algorithms to contain coiled-coils with likelihood between 10% and 40% or that were predicted by only one algorithm to contain coiled-coils with likelihood >85% were subjectively classified as "weak." Other regions were assumed to not harbor a coiled-coil motif.

Hydrophobic domain prediction: Sequences were analyzed using the Kyte-Doolittle hydrophobicity plot in the Lasergene

program, Protean (DNASTar, Madison, WI). A sliding window of nine amino acids, the ideal window size for finding hydrophobic domains in globular proteins (KYTE and DOOLITTLE 1982), was used. A moving average trendline with a period of 9 was plotted over the data to assist visualization. Protein regions scoring above a stringent threshold of 2.1 units above the grand average hydropathy for each protein were considered to be hydrophobic.

Leucine-rich repeats: The C-terminal LRR domain consists of a variable number of leucine-rich repeats. The pattern LXXLXXLXXLXX(N/C/T)(X)XLXXIPXX was originally reported as the consensus sequence for these repeats (JONES and JONES 1997). The underlined portion of the consensus sequence matched the examined protein sequences best. For consistency, we reevaluated the LRR descriptions of all R proteins in our data set and manually reannotated Pi-Ta, Dm3, RP3, and RPG1b LRRs.

Analysis of duplicated genome sequences: The DotPlot program in Lasergene's Megalign was used to compare various DNA sequences. The *Bs2* YAC (AY702979) was aligned against itself, using a minimum similarity of 65% and a window size of 50, and the *Rx/Gpa2* contig (AF265664) alignment used 65% similarity and a 75-base window. The solanaceous R genes were aligned against their respective genomic sequence to find local duplications (*Mi* 1.2, U81378; *RB*, AY303171; *R1*, EF514212; *Tm2*, AF536201). Pairwise percentage similarity of duplications was calculated using Megalign's ClustalV. Regions that were repeated one or more times within the *Bs2* BAC were assigned putative identifications using BLASTX on default settings. Transposon identification was performed using CENSOR (KOHANY *et al.* 2006).

Localization of *Bs2*, *Gpa2*, *Me*, and *Mech* loci on a Capsicum linkage map: DNA markers and genes corresponding to resistance gene loci were integrated into the Capsicum linkage map of LIVINGSTONE *et al.* (1999) by the previously described method (BLUM *et al.* 2003). PCR-based markers and RFLP probes were prepared as described below.

***Bs2* locus:** To determine the position of *Bs2* in the pepper linkage map, two *Bs2*-linked markers, A2 and S19, were used. These map 0 and 7 cM from *Bs2*, respectively (TAI *et al.* 1999b). To localize the A2 marker in our linkage map, A2 fragments were amplified from genomic DNA of ECW-123 using A2 STS PCR primers according to TAI *et al.* (1999a). The resulting 528-bp A2 fragment was used as a probe for RFLP hybridizations. To localize the codominant SCAR marker S19 in our linkage map, S19 primers (TAI *et al.* 1999a) were used.

***Gpa2/Rx* loci:** PCR primers (Integrated DNA Technologies, Coralville, IA) were used to amplify a 435-bp fragment from potato *Gpa2* BAC clone 111 (VAN DER VOSSEN *et al.* 2000), provided by J. Bakker, for subsequent use as an RFLP probe to map *Gpa2* in pepper, as described above. This probe corresponded to nucleotides 398–833 of the coding region of *Gpa2* (GenBank AF195939). In addition, two RFLP markers, GP34 (provided by C. Gephardt) and tomato clone CD19, were mapped to more accurately determine the location of the *Gpa2* gene.

***Me* and *Mech* loci:** Previously, RAPD marker Q04_0.3 was mapped in pepper 10.6 cM away from the nematode resistance locus *Me*₃ (LEFEBVRE *et al.* 1997; DJIAN-CAPORALINO *et al.* 2001). Previous mapping also revealed that a second nematode resistance locus, *Me*₄, maps 10 ± 4 cM away from *Me*₃ (DJIAN-CAPORALINO *et al.* 2001), and it was subsequently found that *Me*₁, *Me*₇, *Mech*₁, and *Mech*₂ could be inferred to localize to a region spanning ~17 cM telomeric to Q04_0.3 and ~10 cM centromeric to Q04_0.3. We mapped RAPD marker Q04_0.3 (OpQ04.300) using previously described methods (LEFEBVRE *et al.* 1997) in our segregating population. The map location of marker OpQ04.300 was used to infer the probable map locations of *Me* and *Mech* genes.

Mapping the *Bs2* gene in a tomato linkage map: A 500-bp DNA fragment of the *Bs2* gene was amplified from genomic DNA of *C. chacoense* (PI439414) using the primers *Bs2* L1 and *Bs2* R1 (TAI *et al.* 1999a). Amplification products were cloned and sequenced at the Cornell University Life Sciences Core Laboratory Center and used as an RFLP probe. Polymorphic bands were mapped in tomato using population filters provided by S. Tanksley.

Transient expression: *Rx* tagged with four HA epitope tags was constructed in the pB1 binary vector containing the *Rx* promoter and 3' sequence (*Rx*:4HA) as described (BENDAHDANE *et al.* 2002; PEART *et al.* 2002b). The NBLet sequence was deleted by overlapping PCR to create *Rx*:4HAΔNBLet. Binary vectors were transformed into the *Agrobacterium tumefaciens* strain C58C1 carrying the virulence plasmid pCH32. Agroinfiltration was performed as previously described (BENDAHDANE *et al.* 2000; PEART *et al.* 2002a). GFP fluorescence was evaluated 5 days later using a hand-held UV lamp. Protein extraction and immunoblotting were performed essentially as described by RAIDAN and MOFFETT (2006).

RESULTS

Primary sequence relationships: NB–LRR proteins homologous to *Bs2* were collected using the full-length *Bs2* protein sequence (AAF09256) in a search using BLASTP. Proteins were identified from both monocot and dicot plants and were mostly non-TIR–NB–LRR R proteins; TIR–NB–LRR matches to *Bs2* scored at or above $e = 10^{-19}$. All matches at or below this threshold were checked manually to determine if they had an experimentally established resistance function, thereby eliminating probable pseudogenes. These criteria produced a set of 35 previously characterized non-TIR NB–LRR proteins from both monocot and dicot plants (Table 1).

Amino acid sequence relationships of the NB–ARC region are a common criterion used to compare R proteins (CANNON *et al.* 2002). Aligned NB–ARC amino acid sequences were trimmed to the same length, and a sequence similarity diagram was generated (Figure 1A). Because recombination and sequence exchange drives the evolution of many R genes, we employed a neighbor-joining method for sequence analysis. Although it is not the most sophisticated method, neighbor-joining is not based on a continuum of sequence divergence that is an assumption required for parsimony and other models of phylogeny reconstruction (DOYLE and GAUT 2000). While recombination detection algorithms are being developed for nucleotide alignments, the divergence of our data set limited us to amino acid level comparisons. Figure 1A is therefore not our only measure of orthology, but critical in the organization of sequences for the following analyses.

From these comparisons of primary sequence, *Rx* and *Gpa2* emerged as the R proteins most closely related to *Bs2*. The high bootstrap values supporting this clade provide a high confidence for this grouping, which reflects the sum of random mutation and recombination among these homologs. While a second *Rx* paralog,

TABLE 1
NBS-LRR genes and accession numbers used in this study

Gene	GenBank accession no.		Plant species	Pathogen	Pathogen type
	Nucleotide	Protein			
<i>Bs2</i>	AF202179	AAF09256	<i>C. chacoense</i>	<i>X. campestris</i> pv. <i>vesicatoria</i>	Bacterium
<i>Dm3</i>	AF113947	AAD03156	<i>Lactuca sativa</i>	<i>Bremia lactucae</i>	Fungus
<i>Gpa2</i>	AF195939	AAF04603	<i>Solanum tuberosum</i> subsp. <i>andigena</i>	<i>Globodera pallida</i>	Nematode
<i>Hero</i>	AJ457052	AAF36987	<i>S. lycopersicum</i>	<i>Globodera rostochiensis</i>	Nematode
<i>HRT</i>	AAF36987	AAF36987	<i>A. thaliana</i>	Turnip crinkle virus	Virus
<i>I2</i>	AF118127	AAD27815	<i>S. lycopersicum</i>	<i>Fusarium oxysporum</i> sp. <i>lycopersici</i>	Fungus
<i>I2C-5</i>	AF408704	AAI01986	<i>S. pimpinellifolium</i>	<i>F. oxysporum</i> sp. <i>lycopersici</i>	Fungus
<i>Lr10</i>	AY270157	AAQ01784	<i>T. aestivum</i>	<i>Puccinia triticulturae</i>	Fungus
<i>Lr21</i>	AF532104	AAP74647	<i>T. aestivum</i>	<i>P. triticulturae</i>	Fungus
<i>Mi 1.2</i>	AF039682	AAC67238	<i>S. lycopersicum</i>	<i>Meloidogyne javanica</i>	Nematode
<i>Mla1</i>	AY009938	AAG37354	<i>Hordeum vulgare</i>	<i>Blumeria graminis</i> sp. <i>hordei</i>	Fungus
<i>Mla6</i>	AJ302293	CAC29242	<i>H. vulgare</i> subsp. <i>vulgare</i>	<i>B. graminis</i> sp. <i>hordei</i>	Fungus
<i>Mla12</i>	AY196347	AAO43441	<i>H. vulgare</i>	<i>B. graminis</i> sp. <i>hordei</i>	Fungus
<i>Mla13</i>	AF523678	AAO16000	<i>H. vulgare</i>	<i>B. graminis</i> sp. <i>hordei</i>	Fungus
<i>Pib</i>	AB013448	BAA76282	<i>Oryza sativa</i>	<i>Magnaporthe grisea</i>	Fungus
<i>Pi-ta</i>	AF207842	AAK00132	<i>O. sativa</i>	<i>M. grisea</i>	Fungus
<i>Pm3b</i>	AY325736	AAQ96158	<i>T. aestivum</i>	<i>B. graminis</i> f. sp. <i>tritici</i>	Fungus
<i>Prf</i>	U65391	AAC49408	<i>S. lycopersicum</i>	<i>Pseudomonas syringae</i> pv. <i>tomato</i>	Bacterium
<i>R1</i>	AF447489	AAI39063	<i>Solanum demissum</i>	<i>Phytophthora infestans</i>	Oomycete
<i>RB</i>	AY36128	AAP86601	<i>Solanum bulbocastanum</i>	<i>P. infestans</i>	Oomycete
<i>Rp1-D</i>	AF107293	AAd47197	<i>Zea mays</i>	<i>Puccinia sorghi</i>	Urediniomycetes
<i>Rp3</i>	AF489541	AAN23081	<i>Z. mays</i>	<i>P. sorghi</i>	Urediniomycetes
<i>Rpg1-b</i>	AY452684	AAR19095	<i>Glycine max</i>	<i>P. syringae</i>	Bacterium
<i>Rpi-blb1</i>	AY426259	AAR29069	<i>S. bulbocastanum</i>	<i>P. infestans</i>	Oomycete
<i>RPM1</i>	X87851	CAA61131	<i>A. thaliana</i>	<i>P. syringae</i>	Bacterium
<i>RPP8</i>	AF089710	AAC83165	<i>A. thaliana</i>	<i>Hyaloperonospora parasitica</i>	Oomycete
<i>RPP13-Nd</i>	AF209732	AAF42832	<i>A. thaliana</i>	<i>H. parasitica</i>	Oomycete
<i>RPP13-Rld</i>	AF209731	AAF42831	<i>A. thaliana</i>	<i>H. parasitica</i>	Oomycete
<i>RPS2</i>	U14158	AAA21874	<i>A. thaliana</i>	<i>P. syringae</i>	Bacterium
<i>RPS5</i>	AF074916	AAC26126	<i>A. thaliana</i>	<i>P. syringae</i> pv. <i>maculicola</i>	Bacterium
<i>Rx</i>	AJ011801	CAB50786	<i>S. tuberosum</i> subsp. <i>andigena</i>	Potato virus X	Virus
<i>Rx2</i>	AJ249448	CAB56299	<i>Solanum acaule</i>	Potato virus X	Virus
<i>Sw-5</i>	AY007366	AAG31013	<i>Solanum peruvianum</i>	Tomato spotted wilt virus	Virus
<i>Tm-2²</i>	AF536201	AAQ10736	<i>S. lycopersicum</i>	Tobacco mosaic virus	Virus
<i>Xa1</i>	AB002266	BAA25068	<i>O. sativa</i>	<i>Xanthomonas oryzae</i>	Bacterium

Rx2, has been identified and mapped to potato chromosome V (BENDAHMANE *et al.* 2000), it has more recently been shown that all sequences highly similar to *Rx/Gpa2* in two different diploid potatoes reside within the *Rx/Gpa2* cluster (BAKKER *et al.* 2003). This suggests that the presence of *Rx2* on chromosome V might represent a recent translocation event that is not widely conserved (BAKKER *et al.* 2003).

Predicted structural relationships: The effect of fractionation on phylogeny prompted us to seek other evidence of relationship among NB-LRRs. The N terminal, NB-ARC, and LRR domains of R proteins are further divided into subdomains and motifs. The methods and criteria for annotating these features vary

between reports, so to compare domains of *Bs2* with those of other R proteins, we revisited the domain prediction for all R proteins in this study (Table 1) to fill in missing information and to apply a consistent set of criteria to all sequences. Our analyses focused on key features of the N terminus, NB-ARC, and LRR domains (Figure 1B).

All of the proteins analyzed in this study are referred to as non-TIR R proteins, and the N termini are often reported to contain coiled-coil or leucine zipper domains. The protein sequences were reevaluated for coiled-coils using the programs COILS and Marcoil and a common set of criteria. The COILS program is commonly used for R-protein evaluation and employs a

FIGURE 1.—Sequence similarity relationships among Bs2 homologs. (A) A neighbor-joining consensus tree constructed from NBS domain protein sequences of previously cloned R genes. Bootstrap values $\geq 40\%$ are indicated. Following the protein sequence name are GenBank accession numbers and an abbreviated binomial of the organism: *A.thal* (*A. thaliana*), *C.chac* (*C. chacoense*), *G.max* (*Glycine max*), *H.vul* (*Hordeum vulgare*), *O.sat* (*Oryza sativa*), *S.acu* (*Solanum acaule*), *S.bul* (*Solanum bulbocastum*), *S.dem* (*Solanum demissum*), *S.lyc* (*Solanum lycopersicon*), *S.per* (*Solanum peruvianum*), *S.pimp* (*Solanum pimpinellifolium*), *S.tub* (*Solanum tuberosum*), *T.aes* (*Triticum aestivum*), and *Z.mays* (*Zea mays*). (B) To the right of each taxon description is a scale diagram showing the gene structure. Untranslated regions are represented by horizontal lines, introns by diagonal lines, and exons by colored bars. The colors represent the domains encoded by the sequence according to the key. Resistance genes that are closely related to each other, as shown in the tree, have similar size and placement of domains, introns, exons, and untranslated regions.

sliding window to evaluate the probability that a stretch of amino acids forms a coiled-coil (LUPAS *et al.* 1991; PAN *et al.* 2000b). The program Marcoil uses a hidden Markov model, which can be advantageous for recognizing shorter coiled-coils such as those believed to be found in R proteins (DELORENZI and SPEED 2002). Often the highest scores were at the N terminus as expected, but this domain was spuriously predicted elsewhere in the protein as well. For example, a typical false positive was found in the polyglutamate repeat in the LRR of Hero, which cannot physically form a coiled-coil (ERNST *et al.* 2002; GRUBER *et al.* 2006). We do not attempt to distinguish between regular coiled-coils and the leucine zipper subclass, but note that many leucine zippers reported in the R-gene literature were not predicted to be coiled-coils, even though a requisite pattern of leucine residues was present. In general, the

Marcoil and COILS programs were in agreement with few exceptions. However, the 14-amino-acid window of COILS gave many apparent false positives relative to the 21-amino-acid window.

Figure 2A, panels A–D, shows the predicted coiled-coils for HRT, Bs2, Rx, and Gpa2, with strong predictions in dark blue and weak predictions in light blue, and illustrates some of the previous discrepancies about coiled-coils in the N termini of these proteins. HRT and Gpa2 had been previously reported as having coiled-coils in the N terminus, while Rx was described as possessing a putative coiled-coil with a less conserved consensus, and Bs2 was noted to lack a coiled-coil domain (BENDAHDANE *et al.* 1999; COOLEY *et al.* 2000; TAI and STASKAWICZ 2000; VAN DER VOSSEN *et al.* 2000). On the basis of our updated analyses, however, Bs2 is more likely to possess a coiled-coil than is Gpa2. While a

FIGURE 2.—Comparison of N-terminal features of non-TIR R genes. (A) Coiled-coil prediction. Marcoil and COILS coiled-coil prediction outputs for different matrices or window sizes, as indicated, are represented graphically above the first 165 amino acids of selected protein sequences from Figure 1. The y-axis represents the percentage probability of forming a coiled-coil for each algorithm. The box below each graph represents the strength of the prediction, with dark blue for strongly predicted and light blue for weakly predicted coiled-coils. Side-by-side comparison predicts that HRT will form a coiled-coil, while Bs2, Rx, and Gpa2 will not. MTIDK stands for the coiled-coil proteins used in the prediction matrix: myosins, tropomyosins, intermediate filaments, desmosomal proteins, and kinesins. MTIDK14 (black dashed line) and MTIDK21 (red dashed line) indicate the size in amino acids of the sliding window used by the algorithm. 9FAM includes all nine families known to form coiled-coils and the MTK matrix is a smaller matrix including only myosins, tropomyosins, and kinesins. (B) An alignment of the N-terminal, non-TIR motif region described by BAI *et al.* (2002) and RAIRDAN *et al.* (2008) organized according to groups defined by the tree in Figure 1A. Amino acids are colored according to their properties. A consensus is shown above the alignment with letter size representing conservedness. A general pattern identified on the basis of amino acid properties is shown below the alignment (Φ, aromatic; u, aliphatic; +, basic; -, acidic; P, polar; and x, nonconserved).

distinction may have been made previously between the coiled-coil nature of the N termini of Rx and Gpa2, side-by-side comparison reveals that no substantive difference was detected by current algorithms. In the absence of experimental data demonstrating the existence of a coiled-coil structure in these R proteins, we suggest they should be more conservatively classified simply as non-TIR R proteins.

A motif that is fairly conserved in the N-terminal domain of most characterized non-TIR-NB-LRR proteins is the EDxxD motif, which is found adjacent to potential coiled-coils and forward of the NB-ARC

(BAI *et al.* 2002; RAIRDAN *et al.* 2008). The region WVxxIRELAYDIEDIVDxY was aligned among all of the R genes in our study and grouped according to clades identified in Figure 1A (also see Figure 2B). In general, the groupings produced by the NB-ARC alignment are mirrored in this motif. Previously, synapomorphies within the NB-ARC were found to correlate with the presence or the absence of a TIR N-terminal domain; this result revealed common patterns that can be found within the non-TIR N-terminal domain on the basis of NB-ARC relationships (PAN *et al.* 2000b). Exceptions to this trend are seen within the EDxxD portion of Bs2 and

FIGURE 3.—Duplicated elements in solanaceous R-gene clusters. Dot plots were generated within and between R-gene clusters to reveal duplicated elements. Homologous elements are shaded in the same colors. The scale bar represents size of the elements, while the distance between elements is presented in kilobase pairs. (A) The YAC sequence containing the *Bs2* gene was found to contain six elements duplicated two to three times, which filled much of the contig and 27-kb intron. (B) One of these elements is also present in the BAC sequence containing *Rx* and *Gpa2*. (A–F) The 5'-end of many solanaceous R genes was found duplicated to a proximal position. The length and percentage of sequence identity of these features are indicated. (C) *Mi 1.2* has two such duplications: one is the 5'-end and the other is within the gene at a position comparable to the other R genes that lack the 5' extension found on *Mi 1.2*.

the divergence of Dm3, RPS2, and RPS5. The use of the DiAlign algorithm (MORGENSTERN *et al.* 1998) allowed the motif of these latter sequences to be aligned on the basis of similarity outside the EDxxD motif, but as noted previously, this clade seems to lack the most conserved portion of the motif (RAIRDAN *et al.* 2008). Given this data set, a slightly modified consensus for this region was observed (Figure 2B). Considering amino acid properties, a general pattern is suggested and described in the Figure 2 legend.

The structural annotation was revised for two other R-protein regions. A hydrophobic region within the NB-ARC (GxP or GLPL) has been shown to be important for R-gene function (RAIRDAN and MOFFETT 2006), but also several other regions have been noted as being hydrophobic in first reports of R-gene isolations. Since criteria used by authors vary, we again applied common criteria for prediction and annotation of hydrophobic domains across the R proteins examined. A Kyte–Doolittle plot was used to analyze hydrophobicity (Figure 1B; indicated in purple) (KYTE and DOOLITTLE 1982). LRRs are not necessarily contiguous, which further complicates their delineation. In our analyses, two types of interruptions were found: (1) gaps in Rp1-D and the polyglutamate repeat in Hero and (2) super-

imposition of alternate domains, as predicted by other methods, on the LRR pattern. LRR domains are shown in red in Figure 1B, and our reevaluation was useful in delimiting the ends of these domains as structural features in our analysis.

Sequence relationships of the noncoding regions near and within R genes: We interpret the intron position within R genes (BAI *et al.* 2002; MEYERS *et al.* 2003) as an indicator of orthology relationships. Introns and exons, both within the coding region and in the 5'- and 3'-UTR, are shown in Figure 1B. Visual comparison of the placement of noncoding regions further demonstrates the striking similarity between closely related R genes. We were intrigued by the extreme 27-kb length of the *Bs2* intron. Dot plots aligning the *Bs2* YAC with itself as well as BLAST and CENSOR searches (KOHANY *et al.* 2006) were employed to investigate this phenomenon (supporting information, Figure S1). The *Bs2* intron contains six major duplicated elements (Figure 3 and Table S1). Portions of the intron were found to bear similarity to the internal regions of two Gypsy-type LTR retrotransposons, Ogre (MACAS and NEUMANN 2007) and GYPOT1 (JURKA and SHANKAR 2006a). There were several interruptions in the partial alignment with the Ogre element. The region with similarity to GYPOT1 is

FIGURE 4.—Effect of the Rx NBLet on Rx function. (A) PVX resistance conferred by Rx in its complete genomic context (Rx:4HA) and without the Rx NBLet (Rx:4HA Δ NBLet). Rx constructs were transiently coexpressed in *Nicotiana benthamiana* leaves with an infectious PVX:GFP clone via agroinfiltration. PVX:GFP accumulation was monitored as GFP fluorescence under UV light at 5 d after infiltration. PVX:GFP infiltrated in the absence of Rx is shown at bottom left for comparison. (B) Western blot analysis showing Rx accumulation when expressed with and without NBLet. *N. benthamiana* leaves were infiltrated with *Agrobacterium*-containing Rx constructs, and samples were collected 2 days later. Rx protein levels were detected by anti-HA immunoblotting.

flanked by direct repeats (Figure 3A-e). Much of the intronic region is duplicated to the surrounding regions as well. Duplicated portions of the Ogre element (Figure 3A-f) and the repeats flanking GYPOT1 (Figure 3A-d) indicate local movement of large retroelements as does another region with similarity to a Copia LTR retrotransposon (KOHANY and JURKA 2007). Smaller portions of the intron are also duplicated: Figure 3A-g harbors an Alien element (POZUETAROMERO *et al.* 1996) and Figure 3A-d is found separately within the intron. Nonautonomous Alien DNA transposons were distributed rather evenly across the YAC, but other elements of the same class, Sonata (JURKA 2006a,b; JURKA and SHANKAR 2006b), were much more numerous and tended to cluster. Non-LTR retroelements (YOSHIOKA *et al.* 1993) were also observed in the vicinity of *Bs2* as well as an additional hAT DNA transposon (JURKA and KOHANY 2006). Other duplicated fragments were observed but bore no similarity to transposons. In general, there is an erosion and truncation of transposable-element-related sequences consistent with multiple insertions followed by sequence drift. The sequence encompassing the GYPOT1 element is also found in the flanking regions of other solanaceous R genes, specifically *Rx* and *Gpa2*. Interestingly, most BLAST hits to this particular transposon-related sequence were associated with R-gene clusters in various plants. An abundance of similar retrotransposons has been reported in tomato (DATEMA *et al.* 2008) and has generally been associated with genome expansion.

Another notable duplicated feature is the presence of a fragment of the *Bs2* gene, specifically a portion of the

5'-end of the gene repeated past the 3'-end of the functional gene. Other solanaceous R genes were tested for similar truncated NB-LRRs, or "NBLets" (Figure 3), because of a similar report for *Tm2²* (LANFERMEIJER *et al.* 2003). Only *Rx* and *RGC3*, a pseudogene near *Rx*, share the same type of trailing 5' gene fragment as *Bs2*; the absence of a NBLet for *Gpa2* may be due to the loss of its terminal exon as compared to *Rx*. NBLets were also found for *Mi 1.2*, *R1*, *RB*, and *Tm2²*, but these were 5' to the coding sequence. Tandem duplications have been implicated in affecting gene expression/activity through the generation of small interfering RNAs. This phenomenon has been observed specifically in R-gene clusters (YI and RICHARDS 2007). To determine whether the duplicated region of *Rx* plays a functional role in *Rx*-mediated resistance to PVX, protein levels and resistance responses were compared in transient agro-expression assays between *Rx* genomic constructs with and without this duplication. A 300-bp region encompassing the 3' duplication was deleted from a binary vector containing the *Rx* promoter, the *Rx* coding region fused to four HA epitope tags, and 3' sequences. *Rx* constructs were coexpressed with a GFP-tagged version of PVX, such that in this assay, *Rx* efficacy is inversely correlated with the amount of GFP fluorescence observed (RAIRDAN and MOFFETT 2006). The deletion of the *Rx* NBLet had no notable effect on the ability of *Rx* to confer resistance to PVX or on the level of *Rx* protein expressed (Figure 4) in this system. These results rule out the possibility that the *Rx* NBLet expresses a protein fragment required for *Rx* function and suggests that the NBLet does not alter *Rx* protein expression levels.

Macrosyntentic relationships of *Bs2*, fractionated orthologs, and paralogs: *Bs2* belongs to a large gene family in *Capsicum* as demonstrated by numerous bands in DNA blot analysis (TAI *et al.* 1999b). Therefore, it was not practical to directly map the *Bs2* gene by RFLP or identify paralogs. PCR approaches using portions of the *Bs2* gene have been employed, but could potentially identify paralogs located nearby or duplicated elsewhere in the genome. To test our hypothesis and to identify the position of *Bs2* in the pepper genome unequivocally, we used DNA markers tightly linked to *Bs2* (TAI *et al.* 1999a; TAI and STASKAWICZ 2000).

A2 is a marker that is tightly linked to the *Bs2* gene (0.0 cM) and resides on the YAC clone containing *Bs2* (TAI *et al.* 1999b; TAI and STASKAWICZ 2000). We cloned the A2 genomic DNA fragment from the *C. annuum* cultivar ECW-123R (*Bs2*/*Bs2*) and used it as an RFLP probe. Two dominant polymorphic bands were detected on survey filters for our comparative mapping population (LIVINGSTONE *et al.* 1999; GRUBE *et al.* 2000b). The 1-kb band mapped between markers TG263 and CT121 on the lower arm of pepper chromosome 9 (P9; Figure 5A and Figure 6), while the 2-kb band was assigned to P1.

FIGURE 5.—Molecular marker polymorphisms related to *Bs2* homologs. (A) RFLP polymorphism of STS marker A2 on pepper DNA digested with *Taq*I. Note polymorphic bands of 1 kb, which maps to pepper chromosome P9, and of 2 kb between mapping population parents R Naky and PI159234. (B) SCAR marker S19 amplified using parents. Note polymorphism between mapping population parents R Naky and PI159234. (C) RFLP polymorphism of potato *Gpa2* fragment on pepper DNA digested with *Bst*NI. The bands marked “a” and “b” are found only in mapping population parent PI159234, while the band marked “c” is found only in mapping population parent R Naky. (D) *Bs2* fragment as an RFLP probe on tomato DNA. Note multiple bands of varying intensity. (E) RAPD marker Q4_0.300 in pepper. Note polymorphisms between mapping population parents R Naky and PI159234.

To determine which of these loci corresponded to the *Bs2* gene, another linked marker, S19, was used. S19 is a codominant SCAR marker located 7 cM from *Bs2* and A2 (TAI *et al.* 1999a). The same pair of polymorphic bands identified during the cloning of *Bs2*, 220 bp in ECW and R Naky, 240 bp in ECW-123R and PI159234, were amplified in our mapping parents (Figure 5B). S19 was mapped to P9, ~6 cM below TG263 and centromeric with respect to the position for the 1-kb band of A2 (Figure 6), demonstrating that *Bs2* resides on P9 in a region of the pepper genome that is orthologous to the top arm of potato XII that includes the fractionated orthologs *Rx* and *Gpa2*. This is consistent with results from others who have located several PCR markers with similarity to *Bs2* on this chromosome (OGUNDWIN *et al.* 2005; SUGITA *et al.* 2006; DJIAN-CAPORALINO *et al.* 2007).

To further test our hypothesis of shared synteny between fractionated R-gene orthologs, RFLP probes corresponding to *Gpa2* were mapped in pepper and *Bs2* probes were mapped in tomato. The probe derived from the 5'-end of *Gpa2* hybridized to an average of 11 bands on pepper genomic survey blots. The prominent polymorphic bands were mapped, and all localized to a region on P9 each 3 cM from marker TG263 (Figure 5C and Figure 6). While *Bs2* is a member of a large gene

family in pepper, it produces few bands on tomato genomic DNA survey blots (Figure 5D and TAI *et al.* 1999b). While one of the major polymorphic bands mapped to tomato chromosome 2, the other mapped to tomato chromosome 12, between CT100 and CT129, which tightly flank *Rx* and *Gpa2* in potato and broadly flank *Bs2* and *Gpa2* homologs in pepper (Figure 6).

Tomato and potato are collinear throughout this arm of the chromosome, differing by only a whole-arm inversion. Pepper is collinear with both tomato and potato in this region except CT100 is centromeric in both pepper and tomato, but near the telomere in potato. This deviation signifies the breakpoint between the inversions. Further, this breakpoint, between CT100 and CT129 in potato compared with pepper, is the location of the R-gene cluster, providing a plausible explanation for the dispersal of R genes to the ends of this inverted region (TG180/*Mi* 3 and CT100/*Lv* in tomato Figure 6). This hypothesis predicts that other R genes similar to *Rx*/*Gpa2* are localized near CT129 in pepper, the other end of the inversion breakpoint. The marker OpQ04.300 is linked to nematode resistances and shared the same polymorphism between our mapping parents as observed in previous studies, allowing us to also demonstrate the presence of *Me1*, *Me3*, *Me4*, *Me7*, *Mech1*, and *Mech2* in this region on a comparative map

FIGURE 6.—Comparative genetic map for pepper chromosome P9. Scale representation of tomato chromosome 12, pepper chromosome P9, and potato chromosome XII aligned via shared molecular markers. Markers are placed at LOD 3 or greater unless shown in parentheses, in which case they were placed at LOD 2 or greater. R-gene names are italicized. The location of multiple pepper homologs of *Gpa2* (*pepGpa2*) is indicated with a bracket, as are the positions of the *Me* and *Mech* nematode resistances.

that aligns with other genera with multiple single-copy linkages (Figure 5E and Figure 6) (LIVINGSTONE *et al.* 1999; DJIAN-CAPORALINO *et al.* 2001, 2007).

Critical breakpoints in this updated alignment occur at telomeric/centromeric regions or near R genes. For the *Rx/Gpa2* cluster, a chromosome translocation/inversion breakpoint dispersed the R-gene homologs and associated markers in other genera relative to potato. Others have observed the genomic distribution of R genes as a somewhat random phenomenon (LEISTER *et al.* 1998; PAN *et al.* 2000a; RICHLY *et al.* 2002), but it has been since shown in *Arabidopsis* that R-gene locations are consistent with the rearrangements of their chromosomal context (BAUMGARTEN *et al.* 2003). In the Solanaceae, 22 genome rearrangements distinguish tomato and pepper (LIVINGSTONE *et al.* 1999). The analysis of GRUBE *et al.* (2000b) and subsequent R-gene discovery in tomato, pepper, and potato described here were combined to examine the association of R genes with chromosome breakpoints (Figure 7). In every case, we could associate at least one source of resistance with every breakpoint. Despite the limitation of only including NB-LRRs near breakpoints that have a known phenotype, the sequence relationship between *Hero* and *Prf* is seen to be reflected in their genomic relationship. As shown by BAUMGARTEN *et al.* (2003), this sequence relationship is not expected for every R gene that can be aligned in the genome because clusters are heterogeneous. This hypothesis can be further tested in a comparative system when the completed tomato genome sequence allows these

comparisons to be made at a higher resolution across the Solanaceae.

DISCUSSION

The ability to determine orthology is critical in the application of comparative genomics to questions of R-gene evolution, function, and discovery. Here we investigate the homology relationships of *Bs2*, a major gene in *Capsicum* for resistance to bacterial spot and other non-TIR NB-LRRs. From analyses of sequence, gene architecture, predicted protein structure, and macrosynteny, *Bs2* is a fractionated ortholog of members of the *Rx/Gpa2* locus. In contrast to monocots, recent reports in dicots illustrate fractionated synteny, and microcolinearity can be found across genera for R genes and other tandemly duplicated genes (BALLVORA *et al.* 2007; SCHLUETER *et al.* 2008). Approaches to understanding and utilizing R-gene macrosynteny in the Solanaceae are certainly viable. The cloning of the late blight resistance gene *R3a* from potato based on *I2* in tomato illustrates the potential of these comparative approaches (HUANG *et al.* 2005). Extensions of our model provide for the *a priori* localization of cloned R-gene sequences in one species on the basis of the genomic location of its fractionated ortholog in a model species and a selection criterion for candidate sequences where resistances align in comparative maps. Comparative maps are critical for understanding and identifying rearrangement breakpoints that fragment these relationships.

FIGURE 7.—Colocalization of R-gene clusters and chromosome breakpoints on comparative maps of tomato and pepper. Tomato chromosomes (solid) and pepper chromosomes (open) are shown aligned on the basis of shared markers as presented in LIVINGSTONE *et al.* (1999). A meta-analysis of the R-gene position in solanaceous genomes by GRUBE *et al.* (2000b) was continued with a focus on R genes that occurred at chromosome breakpoints on the comparative map. A subset of pepper R genes are shown next to pepper chromosomes, tomato R genes are shown next to the tomato chromosomes, and potato R genes are underlined. Representative RFLP markers linked to the R genes are given in parentheses. Chromosomal inversions are indicated with a circular arrow, and chromosomes that are separated by translocations are connected by dotted lines.

The clustering of R genes and the complex recombination of paralogs within clusters pose a special challenge in studies of their evolution. This recombination results in varying levels of sequence exchange through a form of *in vivo* DNA shuffling that generates diversity as well as gain and loss of R genes (MICHELMORE and MEYERS 1998; SONG *et al.* 2002). We qualify this claim of orthology and acknowledge the “fractionation” of the evolutionary history of R genes (recombination, duplication, and deletion) that also does not allow for the creation of a true phylogeny by conventional methods. These limitations lead to our application of additional criteria that can be used to support a common history. It has been noted that for some regions of the R genes, sequence similarity is not sufficient to allow the pairing required for conventional recombination (WICKER *et al.* 2007). The frequently reported association of transposable-element-derived sequences within R-gene clusters provides the requisite conserved sequences for recombination to occur, and the presence of transposons elsewhere in the genome would provide a means for

interchromosomal sequence exchange (MEYERS *et al.* 2003).

The relationship of R-gene clusters and chromosome rearrangement warrants further investigation. The breakpoint of the translocation that differentiates pepper chromosome 9, potato chromosome XII, and tomato chromosome 12 is apparently in or near an R-gene cluster (LIVINGSTONE *et al.* 1999; GRUBE *et al.* 2000b). This phenomenon is repeated throughout comparative maps of the Solanaceae and is summarized in Figure 7. Every chromosomal rearrangement breakpoint is associated with one or more R genes or mapped resistances. Previously, it has been shown that genomic rearrangement and duplication are significant sources of R-gene dispersal and duplication, which is further complicated by ancestral polyploids (BAUMGARTEN *et al.* 2003; AMELINE-TORREGROSA *et al.* 2008). Increased sequence information will provide resolution of the precise relationship of R-gene clusters, embedded transposons, and chromosome breakpoints that may be detected in comparisons between related genera. These processes

also may explain the dramatic expansion observed in some R-gene clusters (MEYERS *et al.* 1998a). The translocation breakpoint proposed within the *Rx/Gpa2* cluster would result in these sequences being dispersed to the centromeric and telomeric regions of the lower arm of P9 (Figure 6). Two other modes of expansion were witnessed in dot plots of the genomic sequences of these R-gene regions that would further disperse R genes in the genome (Figure S1 and Figure S2). This hypothesis can be further tested in a comparative system when genome sequence allows these comparisons to be made at a higher resolution across the Solanaceae.

The clustering of these duplicated sequences also leads to unique regulatory and functional properties (FRIEDMAN and BAKER 2007; TAM *et al.* 2008). Small RNAs have been described as coordinately regulating these R-gene clusters at a post-transcriptional level (Yi and RICHARDS 2007). We speculated that NBLets adjacent to functional R genes may also have a role in this regulatory process. Our test of the effect on expression and function of *Rx* with and without an adjacent NBLet did not detect any differences, indicating that the *Rx* mRNA expression, stability, and translation are unaffected by the NBLet *per se*. It remains possible, however, that the NBLet may affect local chromatin structure in its endogenous context, which in turn could affect R-gene expression levels (FRIEDMAN and BAKER 2007). NBLets are not exclusive to the non-TIR NB-LRR class of R genes; they have been noted in the TIR class as well (GRAHAM *et al.* 2002). *Xa21*, a member of a third class of R genes, has been shown to have a highly conserved 5' domain that is important for mediating recombination between genetically linked paralogs (SONG *et al.* 1997).

A subset of R genes, including *Bs2* and *Rx*, provide durable resistances, but as others are overcome, there is a need for new sources of resistance and to reconsider approaches to using these genes in crop improvement (McDOWELL and WOFFENDEN 2003; LECOQ *et al.* 2004). According to our model, large-scale sequencing of the easily obtained NB-ARC domains from a new source of resistance can be grouped by sequence similarity, and these groups will reflect corresponding R-gene clusters on comparative maps of reference plants. This strategy can be tested in application to the cloning of *Lv* and *Mi 3* in tomato using homology to *Bs2* and *Rx/Gpa2* as references (Figure 6). The importance of the genetic background of a plant is linked to the complexity of R-gene clustering and mechanism and poses different challenges to plant breeders. The introgression of a new resistance gene will occur at regions of the genome that may already contain resistance genes (MICHELMORE 2003). So-called "jackpot" cultivars can be seen as a source of cassettes of resistances and contain clusters of many tightly linked resistances (GRUBE *et al.* 2000a). However, merging selected genes of these clusters is a much more daunting prospect.

A major barrier to understanding R-gene similarities and function is the lack of structural information. Successes in this area have so far capitalized on regions of shared similarity and homology modeling in the NB and ARC domains and, to some degree, the highly divergent LRR, but the N-terminal domain of the non-TIR class lacks this benefit (McHALE *et al.* 2006; TAKKEN *et al.* 2006; CHATTOPADHYAYA and PAL 2008). While some motifs have been found in these variable regions, the major feature ascribed to these proteins, a coiled-coil domain, has never been demonstrated, only computationally predicted. Since oligomerization of NB-LRR-associated coiled-coil domains has not been reported, and cellular proteins that interact with this domain show no common structures, it would seem that the existence of a coiled-coil structure either has a role in protein conformation or is simply an artifact of the prediction programs (DEYOUNG and INNES 2006).

The convoluted history of R-gene diversity is being explicated with increasing resolution. Comparative studies are one tool to investigate shared aspects of R genes, but often reveal striking differences that are fundamental to their evolution and mode of action; R proteins at once must be highly adaptive to changing pathogens, yet retain sufficient similarity to interface with host proteins and signal transduction networks. Elucidating the mechanisms of genome-level processes that have operated in different lineages is a key step both in reaching translational goals and in determining the factors that govern the evolution of this gene family.

We thank B. Staskawicz and R. Freedman for providing the Capsicum YAC clone sequence, C. Gephardt for providing us with the GP34 potato clone, J. Bakker for the potato *Gpa2* clone, and R. Stahl for providing us with *C. annuum* ECW123 seed. We are grateful to Greg Rairdan for generating initial *Rx* constructs and to M. Sacco for coining the term NBLet. Our gratitude to R. Grube, B. Baker, A. Bent, and J. Rouppe van der Voort for helpful conversations regarding the mapping of R genes in the Solanaceae and K. Perez for critical review of the manuscript. This work was supported in part by the National Science Foundation (NSF; DBI-0218166 and IOB-0343327 to M.J. and IOS-0744652 to P.M.). M.M. was supported by a Barbara McClintock Award (Robert Rabson), the Olin Fellowship, the College of Agriculture and Life Sciences (University of Wisconsin, Madison) Dean's fund, and a gift from Kalsec. S.M.C. was supported by an NSF Graduate Research Fellowship. B.-C.K. received support from U. S. Department of Agriculture Initiative for Future Agricultural and Food Systems Award no. 2001-52100-113347 and NSF Plant Genome Award no. 0218166.

LITERATURE CITED

- AMELINE-TORREGROSA, C., B.-B. WANG, M. S. O'BLENESS, S. DESHPANDE, H. ZHU *et al.*, 2008 Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* **146**: 5–21.
- BAI, J. F., L. A. PENNILL, J. C. NING, S. W. LEE, J. RAMALINGAM *et al.*, 2002 Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res.* **12**: 1871–1884.
- BAKKER, E., P. BUTTERBACH, J. ROUPPE VAN DER VOORT, E. VAN DER VOSSEN, J. VAN VLIET *et al.*, 2003 Genetic and physical mapping of homologues of the virus resistance gene *Rx1* and the cyst nematode resistance gene *Gpa2* in potato. *Theor. Appl. Genet.* **106**: 1524–1531.

- BALLVORA, A., A. JOCKER, P. VIEHOVER, H. ISHIHARA, J. PAAL *et al.*, 2007 Comparative sequence analysis of Solanum and Arabidopsis in a hot spot for pathogen resistance on potato chromosome V reveals a patchwork of conserved and rapidly evolving genome segments. *BMC Genomics* **8**: 112.
- BAUMGARTEN, A., S. CANNON, R. SPANGLER and G. MAY, 2003 Genome-level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics* **165**: 309–319.
- BENDAHMANE, A., K. KANYUKA and D. C. BAULCOMBE, 1999 The *Rx* gene from potato controls separate virus resistance and cell death responses. *Plant Cell* **11**: 781–791.
- BENDAHMANE, A., M. QUERCI, K. KANYUKA and D. C. BAULCOMBE, 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.
- BENDAHMANE, A., G. FARNHAM, P. MOFFETT and D. C. BAULCOMBE, 2002 Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the *Rx* locus of potato. *Plant J.* **32**: 195–204.
- BLUM, E., M. MAZOUREK, M. O'CONNELL, J. CURRY, T. THORUP *et al.*, 2003 Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum*. *Theor. Appl. Genet.* **108**: 79–86.
- CANNON, S. B., H. Y. ZHU, A. M. BAUMGARTEN, R. SPANGLER, G. MAY *et al.*, 2002 Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *J. Mol. Evol.* **54**: 548–562.
- CHATTOPADHYAYA, R., and A. PAL, 2008 Three-dimensional models of NB-ARC domains of disease resistance proteins in tomato, *Arabidopsis*, and flax. *J. Biol. Struct. Dyn.* **25**: 357–371.
- COOLEY, M. B., S. PATHIRANA, H. J. WU, P. KACHROO and D. F. KLESIG, 2000 Members of the *Arabidopsis HRT/RPP8* family of resistance genes confer resistance to both viral and oomycete pathogens. *Plant Cell* **12**: 663–676.
- DATEMA, E., L. A. MUELLER, R. BUELS, J. J. GIOVANNONI, R. G. F. VISSER *et al.*, 2008 Comparative BAC end sequence analysis of tomato and potato reveals overrepresentation of specific gene families in potato. *BMC Plant Biol.* **8**: 34.
- DELORENZI, M., and T. SPEED, 2002 An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**: 617–625.
- DEYOUNG, B. J., and R. W. INNES, 2006 Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat. Immunol.* **7**: 1243–1249.
- DJIAN-CAPORALINO, C., L. PIJAROWSKI, A. FAZARI, M. SAMSON, L. GAVEAU *et al.*, 2001 High-resolution genetic mapping of the pepper (*Capsicum annuum* L.) resistance loci *Me₃* and *Me₄* conferring heat-stable resistance to root-knot nematodes (*Meloidogyne* spp.). *Theor. Appl. Genet.* **103**: 592–600.
- DJIAN-CAPORALINO, C., A. FAZARI, M. J. ARGUEL, T. VERNIE, C. VANDECASTEELE *et al.*, 2007 Root-knot nematode (*Meloidogyne* spp.) *Me* resistance genes in pepper (*Capsicum annuum* L.) are clustered on the P9 chromosome. *Theor. Appl. Genet.* **114**: 473–486.
- DOYLE, J. J., and B. S. GAUT, 2000 Evolution of genes and taxa: a primer. *Plant Mol. Biol.* **42**: 1–23.
- ELLIS, J., P. DODDS and T. PRYOR, 2000 Structure, function and evolution of plant disease resistance genes. *Curr. Opin. Plant Biol.* **3**: 278–284.
- ERNST, K., A. KUMAR, D. KRISELEIT, D. U. KLOOS, M. S. PHILLIPS *et al.*, 2002 The broad-spectrum potato cyst nematode resistance gene (*Hero*) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. *Plant J.* **31**: 127–136.
- FLUHR, R., 2001 Sentinels of disease: plant resistance genes. *Plant Physiol.* **127**: 1367–1374.
- FRIEDMAN, A. R., and B. J. BAKER, 2007 The evolution of resistance genes in multi-protein plant resistance systems. *Curr. Opin. Genet. Dev.* **17**: 493–499.
- GALE, M. D., and K. M. DEVOS, 1998 Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* **95**: 1971–1974.
- GLAZE BROOK, J., 2005 Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **43**: 205–227.
- GRAHAM, M. A., L. F. MAREK and R. C. SHOEMAKER, 2002 Organization, expression and evolution of a disease resistance gene cluster in soybean. *Genetics* **162**: 1961–1977.
- GRUBE, R. C., J. R. BLAETH, M. S. ARNEDEO ANDRÉS, C. CARANTA and M. K. JAHN, 2000a Identification and comparative mapping of a dominant potyvirus resistance gene cluster in *Capsicum*. *Theor. Appl. Genet.* **101**: 852–859.
- GRUBE, R. C., E. R. RADWANSKI and M. JAHN, 2000b Comparative genetics of disease resistance within the Solanaceae. *Genetics* **155**: 873–887.
- GRUBER, M., J. SODING and A. N. LUPAS, 2006 Comparative analysis of coiled-coil prediction methods. *J. Struct. Biol.* **155**: 140–145.
- HUANG, S., E. A. G. VAN DER VOSSEN, H. KUANG, V. G. VLEESHOUWERS, N. ZHANG *et al.*, 2005 Comparative genomics enabled the isolation of the *R3a* late blight resistance gene in potato. *Plant J.* **42**: 251–261.
- HULBERT, S. H., C. A. WEBB, S. M. SMITH and Q. SUN, 2001 Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* **39**: 285–312.
- JONES, D. A., and J. D. G. JONES, 1997 The role of leucine-rich repeat proteins in plant defences, pp. 89–167 in *Advances in Botanical Research*, edited by J. H. ANDREWS, I. C. TOMMERUP and J. A. CALLOW. Academic Press, San Diego/New York/London.
- JURKA, J., 2006a SONATA1: non-autonomous DNA transposon from *Solanum* species. *Repbases Reports* **6**: 267.
- JURKA, J., 2006b SONATA2: non-autonomous DNA transposon from *Solanum* species. *Repbases Reports* **6**: 268.
- JURKA, J., and O. KOHANY, 2006 hAT-2, SD hAT transposon from *Solanum demissum*. *Repbases Reports* **6**: 492.
- JURKA, J., and R. SHANKAR, 2006a GYPOT1: gypsy-type element from black cottonwood. *Repbases Reports* **6**: 488.
- JURKA, J., and R. SHANKAR, 2006b SONATA3: non-autonomous DNA transposon from *Solanum* species. *Repbases Reports* **6**: 508.
- KOHANY, O., and J. JURKA, 2007 LTR retrotransposons from grapevine. *Repbases Reports* **7**: 696.
- KOHANY, O., A. J. GENTLES, L. HANKUS and J. JURKA, 2006 Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* **7**: 474.
- KUANG, H., S. S. WOO, B. C. MEYERS, E. NEVO and R. W. MICHELMORE, 2004 Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* **16**: 2870–2894.
- KUMAR, S., K. TAMURA, I. B. JAKOBSEN and M. NEI, 2001 MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244–1245.
- KYTE, J., and R. F. DOOLITTLE, 1982 A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105–132.
- LANFERMEIJER, F. C., J. DIJKHUIS, M. J. G. STURRE, P. DE HAAN and J. HILLE, 2003 Cloning and characterization of the durable tomato mosaic virus resistance gene *Tm-2²* from *Lycopersicon esculentum*. *Plant Mol. Biol.* **52**: 1039–1051.
- LECOQ, H., B. MOURY, C. DESBIEZ, A. PALLOIX and M. PITRAT, 2004 Durable virus resistance in plants through conventional approaches: a challenge. *Virus Res.* **100**: 31–39.
- LEFEBVRE, V., C. CARANTA, S. PFLIEGER, B. MOURY, A. M. DAUBEZE *et al.*, 1997 Updated intraspecific maps of pepper. *Capsicum & Eggplant Newsletter* **13**: 35–41.
- LEISTER, D., J. KURTH, D. A. LAURIE, M. YANO, T. SASAKI *et al.*, 1998 Rapid reorganization of resistance gene homologues in cereal genomes. *Proc. Natl. Acad. Sci. USA* **95**: 370–375.
- LEVY, M., O. EDELBAUM and I. SELA, 2004 Tobacco mosaic virus regulates the expression of its own resistance gene *N*. *Plant Physiol.* **135**: 2392–2397.
- LIVINGSTONE, K. D., V. K. LACKNEY, J. R. BLAETH, R. VAN WIJK and M. K. JAHN, 1999 Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics* **152**: 1183–1202.
- LUPAS, A., M. VANDYKE and J. STOCK, 1991 Predicting coiled coils from protein sequences. *Science* **252**: 1162–1164.
- MACAS, J., and P. NEUMANN, 2007 Ogre elements: a distinct group of plant Ty3/gypsy-like retrotransposons. *Gene* **390**: 108–116.
- MARTIN, G. B., A. J. BOGDANOVE and G. SESSA, 2003 Understanding the functions of plant disease resistance proteins. *Annu. Rev. Plant Biol.* **54**: 23–61.
- MCDOWELL, J. M., and S. A. SIMON, 2006 Recent insights into *R* gene evolution. *Mol. Plant Pathol.* **7**: 437–448.

- McDOWELL, J. M., and B. J. WOFFENDEN, 2003 Plant disease resistance genes: recent insights and potential applications. *Trends Biotechnol.* **21**: 178–183.
- McHALE, L., X. TAN, P. KOEHL and R. W. MICHELMORE, 2006 Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* **7**: 212.
- MEYERS, B. C., D. B. CHIN, K. A. SHEN, S. SIVARAMAKRISHNAN, D. O. LAVELLE *et al.*, 1998a The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* **10**: 1817–1832.
- MEYERS, B. C., K. A. SHEN, P. ROHANI, B. S. GAUT and R. W. MICHELMORE, 1998b Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *Plant Cell* **10**: 1833–1846.
- MEYERS, B. C., A. KOZIK, A. GRIEGO, H. H. KUANG and R. W. MICHELMORE, 2003 Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**: 809–834.
- MICHELMORE, R. W., 2003 The impact zone: genomics and breeding for durable disease resistance. *Curr. Opin. Plant Biol.* **6**: 397–404.
- MICHELMORE, R. W., and B. C. MEYERS, 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., B. C. MEYERS, R. W. MICHELMORE and B. S. GAUT, 2002 Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res.* **12**: 1305–1315.
- MORGENSTERN, B., K. FRECH, A. DRESS and T. WERNER, 1998 DIALIGN: finding local similarities by multiple sequence alignment. *Bioinformatics* **14**: 290–294.
- NURNBERGER, T., F. BRUNNER, B. KEMMERLING and L. PIATER, 2004 Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol. Rev.* **198**: 249–266.
- OGUNDIWIN, E. A., T. F. BERKE, M. MASSOUDI, L. L. BLACK, G. HUESTIS *et al.*, 2005 Construction of 2 intraspecific linkage maps and identification of resistance QTLs for *Phytophthora capsici* root-rot and foliar-blight diseases of pepper (*Capsicum annuum* L.). *Genome* **48**: 698–711.
- PAN, Q. L., Y. S. LIU, O. BUDAI-HADRIAN, M. SELA, L. CARMEL-GOREN *et al.*, 2000a Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and *Arabidopsis*. *Genetics* **155**: 309–322.
- PAN, Q. L., J. WENDEL and R. FLUHR, 2000b Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* **50**: 203–213.
- PATERSON, A. H., J. E. BOWERS, D. G. PETERSON, J. C. ESTILL and B. A. CHAPMAN, 2003 Structure and evolution of cereal genomes. *Curr. Opin. Genet. Dev.* **13**: 644–650.
- PEART, J. R., G. COOK, B. J. FEYS, J. E. PARKER and D. C. BAULCOMBE, 2002a An *EDSI* orthologue is required for N-mediated resistance against tobacco mosaic virus. *Plant J.* **29**: 569–579.
- PEART, J. R., R. LU, A. SADANANDOM, I. MALCUIT, P. MOFFETT *et al.*, 2002b Ubiquitin ligase-associated protein SGT1 is required for host and nonhost disease resistance in plants. *Proc. Natl. Acad. Sci. USA* **99**: 10865–10869.
- POZUETAROMERO, J., G. HOULNE and R. SCHANTZ, 1996 Non-autonomous inverted repeat Alien transposable elements are associated with genes of both monocotyledonous and dicotyledonous plants. *Gene* **171**: 147–153.
- RAIRDAN, G. J., and P. MOFFETT, 2006 Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation. *Plant Cell* **18**: 2082–2093.
- RAIRDAN, G., and P. MOFFETT, 2007 Brothers in arms? Common and contrasting themes in pathogen perception by plant NB-LRR and animal NACHT-LRR proteins. *Microbes Infect.* **9**: 677–686.
- RAIRDAN, G. J., S. M. COLLIER, M. A. SACCO, T. T. BALDWIN, T. BOETRICH *et al.*, 2008 The coiled-coil and nucleotide binding domains of the potato Rx disease resistance protein function in pathogen recognition and signaling. *Plant Cell* **20**: 739–751.
- RICHLY, E., J. KURTH and D. LEISTER, 2002 Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. *Mol. Biol. Evol.* **19**: 76–84.
- SCHLUETER, J. A., B. E. SCHEFFLER, S. JACKSON and R. C. SHOEMAKER, 2008 Fractionation of synteny in a genomic region containing tandemly duplicated genes across *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. *J. Hered.* **99**: 390–395.
- SMITH, S. M., A. J. PRYOR and S. H. HULBERT, 2004 Allelic and haplotypic diversity at the *Rp1* rust resistance locus of maize. *Genetics* **167**: 1939–1947.
- SONG, R., V. LLACA and J. MESSING, 2002 Mosaic organization of orthologous sequences in grass genomes. *Genome Res.* **12**: 1549–1555.
- SONG, W. Y., L. Y. PI, G. L. WANG, J. GARDNER, T. HOLSTEN *et al.*, 1997 Evolution of the rice *Xa21* disease resistance gene family. *Plant Cell* **9**: 1279.
- SUGITA, T., K. YAMAGUCHI, T. KINOSHITA, K. YUJI, Y. SUGIMURA *et al.*, 2006 QTL analysis for resistance to phytophthora blight (*Phytophthora capsici* Leon.) using an intraspecific doubled-haploid population of *Capsicum annuum*. *Breed. Sci.* **56**: 137–145.
- TAI, T., and B. J. STASKAWICZ, 2000 Construction of a yeast artificial chromosome library of pepper (*Capsicum annuum* L.) and identification of clones from the *Bs2* resistance locus. *Theor. Appl. Genet.* **100**: 112–117.
- TAI, T., D. DAHLBECK, R. E. STALL, J. PELEMAN and B. J. STASKAWICZ, 1999a High-resolution genetic and physical mapping of the region containing the *Bs2* resistance gene of pepper. *Theor. Appl. Genet.* **99**: 1201–1206.
- TAI, T. H., D. DAHLBECK, E. T. CLARK, P. GAJIWALA, R. PASION *et al.*, 1999b Expression of the *Bs2* pepper gene confers resistance to bacterial spot disease in tomato. *Proc. Natl. Acad. Sci. USA* **96**: 14153–14158.
- TAKKEN, F. L. W., M. ALBRECHT and W. I. L. TAMELING, 2006 Resistance proteins: molecular switches of plant defence. *Curr. Opin. Plant Biol.* **9**: 383–390.
- TAM, O. H., A. A. ARAVIN, P. STEIN, A. GIRARD, E. P. MURCHISON *et al.*, 2008 Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**: 534–538.
- VAN DER BIEZEN, E. A., and J. D. G. JONES, 1998 The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr. Biol.* **8**: 226–227.
- VAN DER VOSSEN, E. A. G., J. VAN DER VOORT, K. KANYUKA, A. BENDAHDANE, H. SANDBRINK *et al.*, 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.
- VAN OOIJEN, G., H. A. VAN DEN BURG, B. J. C. CORNELISSEN and F. L. W. TAKKEN, 2007 Structure and function of resistance proteins in solanaceous plants. *Annu. Rev. Phytopathol.* **45**: 43–72.
- WICKER, T., N. YAHIAOUI and B. KELLER, 2007 Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. *Plant J.* **51**: 631–641.
- WISSER, R. J., P. J. BALINT-KURTI and R. J. NELSON, 2006 The genetic architecture of disease resistance in maize: a synthesis of published studies. *Phytopathology* **96**: 120–129.
- YI, H., and E. J. RICHARDS, 2007 A cluster of disease resistance genes in *Arabidopsis* is coordinately regulated by transcriptional activation and RNA silencing. *Plant Cell* **19**: 2929–2939.
- YOSHIOKA, Y., S. MATSUMOTO, S. KOJIMA, K. OHSHIMA, N. OKADA *et al.*, 1993 Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific transfer-RNAs. *Proc. Natl. Acad. Sci. USA* **90**: 6562–6566.

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.109.101022/DC1>

The Fractionated Orthology of *Bs2* and *Rx/Gpa2* Supports Shared Synteny of Disease Resistance in the Solanaceae

**Michael Mazourek, Elizabeth T. Cirulli, Sarah M. Collier, Laurie G. Landry,
Byoung-Cheorl Kang, Edmund A. Quirin, James M. Bradeen, Peter Moffett
and Molly M. Jahn**

Copyright © 2009 by the Genetics Society of America
DOI: 10.1534/genetics.109.101022

FIGURE S1.—The YAC sequence containing the *Bs2* gene was compared against itself using a window of 50 basepairs and sequence identity cutoff of 65%. The units on the axes are kilobases. The diagram of the *Bs2* YAC from FIGURE 3A is shown along these axes in its orientation as found in GenBank.

FIGURE S2.—The contig containing *Rx*, *Gpa2* and two pseudogenes was aligned against itself. A window of 75 basepairs and a sequence identity cutoff of 65% was used for visualization. Colored dots represent the genes *Rx* (red), *Gpa2* (yellow), and pseudogenes (blue). The units for the labeled axes are kilobases.

TABLE S1**Transposon related sequences identified in YAC AY702979**

Element	Direction	From (bp)	To (bp)	Similarity	Class
Alien	Direct	2676	2959	72%	Nonautonomous DNA Transposon
Sonata1	Direct	6760	6853	74%	Nonautonomous DNA Transposon
Sonata2	Direct	8920	9121	76%	Nonautonomous DNA Transposon
Sonata1	Direct	17216	17283	84%	Nonautonomous DNA Transposon
TS	Direct	18625	18707	74%	Non-LTR Retrotransposon
Sonata2	Direct	21439	21532	73%	Nonautonomous DNA Transposon
hAT-2_SD	Direct	23821	24887	68%	DNA Transposon
Sonata3	Direct	28573	28658	74%	Nonautonomous DNA Transposon
Sonata1	Direct	28803	28881	79%	Nonautonomous DNA Transposon
GYPOT1_I	Direct	41227	44107	69%	LTR Retrotransposon (Gypsy)
TS2	Direct	49709	50349	70%	SINE retrotransposon (non-LTR)
Ogre-SD1_I	Direct	50998	51207	58%	LTR Retrotransposon (Gypsy)
Ogre-SD1_I	Direct	51335	51756	68%	-continued-
Ogre-SD1_I	Direct	51812	52118	74%	-continued-
Ogre-SD1_I	Direct	52830	53297	72%	-continued-
Ogre-SD1_I	Direct	53351	53651	69%	-continued-
Ogre-SD1_I	Direct	53905	54011	74%	-continued-
Ogre-SD1_I	Direct	55404	55752	75%	-continued
Alien	Direct	57005	57327	77%	Nonautonomous DNA Transposon
SINE_SO	Complementary	64813	64967	79%	Non-LTR Retrotransposon/SINE
Copia16-VV_I	Complementary	70059	74698	74%	LTR Retrotransposon (Copia)
Ogre-SD1_I	Complementary	78998	79968	69%	LTR Retrotransposon (Gypsy)
Ogre-SD1_I	Complementary	80690	81017	74%	-continued-
Ogre-SD1_I	Complementary	81076	81486	69%	-continued
Alien	Complementary	84392	84712	77%	Nonautonomous DNA Transposon
Sonata3	Complementary	100243	100351	68%	Nonautonomous DNA Transposon
Sonata2	Complementary	100478	100697	73%	Nonautonomous DNA Transposon
hAT-2_SD	Direct	102261	103322	68%	DNA Transposon