# Genome-Level Evolution of Resistance Genes in *Arabidopsis thaliana*

## Andrew Baumgarten,[1] Steven Cannon,[2] Russ Spangler[3] and Georgiana May[3]

*Graduate Program in Plant Biological Sciences, University of Minnesota, Saint Paul, Minnesota 55108*

## ABSTRACT

Pathogen resistance genes represent some of the most abundant and diverse gene families found within plant genomes. However, evolutionary mechanisms generating resistance gene diversity at the genome level are not well understood. We used the complete *Arabidopsis thaliana* genome sequence to show that most duplication of individual NBS-LRR sequences occurs at close physical proximity to the parent sequence and generates clusters of closely related NBS-LRR sequences. Deploying the statistical strength of phylogeographic approaches and using chromosomal location as a proxy for spatial location, we show that apparent duplication of NBS-LRR genes to ectopic chromosomal locations is largely the consequence of segmental chromosome duplication and rearrangement, rather than the independent duplication of individual sequences. Although accounting for a smaller fraction of NBS-LRR gene duplications, segmental chromosome duplication and rearrangement events have a large impact on the evolution of this multigene family. Intergenic exchange is dramatically lower between NBS-LRR sequences located in different chromosome regions as compared to exchange between sequences within the same chromosome region. Consequently, once translocated to new chromosome locations, NBS-LRR gene copies have a greater likelihood of escaping intergenic exchange and adopting new functions than do gene copies located within the same chromosomal region. We propose an evolutionary model that relates processes of genome evolution to mechanisms of evolution for the large, diverse, NBS-LRR gene family.

I N the dynamic interaction between plants and pathogens, plants counter the evolution of pathogen avirulence genes with the evolution of new resistance genes (FLOR 1956; CLAY and KOVER 1996; STAHL *et al.* 1999). Although our understanding of the molecular pathways and the genomic architecture underlying pathogen resistance traits has greatly advanced over the last decade (BAKER *et al.* 1997; LEISTER *et al.* 1998; MEYERS *et al.* 1999; GRUBE *et al.* 2000; PAN *et al.* 2000), the genome-level mechanisms generating resistance gene diversity over relatively short time scales are not well understood. To address this problem, we borrowed from phylogeographic approaches (AVISE 2000) to determine the evolutionary mechanisms by which numerous and diverse pathogen resistance genes have become dispersed across plant genomes. In this article, we focus on genes encoding proteins containing nucleotide binding site and leucine rich repeat domains (NBS-LRR) because these are the most common class of the resistance genes thus far characterized within plant genomes (reviewed in BAKER *et al.* 1997; HULBERT *et al.* 2001). Like many resistance

and defense genes, members of the NBS-LRR gene family are found widely distributed throughout plant genomes (LEISTER *et al.* 1998; MEYERS *et al.* 1999; GRUBE *et al.* 2000; WANG *et al.* 2001) but are clustered within discrete chromosomal regions. Within such clusters, diverse NBS-LRR genes often control recognition of evolutionarily diverse pests and pathogens (MEYERS *et al.* 1999; GEBHARDT and VALKONEN 2001). To account for the diversity of resistance genes, it has been suggested that resistance gene families evolve by a birth-death process rather than in a concerted fashion (MICHELMORE and MEYERS 1998). Further, several authors have proposed that sequence similarity and function of orthologous resistance genes is conserved because sequence exchange between alleles occurs more frequently than sequence exchange between paralogous gene copies (MICHELMORE and MEYERS 1998; NOËL *et al.* 1999; PARNISKE and JONES 1999).

While molecular mechanisms generating variation in NBS-LRR gene copies within chromosomal clusters are becoming better understood, controversy surrounds models of the evolution of diverse NBS-LRR genes and their observed distribution within plant genomes. In this study, we used the complete sequence of *Arabidopsis thaliana* to test two distinct models for the evolution of NBS-LRR genes on the genomic level. Phylogenetic and comparative mapping studies of resistance gene analogs within the grasses and other plant species suggest a rapid rearrangement model where frequent duplication of individual resistance genes and transposition to ec-

[1] *Corresponding author:* Department of Ecology, Evolution and Behavior, 100 Ecology, 1987 Upper Buford Circle, University of Minnesota, St. Paul, MN 55108. E-mail: baum0217@tc.umn.edu

[2] *Present address:* Department of Plant Pathology, University of Minnesota, Room 495 BorH 6030, 1991 Upper Buford Circle, St. Paul, MN 55108.

[3] *Present address:* Department of Ecology, Evolution and Behavior, University of Minnesota, 100 Ecology, 1987 Upper Buford Circle, St. Paul, MN 55108.

topic chromosomal locations scatters closely related gene copies across the genome, reminiscent of mobile genetic elements (GALLEGO *et al.* 1998; LEISTER *et al.* 1998; DUBCOVSKY *et al.* 2001; RICHLY *et al.* 2002). In contrast, comparative mapping studies within the family Solanaceae provide evidence for a conserved synteny model in which duplication of NBS-LRR genes to nonsyntenic regions is uncommon (GRUBE *et al.* 2000; PAN *et al.* 2000). The conserved synteny and rapid rearrangement models are not mutually exclusive but are used here as a heuristic tool to generate testable predictions for the distribution of paralogous NBS-LRR sequences across plant genomes. The rapid rearrangement model predicts that closely related NBS-LRR sequences, the products of recent gene duplication events, will be found across ectopic chromosomal locations. The conserved synteny model predicts that closely related NBS-LRR genes will be located near each other on the same chromosome and that relocation of NBS-LRR gene copies to new chromosome regions is the result of genome-level evolutionary processes such as polyploid events, chromosome doubling, or recombination between overlapping inversions. The genomes of many plant species such as soybean (GRANT *et al.* 2000), maize (WILSON *et al.* 1999), and *A. thaliana* (VISION *et al.* 2000) consist of several duplicated chromosomal regions.

We test the fit of data to the differing predictions presented by the conserved synteny and rapid rearrangement model by reconstructing the history of NBS-LRR gene duplication and rearrangement within the *A. thaliana* genome. Our data set and resulting gene tree are much the same as that of RICHLY *et al.* (2002), but in this article we draw on the statistical power of phylogenetic and phylogeographic approaches to examine the correlation between the phylogenetic relationship of NBS-LRR sequences and their physical location in the genome. We find that most ectopic relocation of NBS-LRR gene copies is almost entirely explained by duplication and rearrangement of the chromosome segments on which those genes reside, rather than the independent duplication of individual sequences as inferred in several reports (LEISTER *et al.* 1998; PAN *et al.* 2000; RICHLY *et al.* 2002). Further, we demonstrate for the first time the impact of physical chromosomal location on the frequency of intergenic exchange between NBS-LRR gene copies. We develop a genome-wide model for the evolution of NBS-LRR genes in *A. thaliana* by bringing together the relationship of NBS-LRR gene duplication, segmental chromosomal duplication, and levels of intergenic recombination events.

## MATERIALS AND METHODS

**NBS-LRR genealogical reconstruction:** We extracted 149 NBS-LRR gene sequences from the MIPS *A. thaliana* (Columbia) database (http://www.mips.biochem.mpg.de/proj/thal/db/index.html). The NBS-LRR sequences include both those encoding the Toll interleukin receptor motif (TIR genes) and sequences not encoding the TIR motif (non-TIR genes). Except for the region encompassing the TIR, the N-terminal region through the NBS region of NBS-LRR genes is fairly well conserved and suitable for describing evolutionary relationships among diverse sequences. We aligned the amino acid sequence of this region using an iterative hidden Markov model (HMM) algorithm found in the program HMMPRO (BALDI *et al.* 1994; http://www.netid.com). Gap characters inserted in the region of the TIR motif to achieve alignment were subsequently removed and an alignment of 425 amino acid sites resulted. The genealogy of all 149 NBS-LRR sequences was generated using the Bayesian likelihood program, Mr. Bayes (HUELSENBECK 2000), to create a 50% majority rule consensus tree of 8000 estimated maximum likelihood (ML) trees. Statistical support for the topology of the tree was evaluated using PAUP *4.0b8 (SWOFFORD 2001) to calculate the frequency of branches present among the 8000 ML trees. Branches present in <50% of these trees were collapsed to polytomies. The consensus tree (Figure 1) was midpoint rooted at the ancient division of TIR genes and non-TIR genes (MEYERS *et al.* 1999). The physical position of each NBS-LRR sequence within the *A. thaliana* genome was determined from the The Arabidopsis Information Resource website (http://www.arabidopsis.org; HUALA *et al.* 2001) and mapped onto the consensus tree as an unordered character state using MacClade 4.0 (MADDISON and MADDISON 2000). Phylogenetic clades (clades A–K, Figure 1) were delimited to include NBS-LRR sequences similar enough to allow full-length nucleotide sequence alignment [average pairwise distances within a clade were 0.57–1.57 substitutions per amino acid site; FELSENSTEIN (2001)] but such that each clade also contains sequences located on at least two different chromosomes.

**Genome location of NBS-LRR genes and of duplicated chromosome segments:** The relative frequency at which NBS-LRR genes duplicate to syntenic or to nonsyntenic genomic regions was estimated by analyzing changes in chromosome location state reconstructed at each node of the genealogy using MacClade 4.0 (MADDISON and MADDISON 2000). Genome location of NBS-LRR sequences was first defined by their chromosomal location and next, by their location within clusters of NBS-LRR genes located in close physical proximity on the chromosome. In our analyses, chromosomal clusters of NBS-LRR genes were defined by the occurrence of sequences within 2 Mb of one another (2-Mb linked clusters) and are largely the same as those evident in MEYERS *et al.* (1999; http://www.niblrrs. ucdavis.edu/At_RGenes/).

We used the previously identified breakpoints of segmentally duplicated regions within the *A. thaliana* genome to determine which NBS-LRR gene sequences were found within duplicate regions of the genome (ARABIDOPSIS GENOME INITIATIVE 2000; BLANC *et al.* 2000; GRANT *et al.* 2000; VISION *et al.* 2000). For our purposes, breakpoints of segmentally duplicated regions were defined by the position of the most telomeric and centromeric genes found within the segment. Because the breakpoints of more divergent duplicated regions are difficult to identify by the similarity searches used in the above studies, we accommodated uncertainty in breakpoints by including NBS-LRR genes falling within 0.5 Mb of estimated breakpoints. Further, we used comparative maps between *A. thaliana* and *Brassica nigra* (LAGERCRANTZ *et al.* 1998) to determine if NBS-LRR genes presently found in different regions of the *A. thaliana* genome might have once been clustered together within an ancestral genome. In these comparative maps, breakpoints were estimated using marker sequences mapped in both *A. thaliana* and *B. nigra* genomes, and as above, NBS-LRR genes were included within a chromosome segment if they fell within 0.5 Mb of the mapped sequence in the *A. thaliana* genome.

**Association test for the role of segmental duplication and rearrangement:** We employed an index of association test (IA; Brown *et al.* 1980; Agapow and Burt 1999; http://www.bio. ic.ac.uk/evolve/software/multilocus) to provide a second test of the data's fit to the conserved synteny or rapid rearrangement models. The IA is a summary statistic used to describe the extent of multilocus associations of alleles across populations. For the purpose of the IA test, we treated 2-Mb linked clusters as populations and plotted the occurrence of distantly related NBS-LRR sequences representing different phylogenetic clades (Figure 1, A–K) within these clusters and across chromosomal locations. Sequences representing the different phylogenetic clades were recorded as a single data point within the 2-Mb linked clusters. This approach also limited the effect of variation in numbers of NBS-LRR genes within the same 2-Mb linked cluster on IA test significance values. The statistical significance of observed IA values was obtained by comparing the observed IA values with a null distribution of association values derived from 1000 permutations of NBS-LRR sequence locations across 2-Mb linked clusters. The permutations serve to create expectations for association of gene copies by chance. Sensitivity analyses were conducted by defining smaller clusters of NBS-LRR genes on the basis of sequences located within 0.5 Mb (0.5-Mb linked clusters) of one another and modifying the number of sequences included within a phylogenetic clade.

**Intergenic sequence exchange:** We examined nucleotide alignments of NBS-LRR genes for evidence of intergenic sequence exchange. The full-length DNA sequences of NBS-LRR genes, both with and without introns, were aligned using CLUSTALW (Thompson *et al.* 1994). If lengths of aligned sequences varied greatly such that large gaps were introduced into the alignment, the shortest sequence was removed from the data set and the sequences were realigned. All nucleotide sites in the resulting alignments were examined with GENE-CONV (Sawyer 2000), which searches for tracts of shared, phylogenetically informative sites between three or more aligned sequences and thus detects sequence exchange events resulting from reciprocal exchange, unequal crossover, or gene conversion events. Only intergenic exchange events maintaining global Bonferroni corrected *P* values ≤0.05 were considered significant. Sequence mismatch penalties were varied at 0, 1, and 2 to provide the range of detected intergenic exchange events shown in Tables 1 and 2.

## RESULTS

**Genome location of NBS-LRR gene duplication events:** The relative frequency at which NBS-LRR gene duplication occurs to syntenic or to nonsyntenic regions of the genome was estimated by counting the changes of chromosome location state reconstructed at each node of the genealogy and dividing by the total number of nodes involved (Figure 1). Since the *A. thaliana* sequence represents a single haploid genome, each node within the NBS-LRR genealogy represents a gene duplication event. Thus, we counted each character state change in chromosomal location as duplication to a new location. As a conservative test, we counted only those nodes at which changes of location state or no changes of location state were unambiguously assigned by using MacClade 4.0 (Maddison and Maddison 2000). Implicitly, our approach uses the branches with >50% representation on

the consensus tree as an estimate of support for character state changes at each node. An alternative approach using the 8000 ML trees to generate confidence intervals on the inferences of chromosomal location state changes is worthy of development but beyond the scope of our work.

First, we conducted analyses to estimate the fraction of duplication events moving gene copies to the same or different chromosomes. We found that 81.1% of NBS-LRR duplications (90 of 111 reconstructed character states) resulted in gene copies located on the same chromosome as the most closely related paralogous sequence. To further refine the analysis, duplication within and between 2-Mb linked clusters was examined by assigning location within 2-Mb linked clusters as a character state of each sequence and counting changes in character state reconstructed at each node with Mac-Clade 4.0 (Maddison and Maddison 2000). Even using more restricted chromosomal regions, we found that 79.8% of duplication events (71 of 89 reconstructed character states) occur close to one another and within the same 2-Mb linked cluster. These data demonstrate a strong correlation between a close phylogenetic relationship and physical proximity of NBS-LRR gene copies, an observation expected under the model of conserved synteny.

**Role of segmental duplication and rearrangement:** The majority of observations from our genealogical analyses support the conserved synteny model but do not fully exclude the rapid rearrangement model because some closely related NBS-LRR genes were also found in different regions of the *A. thaliana* genome (Figure 1). Of the 89 duplication events resolved in the phylogenetic analyses, 18 apparently involved duplication to a new, ectopic chromosomal location. Infrequent duplication of individual sequences to a new chromosome location coupled with more frequent duplication within a region could result in much the same distribution of NBS-LRR sequences as we present above. However, the *A. thaliana* genome has endured multiple chromosome duplication and rearrangement events, suggesting that segmental duplication and rearrangement might have distributed NBS-LRR gene copies across different chromosome regions (Lagercrantz 1998; Koch *et al.* 1999; Arabidopsis Genome Initiative 2000; Blanc *et al.* 2000; Grant *et al.* 2000; Vision *et al.* 2000).

Using previously identified breakpoints of segmentally duplicated regions within the *A. thaliana* genome (Lagercrantz 1998; Arabidopsis Genome Initiative 2000; Blanc *et al.* 2000; Grant *et al.* 2000; Vision *et al.* 2000), we determined that 15 of the 18 apparent ectopic duplications also fell within segmentally duplicated regions of the genome. Only 3 NBS-LRR gene duplication events of 89 total events are left unexplained by the conserved synteny model and could possibly reflect independent duplication and rearrangement of NBS-LRR genes as predicted by the rapid rearrangement model.

Thus, we show that almost all NBS-LRR duplication events involving apparent relocation to a new chromosomal region are explained by segmental duplication and rearrangement. Taking the results of phylogenetic analyses and segmental duplication together, data strongly support the conserved synteny model because most
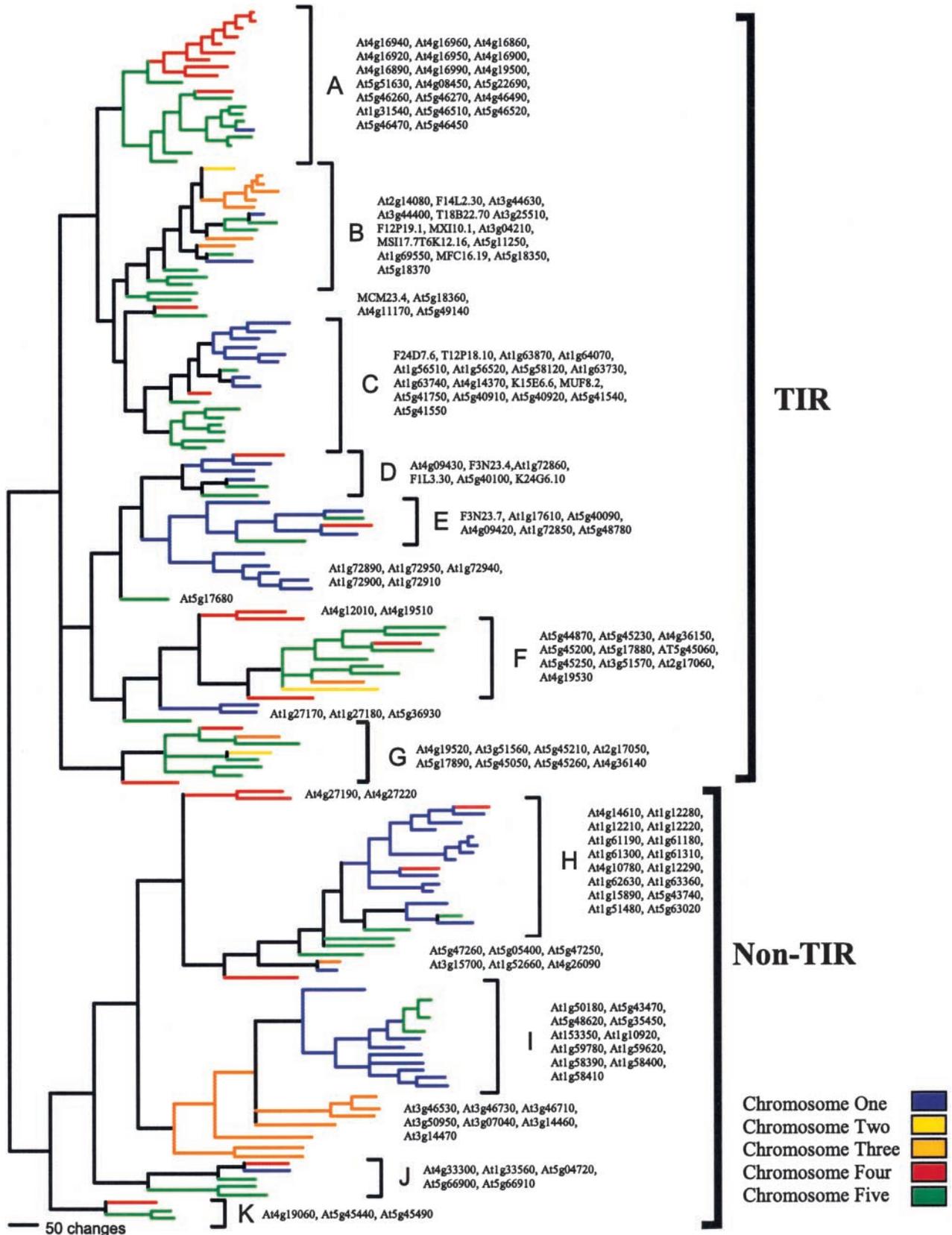
NBS-LRR duplication events (96.6% or 86 of 89 total) occur within local chromosomal regions.

**Association test for the role of segmental duplication and rearrangement:** Our estimates for the contribution of segmental duplication and rearrangement to the genomic distribution of NBS-LRR sequences are based on the reconstruction of recent duplication events on the gene tree. These inferences will be limited by gene loss, intergenic sequence exchange, and ambiguity in assigning character state changes. As a second test of the conserved synteny and rapid rearrangement models, independent of tree topology, we used an IA test (see materials and methods) to examine the correlated occurrence of NBS-LRR gene copies across the genome. The conserved synteny model predicts that duplication and rearrangement of NBS-LRR gene sequences to new chromosomal locations is accomplished by the duplication and rearrangement of entire chromosomal segments on which those NBS-LRR genes reside. If this is true, we will observe the co-occurrence of gene copies across several chromosome locations. For example, if genes A and B are represented together on one chromosomal segment, their gene copies, A' and B', will be represented together on another, duplicate chromosomal segment. In contrast, the rapid rearrangement model predicts that individual NBS-LRR sequences duplicate to new chromosome locations independently of other gene sequences. As a result, genomic locations of distantly related gene copies will not be correlated with each other or with segmental duplications. Note, however, that the ectopic duplication of a single NBS-LRR sequence, followed by rapid duplication within a chromosome location, could generate significant IA values if these recent duplicates were included in the analysis. To avoid confounding recent gene duplication with correlated dispersal, we examined only the association of more distantly related NBS-LRR gene copies. We plotted the genomic location of more distantly related NBS-LRR sequences representing different phylogenetic clades (Figure 1, A–K) onto their sequence position within the *A. thaliana* genome, with each occurrence within 2-Mb linked clusters represented as a single point (Figure 2). From inspection of Figure 2, it is evident that sequences representing different clades co-occur within more than one distinct genomic region. For example, sequences representative of clades D and E are found together in two separate regions of chromosome 1, in one region of chromosome 4, and in two separate regions of chromosome 5. Similarly, sequences representative of clades

F and G are found together in regions of chromosomes 2, 3, 4, and 5. The prevalent associations of distantly related NBS-LRR gene copies across 2-Mb linked clusters led to significant IA tests ($P \leq 0.05$) and significance levels were not sensitive to varying the size of linked clusters (2- or 0.5-Mb linked clusters) or varying the delimitation of phylogenetic clades on Figure 1. The correlated occurrence of divergent NBS-LRR sequences across different genomic regions strongly supports the hypothesis that segmental chromosome duplication and rearrangement have distributed members of the NBS-LRR gene family throughout the *A. thaliana* genome and not the independent duplication and rearrangement of individual genes.

**Intergenic sequence exchange:** We next used GENE-CONV (Sawyer 2000) to investigate the impact of physical location and sequence similarity on the frequency of intergenic exchange. The global GENECONV analysis looks for shared phylogenetically informative character states between two sequences as evidence of recombination and results are insensitive to varying substitution models. GENECONV may be considered conservative in its estimate of recombination rates but is useful in detecting intergenic recombination between divergent sequences (Posada and Crandall 2001). As above, closely related NBS-LRR genes were defined as belonging to the same phylogenetic clade (Figure 1, A–K) and physical location was defined by 2-Mb linked clusters. We first examined the frequency of exchange between distantly related sequences and could not detect significant sequence exchange events even when we restricted our attention to sequences lying in close chromosomal proximity and used relaxed sequence mismatch parameters.

Subsequent tests for sequence exchange were conducted by comparing more closely related NBS-LRR sequences. The effect of genome location was determined by comparing the frequencies of intergenic exchange events detected for sequences located within and between different 2-Mb linked clusters. Among sequences located within the same 2-Mb linked cluster, many significant intergenic sequence exchange events ($P \leq 0.05$) were found regardless of parameters defined. The inclusion of intron sequences provided a larger number of polymorphic sites and increased the number and length of sequence exchange events detected (Table 1). Between closely related NBS-LRR sequences located in different 2-Mb linked clusters, a number of significant sequence exchange events ($P \leq 0.05$) were

Figure 1.—Phylogeny of *A. thaliana* NBS-LRR gene sequences. Phylogeny shown represents the 50% majority-rule consensus tree of 8000 estimated ML trees and is rooted at the ancient divergence of sequences containing the TIR motif and those sequences lacking a TIR motif (non-TIR). Branches are colored to show the chromosome location reconstructed for that branch by MacClade 4.0 (Maddison and Maddison 2000). Black branches represent chromosome locations that could not be unambiguously reconstructed by MacClade 4.0. Phylogenetic clades (A–K) were delimited to define more closely related sequences within a clade and more distantly related sequences falling in different clades.
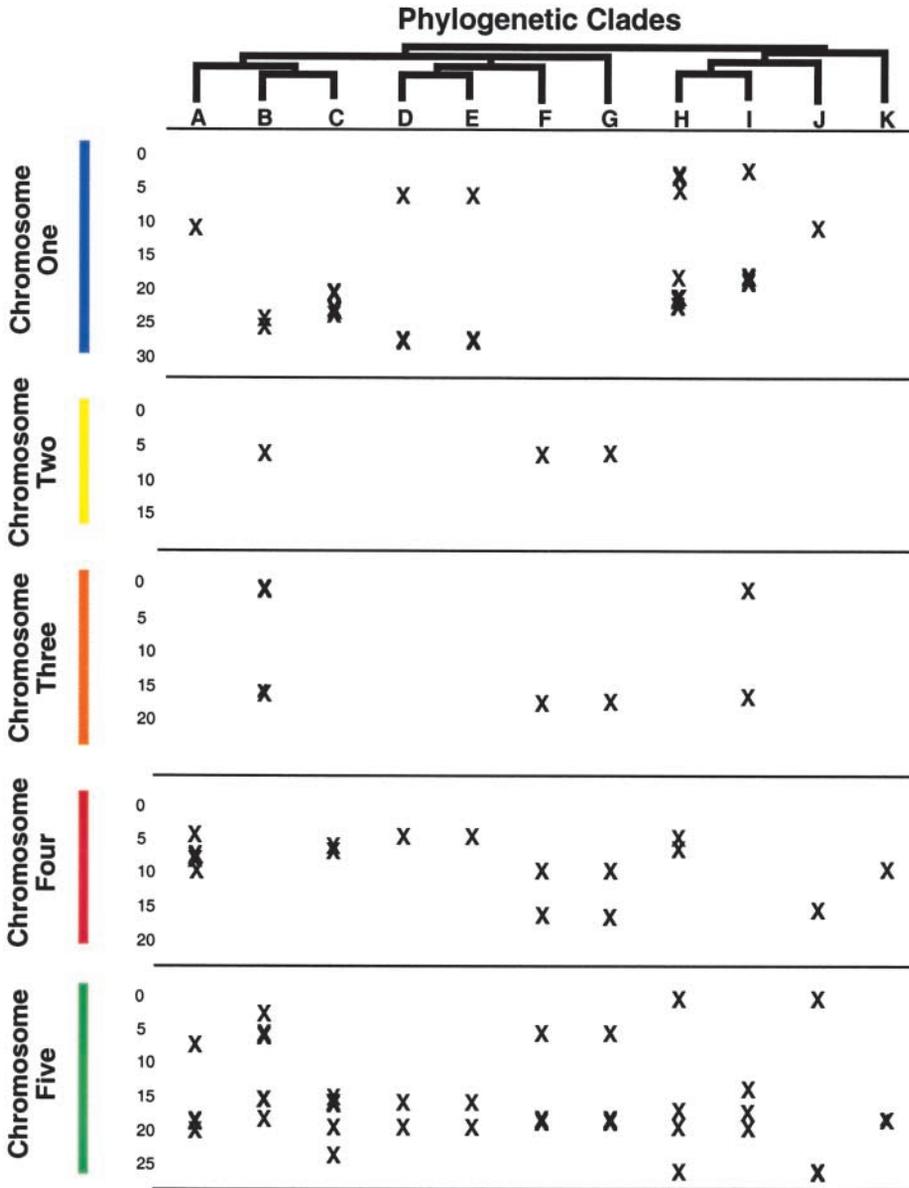
FIGURE 2.—Genomic position of sequences representing different phylogenetic clades within the NBS-LRR phylogeny. Sequence membership in phylogenetic clades is plotted on the *x*-axis and the chromosome location and position of the sequence within the chromosome is plotted on the *y*-axis. Position within chromosomes is in megabases from the telomere of the top arm of each chromosome. An X plots the occurrence of a sequence representing a phylogenetic clade at the corresponding chromosomal location. The relationship of clades is illustrated by a skeleton phylogeny at the top of the diagram and has the same topology as Figure 1. Phylogenetic clades are delimited as in Figure 1, except that additional sequences were included in clades C, H, and I to conduct the index of association tests.

also detected and again, inclusion of intron sequences allowed more events to be detected (Table 2). Considerable variation in the frequency and length of recombination events is evident along chromosomes (Table 1) and across different clades (Table 2). Comparing the two sets of results, the average frequency of exchange detected among sequences residing within the same 2-Mb linked cluster was ~20-fold higher and involved longer exchange tracts (Table 1) than exchanges between sequences residing in different 2-Mb linked clusters (Table 2). Most intergenic exchange between NBS-LRR sequences on different chromosomes involved sequences located within duplicated genomic regions, suggesting that some of these exchanges might have occurred before segmental duplication or rearrangement events. Mutations and gene loss following a duplication event would obscure the history of sequence exchange and lower the probability of detecting ancestral

exchanges. Altogether, our results show that frequency of exchange detected between NBS-LRR sequences from different chromosomal regions is detectable but much lower than the frequency of intergenic exchange observed within chromosomal regions.

## DISCUSSION

Using a novel phylogeographic approach treating chromosomal regions as geographic populations, we provide strong support for a conserved synteny model of NBS-LRR gene evolution in which most NBS-LRR gene duplications occur within restricted chromosomal regions.

Previously, the observation that closely related resistance gene sequences are located across widespread chromosomal locations has been interpreted as support for a model of rapid rearrangement of duplicated resis-

**TABLE 1**

**Number and length of significant ($P < 0.05$) intergenic exchange events detected between NBS-LRR sequences located within chromosome clusters and grouped within the same phylogenetic clade (Figure 1, A–K)**

| Cluster location[a] | Phylogenetic clade | Exon sequence | | Exon and intron sequence | |
|---|---|---|---|---|---|
| | | Exchanges per pairwise comparison | Exchange lengths (bp) | Exchanges per pairwise comparison | Exchange lengths (bp) |
| Chromosome 1: 4.2 Mb | H | 0.10 | 52–80 | 0.10 | 52–106 |
| Chromosome 1: 19.5 Mb | C | 0.10–0.19 | 68–411 | 0.05–0.33 | 37–334 |
| Chromosome 1: 21.1 Mb | H | 0.20–0.38 | 7–496 | 0.42–1.62 | 43–2623 |
| Chromosome 1: 22.3 Mb | I | 0.07–0.13 | 43–277 | 0.4–0.73 | 24–651 |
| Chromosome 1: 27.1 Mb | E | 0.05–0.10 | 23–308 | 0.05–0.10 | 23–302 |
| Chromosome 3: 17.0 Mb | B | 1.10–1.80 | 17–1232 | 1.70–3.50 | 26–1917 |
| Chromosome 3: 17.6 Mb | I | 0.17 | 16–25 | 0.00 | NA |
| Chromosome 4: 7.1 Mb | A | 1.50–1.90 | 20–1522 | 1.53–3.86 | 23–2969 |
| Chromosome 5: 6.1 Mb | B | 0.00–0.01 | 83 | 0.00–0.10 | 83 |
| Chromosome 5: 16.7 Mb | C | 0.14–0.19 | 53–831 | 0.23–0.47 | 31–1160 |
| Chromosome 5: 18.2 Mb | F | 0.00 | NA | 0.00 | NA |
| Chromosome 5: 18.3 Mb | G | 0.00 | NA | 0.00 | NA |
| Chromosome 5: 18.7 Mb | A | 0.35–0.46 | 23–2081 | 1.07–1.25 | 21–2697 |
| Average | | 0.41 | 252 | 0.77 | 288 |
| Standard deviation | | 0.59 | 401 | 1.06 | 500 |

Reported number and exchange lengths represent the range of values for intergenic exchange events detected using sequence mismatch penalties of 0, 1, and 2. Values of intergenic exchange represent the number of exchange events detected per pairwise comparison. NA, not applicable.

[a] Midpoint position of cluster on the chromosome in megabases from the telomere.

tance genes (Gallego *et al.* 1998; Leister *et al.* 1998; Pan *et al.* 2000; Dubcovsky *et al.* 2001; Richly *et al.* 2002), as might our evidence for sequence exchange between genes in ectopic locations. In contrast, we derive support for the conserved synteny model by showing that duplication and rearrangement of chromosomal segments, rather than duplication of individual gene sequences, drives the genomic distribution and apparent ectopic duplication of NBS-LRR gene sequences.

Our finding that most recent gene duplication events have occurred within a local chromosomal region is supported by results from comparative mapping and plant genome projects (see Hulbert *et al.* 2001), which show that several closely related gene copies are clustered together within local chromosomal regions. Such resistance gene clusters may themselves be located within larger chromosomal regions that carry diverse genes encoding resistance to several different pathogens (Gebhardt and Valkonen 2001). Here, we build on previous models for the evolutionary dynamics within NBS-LRR gene clusters (Hulbert 1997; Parniske *et al.* 1997; Michelmore and Meyers 1998; Noël *et al.* 1999; Parniske and Jones 1999; Chin *et al.* 2001; Dodds *et al.* 2001; Sun *et al.* 2001) to develop a genome-level model for the evolutionary dynamics of NBS-LRR genes.

We took advantage of detailed information for segmental duplication and rearrangement in the evolution of the *A. thaliana* genome (Lagercrantz 1998; Arabidopsis Genome Initiative 2000; Blanc *et al.* 2000; Grant *et al.* 2000; Vision *et al.* 2000) to correlate segmental duplication and rearrangement events with gene duplication events. Tracking changes in chromosomal

**Number and length of significant ($P < 0.05$) intergenic exchange events detected between NBS-LRR sequences located on different chromosomes and grouped within the same phylogenetic clade (Figure 1, A–K)**

| Phylogenetic clade | Exon sequence | | Exon and intron sequence | |
|---|---|---|---|---|
| | Exchanges per pairwise comparison | Exchange length (bp) | Exchanges per pairwise comparison | Exchange length (bp) |
| A | 0.01–0.07 | 593 | 0.00–0.04 | 17–1085 |
| B | 0.00–0.02 | 36–93 | 0.00–0.02 | 72 |
| C | 0.00–0.05 | 288–440 | 0.00–0.07 | 78–1012 |
| D | 0.00–0.05 | 42–132 | 0.00–0.05 | 33–196 |
| E | 0.00 | 0 | 0.00–0.19 | 14–256 |
| F | 0.00–0.04 | 38–190 | 0.00–0.04 | 52–146 |
| G | 0.03–0.05 | 46–104 | 0.06–0.08 | 20–104 |
| H | 0.00 | 0 | 0.00 | 123–1262 |
| I | 0.00 | 0 | 0.00–0.03 | 128–230 |
| J | 0.00 | 0 | 0.00 | 0 |
| K | 0.00 | 0 | 0.00 | 0 |
| | | Average | | |
| | 0.02 | 152 | 0.03 | 222 |
| | | Standard deviation | | |
| | 0.03 | 156 | 0.05 | 257 |

Reported number and exchange lengths represent range of values for intergenic exchange events detected using sequence mismatch penalties of 0, 1, and 2. Values of intergenic exchange represent the number of exchange events detected per pairwise comparison.

location associated with gene duplication events on the genealogy, we demonstrate that ~80% of duplication events occur within relatively restricted chromosomal regions while the remaining 20% involve duplication to new chromosomal locations.

Our estimate for the level of local duplication (80%) is larger than that of RICHLY *et al.* (2002) who estimated that 50% of gene copies remain in local chromosomal regions. Differences in these estimates arise from two sources. RICHLY *et al.* (2002) defined clusters of NBS-LRR sequences as genes occurring within eight open reading frames of each other, regions that may be smaller than the physically delimited distances (2- or 0.5-Mb linked clusters) used in this study. More importantly, the pairwise approach of RICHLY *et al.* (2002) considers changes in chromosome location states only at terminal branches of the genealogy. Using character state reconstruction accounts for the inherent correlation of events within a genealogy and thus avoids overestimation of character state changes. Both the results presented here and those of RICHLY *et al.* (2002) estimate that the majority of NBS-LRR duplication events occur locally but the studies differ in their attribution of evolutionary mechanisms for genome-wide dispersal of gene copies. The most striking results were obtained when we included additional information concerning the evolutionary history of segmental duplication and rearrangement within the *A. thaliana* genome (LAGERCRANTZ 1998; GRANT *et al.* 2000; VISION *et al.* 2000). Without incorporation of genome-level processes, each NBS-LRR gene copy

that lands in an ectopic location will be counted as an independent and rapid rearrangement event (LEISTER *et al.* 1998; RICHLY *et al.* 2002). With genome-level data, we find that most all NBS-LRR gene duplications involving ectopic relocation (16% of the total number of duplications) are explained by duplication of the chromosomal segments on which they reside. The conserved synteny model accounts for the genome-wide distribution of 96% of the total NBS-LRR gene duplications.

The conclusions resulting from phylogenetic reconstruction were further supported by results for the co-occurrence of distantly related genes within more than one distinct genomic region. We reasoned that if segmental duplication and rearrangement drive the genomic distribution of NBS-LRR genes, then all genes located on these segments will be duplicated together. As a result, gene copies will occur together in different genome locations more often than expected under independent duplication and rearrangement of individual genes. We treated NBS-LRR gene clusters as populations across the genome and examined the correlated occurrence of NBS-LRR gene copies in these populations using an index of association test. The index of association provides a robust statistical test of whether observed association of sequences in different chromosomal regions deviates significantly from random associations expected under the rapid rearrangement model. The results demonstrate significant associations of genes representing different phylogenetic clades across several genomic locations. The impact of intergenic exchange on

the IA results is negligible because measurable levels of intergenic exchange do not occur between divergent sequences representative of different phylogenetic clades (see below). Together, the results of phylogenetic and association tests demonstrate that segmental duplication and rearrangement are the most important mechanisms in distributing NBS-LRR gene family members across the genome.

**Intergenic exchange between NBS-LRR genes:** Finding that the genomic distribution of NBS-LRR sequences is dependent on segmental duplication and rearrangement has profound implications for mechanisms of molecular evolution in this gene family. In this study, we examined the relationship of sequence similarity and genomic position to the frequency of intergenic exchange between NBS-LRR gene copies. Evidence of numerous sequence exchange events was detected between closely related NBS-LRR genes, confirming the results of CHIN *et al.* (2001) and SUN *et al.* (2001), while no intergenic exchange events were detected between NBS-LRR sequences representative of different phylogenetic clades, even when those sequences were in proximity to one another. Similarly, the recent study of positive selection on NBS-LRR genes in *A. thaliana* (MONDRAGÓN-PALOMINO *et al.* 2002) detected little to no recombination between more distantly related NBS-LRR genes located in different chromosomal regions. Interestingly, the position of recombination events we detected within NBS-LRR coding sequences did not correlate with the codon sites shown by MONDRAGÓN-PALOMINO *et al.* (2002) to be under positive selection, indicating that convergent evolution did not significantly inflate our estimates of recombination events between NBS-LRR genes.

In our study, the number and length of detected intergenic exchange events varied among chromosomal clusters and phylogenetic clades. Resolution of major clades within the genealogy was not impeded by intergenic exchange but exchange rates are high enough to contribute to homoplasy and phylogenetic grouping of sequences within clades (WIUF and HEIN 1999, 2000; SCHIERUP and HEIN 2000), similar to the results of DODDS *et al.* (2001) for the *N* locus of flax. However, as previously shown for the major histocompatibility complex (MHC; TAKAHATA and SATTA 1998; MARTINSOHN *et al.* 1999), we infer that intergenic exchange rates have not been high enough to cause the concerted evolution of NBS-LRR sequences since phylogenetic relationships are resolved and polytomies are not common within clades.

While several studies have examined intergenic exchange between NBS-LRR gene copies within the same chromosome region (HULBERT 1997; MICHELMORE and MEYERS 1998; CHIN *et al.* 2001; SUN *et al.* 2001), few studies have compared rates of sequence exchange within and between different chromosome regions (see PARNISKE and JONES 1999). Examining evolution within gene clusters, birth-death models used to describe NBS-LRR (MICHELMORE and MEYERS 1998), and MHC gene evo-

lution (TAKAHATA and SATTA 1998; MARTINSOHN *et al.* 1999) imply that intergenic exchange is limited and that mutation, interallelic sequence exchange, and divergent selection drive the diversification of gene copies within physical clusters. While the complete genome sequence of Arabidopsis does not allow evaluation of interallelic exchange, we did detect significant intergenic exchange between paralogous NBS-LRR genes within local chromosomal regions and our conclusions contrast somewhat with the MICHELMORE and MEYERS (1998) model. Finding intergenic exchange is not itself surprising, since the rates of intergenic sequence exchange between duplicate genes found in close physical proximity have been estimated well above the neutral mutation rate in plants (HULBERT 1997; PARNISKE *et al.* 1997; JELESKO *et al.* 1999; CHIN *et al.* 2001; DODDS *et al.* 2001; SUN *et al.* 2001). Sequence exchange rates and their effects on NBS-LRR gene function vary widely across different plant systems (HULBERT 1997; PARNISKE *et al.* 1997; CHIN *et al.* 2001; DODDS *et al.* 2001; SUN *et al.* 2001) and different chromosomal regions (this study) but could contribute to gene structural and functional conservation among members of a physical cluster.

Given our observation of frequent intergenic exchange, the divergence of duplicate NBS-LRR genes within clusters must be rapid to escape relentless sequence homogenization by intergenic exchange (WALSH 1985, 1987). Divergence between paralogous gene copies has been proposed to occur by relaxation of selection on gene copies following duplication (LYNCH and CONERY 2000) and interallelic recombination at rates greater than intergenic recombination (MICHELMORE and MEYERS 1998). Regardless of mechanisms generating sequence diversity, strong diversifying selection must push the evolution of paralogous gene copies at a rate greater than the rate of conversion or accumulation of mutations to pseudogenes for sequences to escape homogenization by intergenic exchange (WALSH 1985, 1995). Our results do not directly address selection levels but we note that divergent NBS-LRR gene sequences are found in close physical proximity and demonstrate that these experience few exchange events. For example, sequences from clades D and E are always found within close physical proximity. An extreme example is found on the bottom arm of chromosome 5, where sequences representing all clades except J are found together in a "supercluster" (RICHLY *et al.* 2002). These dissimilar and yet clustered genes are likely records of strong diversifying selection (MEYERS *et al.* 1999; MONDRAGÓN-PALOMINO *et al.* 2002).

**Effect of segmental duplication on the rates of intergenic exchange:** Segmental duplication and rearrangement account for a small proportion of NBS-LRR gene duplications but put gene copies into a new evolutionary context, out of the reach of intergenic exchange. We present the first empirical evidence validating models of OHTA and DOVER (1983), which predicted lower lev-

els of intergenic exchange rates between gene copies located on different chromosomes, as we show that intergenic sequence exchange is much more limited among duplicate NBS-LRR sequences located in different chromosomal regions than among sequences located within the same region. The limited intergenic exchange that we inferred between NBS-LRR genes in different genomic regions could result from intergenic exchange that occurred before segmental duplication or actual intergenic exchange between NBS-LRR genes in different genomic locations. Nevertheless, as Ohta and Dover realized, distributing duplicate sequences to different chromosomal regions reduces the introduction of new sequence into a cluster of gene copies by intergenic exchange. With lower levels of sequence exchange between gene clusters, divergence between sequences in different clusters will accelerate relative to variation within clusters. Given the frequencies of exchange that we estimate within and between chromosomal clusters, Ohta and Dover's (1983) model predicts that the probability of identity between gene copies located in different genomic regions will be at least fourfold lower than identity among sequences within clusters. Segmental duplication and rearrangement will allow duplicate sequences to escape homogenization of intergenic exchange under selection levels lower than would be necessary for escape from homogenization within a cluster (Walsh 1987).

**A model for the genomic evolution of the NBS-LRR gene family:** We present a genome-level model in which the evolutionary trajectory of NBS-LRR genes is primarily governed by two mechanisms. First, most of the dynamic variation in NBS-LRR gene copy number occurs within local chromosomal regions. New NBS-LRR genes can arise and be lost through unequal crossing over, conversion, and an accumulation of mutations leading to either a pseudogene or a new function (Walsh 1995; Michelmore and Meyers 1998; Lynch and Force 2000). Infrequently, new resistance specificities evolve and are maintained through time by balancing selection (Song *et al.* 1997; Stahl *et al.* 1999; Stahl and Bishop 2000) against the grind of homogenization and gene loss. Second, although accounting for a smaller fraction of gene duplication events, segmental duplication will have an impact on NBS-LRR gene family diversification not previously recognized. NBS-LRR gene copies in new chromosomal locations will more often escape intergenic exchange and adopt new functions under levels of selection lower than required for gene copies within a cluster. Further, segmental duplication could allow the preservation of many alleles that would otherwise be maintained at a single NBS-LRR locus (Otto and Young 2002). The distribution of NBS-LRR sequences across the *A. thaliana* genome is the result of a dynamic history of gene duplication within chromosomal regions coupled with segmental duplication and rearrangement of larger chromosomal regions. The genome-level model presented here describes the relationship between genomic architecture and evolutionary processes in an important gene family and thus provides a model to be tested in other systems and compared to the findings for *A. thaliana.*

## LITERATURE CITED

Agapow, P. M., and A. Burt, 1999   *MultiLocus v1.1.* Imperial College, Ascot, UK.

Arabidopsis Genome Initiative, 2000   Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature **408:** 796–815.

Avise, J., 2000   *Phylogeography: The History and Formation of Species.* Harvard University Press, Cambridge, MA.

Baker, B., P. Zambryski, B. Statskawicz and S. P. Dinesh-Kumar, 1997   Signaling in plant-microbe interactions. Science **276:** 726–733.

Baldi, P., Y. Chauvin, T. Hunkapiller and T. A. McClure, 1994   Hidden Markov models of biological primary sequence information. Proc. Natl. Acad. Sci. USA **91**(2): 1059–1063.

Blanc, P., A. Barakat, R. Guyot, R. Cooke and M. Delseny, 2000   Extensive duplication and reshuffling in the *Arabidopsis* genome. Plant Cell **12:** 1093–1101.

Brown, A. H. D., M. W. Feldman and E. Nevo, 1980   Multi-locus population structure. Genetics **96:** 523–536.

Chin, D. B., R. Arroyo-Garcia, O. Ochoa, R. V. Kesseli, D. O. Lavelle *et al.*, 2001   Recombination and spontaneous mutation at the major cluster of resistance genes in lettuce (*Lactuca sativa*). Genetics **157:** 831–849.

Clay, K., and P. A. Kover, 1996   The red queen hypothesis and plant/pathogen interactions. Annu. Rev. Phytopathol. **34:** 29–50.

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

Dubcovsky, J., W. Ramakrishna, P. J. SanMiguel, L. Yan, B. A. Shiloff *et al.*, 2001   Comparative sequence analysis of collinear barley and rice bacterial artificial chromosomes. Plant Physiol. **125:** 1342–1353.

Felsenstein, J., 2001   *PHYLIP (Phylogeny Inference Package) Version 3.5c.* University of Washington, Seattle.

Flor, H. H., 1956   Complementary genetic systems in flax and flax rust. Adv. Genet. **8:** 29–54.

Gallego, F., C. Feuillet, M. Messmer, A. Penger, A. Graner *et al.*, 1998   Comparative mapping of the two wheat leaf rust resistance loci *Lr1* and *Lr10* in rice and barley. Genome **41:** 328–336.

Gebhardt, C., and J. P. T. Valkonen, 2001   Organization of genes controlling disease resistance in the potato genome. Annu. Rev. Phytopathol. **39:** 79–102.

Grant, D., P. Cregan and R. C. Shoemaker, 2000   Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proc. Natl. Acad. Sci. USA **97:** 4168–4173.

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

Huala, E., A. W. Dickerman, M. Garcia-Hernandez, W. Danforth, L. Reiser *et al.*, 2001   The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. **29:** 102–105.

Huelsenbeck, J. P., 2000   *Mr. Bayes: Bayesian Inference of Phylogeny.* University of Rochester, Rochester, NY (http://morphbank.ebc.uu.se/mrbayes/).

HULBERT, S., 1997   Structure and evolution of the *rp1* complex conferring rust resistance in maize. Annu. Rev. Phytopathol. **35:** 293–310.

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.

JELESKO, J. G., R. HARPER, M. FURUYA and W. GRUISSEM, 1999   Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **96:** 10302–10307.

KOCH, M., J. BISHOP and T. MITCHELL-OLDS, 1999   Molecular systematics and evolution of *Arabidopsis* and *Arabis*. Plant Biol. **1:** 529–537.

LAGERCRANTZ, U., 1998   Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics **150:** 1217–1228.

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.

LYNCH, E., and A. FORCE, 2000   The probability of duplicate gene preservation by subfunctionalization. Genetics **154:** 459–473.

LYNCH, M., and J. S. CONERY, 2000   The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1155.

MADDISON, D. R., and W. P. MADDISON, 2000   *MacClade 4.0: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, MA.

MARTINSOHN, J. T., A. B. SOUSA, L. A. GUETHLEIN and J. C. HOWARD, 1999   The gene conversion hypothesis of MHC evolution: a review. Immunogenetics **50:** 168–200.

MEYERS, B. C., A. W. DICKERMAN, R. W. MICHELMORE, S. SIVARAMAKRISHNAN, B. W. SOBRAL *et al.*, 1999   Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Mol. Plant-Microbe Interact. **20:** 317–332.

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.

MONDRAGÓN-PALOMINO, M., B. MEYERS, R. W. MICHELMORE and B. S. GAUT, 2002   Patterns of positive selection in the complete NBS-LRR gene family in *Arabidopsis thaliana*. Genet. Res. **12** (1): 1305–1315.

NOËL, L., T. L. MOORES, E. A. VAN DER BIEZEN, M. PARNISKE, M. J. DANIELS *et al.*, 1999   Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. Plant Cell **11:** 2099–2111.

OHTA, T., and G. DOVER, 1983   Population genetics of multigene families that are dispersed into two or more chromosomes. Proc. Natl. Acad. Sci. USA **80** (13): 4079–4083.

OTTO, S., and P. YOUNG, 2002   The evolution of gene duplicates. Adv. Genet. **46:** 451–483.

PAN, Q., Y.-S. LIU, O. BUDAI-HADRIAN, M. SELA, L. CARMEL-GOREN *et al.*, 2000   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.

PARNISKE, M., and J. D. G. JONES, 1999   Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proc. Natl. Acad. Sci. USA **96:** 5850–5855.

PARNISKE, M., K. E. HAMMOND-KOSACK, C. GOLSTEIN, C. M. THOMAS, D. A. JONES *et al.*, 1997   Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell **91:** 821–832.

POSADA, D., and K. A. CRANDALL, 2001   Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA **98:** 13757–13762.

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.

SAWYER, S. A., 2000   *GENECONV: A Computer Package for the Statistical Detection of Gene Conversion*. St. Louis (http://www.math.wustl.edu/~sawyer/geneconv/).

SCHIERUP, M. H., and J. HEIN, 2000   Consequences of recombination on traditional phylogenetic analysis. Genetics **156:** 879–891.

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:** 1279–1287.

STAHL, E. A., and J. G. BISHOP, 2000   Plant-pathogen arms races at the molecular level. Curr. Opin. Plant Biol. **3:** 299–304.

STAHL, E. G., G. DWYER, R. MAURICIO, M. KREITMAN and J. BERGELSON, 1999   Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature **400:** 667–671.

SUN, Q., N. C. COLLINS, M. AYLIFFE, S. SMITH, S. M. DRAKE *et al.*, 2001   Recombination between paralogues at the *rp1* rust resistance locus in maize. Genetics **158:** 423–438.

SWOFFORD, D. L., 2001   *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.

TAKAHATA, N., and Y. SATTA, 1998   Footprints of intragenic recombination at *HLA* loci. Immunogenetics **47:** 430–441.

THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994   The CLUSTALX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **22:** 2673–2680.

VISION, T. J., D. G. BROWN and S. D. TANKSLEY, 2000   The origins of genomic duplications in *Arabidopsis*. Science **290:** 2114–2117.

WALSH, J. B., 1985   How many processed pseudogenes are accumulated in a gene family? Genetics **110:** 345–364.

WALSH, J. B., 1987   Sequence-dependent gene conversion: Can duplicated genes diverge fast enough to escape conversion? Genetics **117:** 543–557.

WALSH, J. B., 1995   How often do duplicated genes evolve new functions? Genetics **139:** 421–428.

WANG, Z., G. TARAMINO, D. YANG, G. LIU, S. V. TINGEY *et al.*, 2001   Rice ESTs with disease-resistance gene- or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. Mol. Gen. Genet. **265:** 302–310.

WILSON, W. A., S. E. HARRINGTON, W. L. WOODMAN, M. LEE, M. E. SORRELLS *et al.*, 1999   Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. Genetics **153:** 453–473.

WIUF, C., and J. HEIN, 1999   The ancestry of a sample of sequences subject to recombination. Genetics **151:** 1217–1228.

WIUF, C., and J. HEIN, 2000   The coalescent with gene conversion. Genetics **155:** 451–462.