Resistance gene evolution
Pamela C Ronald

Plant resistance genes are highly polymorphic and have diverse recognition specificities. These genes often occur as members of clustered gene families that have evolved through duplication and diversification. Regions of nucleotides conserved between family members and flanking sequences facilitate equal or unequal recombination events. Transposition contributes to allelic diversity. Resistance gene clusters appear to evolve more rapidly than other regions of the genome, and domains responsible for recognitional specificity, such as the leucine-rich repeat domain, are subject to adaptive selection.

Addresses
Department of Plant Pathology, 1 Shields Avenue, University of California at Davis, Davis, CA 95616, USA; e-mail: pcronald@ucdavis.edu

http://biomednet.com/elecref~1389526600100294
© Current Biology Ltd ISSN 1369-5266

Abbreviations

dn nonsynonymous substitutions
ds synonymous substitutions
kb kilobase
LRR leucine-rich repeat
MHC major histocompatibility complex
pb base pair
R resistance
RLA resistance-like analog
TE transposable element

Introduction

The continued survival of most organisms depends on the presence of specific genetic systems that maintain diversity in the face of a changing environment. Classic examples include antigenic variation in trypanosomes and immunoglobulin gene formation in mammals. Similarly, most plant species contain a large number of highly polymorphic disease resistance (R) genes, most of which share common structural domains [1]. It has long been speculated that DNA rearrangements play a key role in the evolution of these genes, thus allowing plants to generate new resistances to match the changing pattern of pathogen virulence [2,3]. In support of this hypothesis, studies of the maize disease resistance locus \( r p l \) revealed that recombination of flanking markers was associated with the creation of novel resistance phenotypes [4]. This review focuses on the evolution of R genes using recent information gained from molecular genetic analysis of R genes.

Genomic organization of resistance genes

R genes of different structural classes conferring resistance to diverse pathogens are present in the plant genome in nonoverlapping discrete clusters (groups of genes of related structure and/or function) [5,6**]. Within a discrete cluster, members of an R gene family are often arranged as tandem direct repeats, which is consistent with an origin through gene duplication and their continued evolution through unequal exchange. There are also R loci which consist of a single gene with multiple distinct alleles. For example, the \( L \) rust resistance locus in flax has 13 different specificities [7].

In addition to R genes with known specificities, resistance-like analogs (RGAs sequences, whose function is unknown, map as clusters in rice, \textit{Arabidopsis}, potato, tomato and soybean [8,9*,10,11*]. On the basis of comparative mapping studies of monocot RGAs, Leister \textit{et al.} [9*] suggest that R genes diverge more rapidly than the rest of the genome through sequence divergence or ectopic recombination. For example, using rice and barley RGAs, for mapping the rust resistance loci [7], only 17 loci were identified but only five were found at syntenic map locations. Similarly, the barley \( m l o \) and \( R p g l \) genes, conferring resistance to the powdery mildew and stem rust fungi respectively, are not found in the syntenic region in the rice genome although the order of flanking markers is conserved between barley and rice [12,13]. These results contrast to the synteny observed in most other cereal genes [14].

Although the clustering and rapid evolution of R genes suggests that a gene conferring resistance to one pathogen species could evolve to recognize a different pathogen species, there is no direct evidence yet to support this hypothesis. Future cloning and sequencing of linked genes conferring resistance to different pathogens may eventually demonstrate such a common evolutionary origin.

In plants, leucine-rich repeat (LRR) domains of R gene products show similarity to domains in diverse proteins controlling cell–cell communication in development and signaling, suggesting that both classes of genes may have evolved through duplication and divergence of common ancestors [15–17]. To date, genes controlling development have not been found within R gene clusters.

Duplication and recombination

Duplication plays a central role in creating complex genetic systems [18]. Duplication can create new loci, alter gene family number through recombination, or generate repeated sequences within a gene. For example, studies of the major histocompatibility complex (MHC) showed that human and mouse genomes contain regions that apparently emerged as a result of chromosomal duplication [19]. Similarly, at least two additional clusters of the
resistance gene Cf-9 homologues on the short arm of chromosome 1 of tomato have been found [20] and chromosomal duplications created entirely new clusters of R genes in lettuce [21].

Recombination can lead to amplification or reduction of the number of R gene family members. For example, the presence of two nearly identical functional Cf-2 genes suggests that they arose through a recent gene duplication event [22]. Analysis of the Cf-2/Cf-5 locus, where only a few sequences homologous to Cf genes reside, has revealed a rare disease-susceptible recombinant that arose via an unequal crossover event leading to a reduction of the Cf homologue numbers [22]. Molecular analysis of five Cf-4/Cf-9 disease-sensitive recombinants demonstrated that each was generated by chromosomal mispairing of intergenic sequences and unequal crossing over [23*]. The Xa21 multigene family encoding resistance to bacterial blight in rice contains a large duplication of at least 17 kb; one of the duplicated genes confers the same race-specific resistance as Xa21 [24*,25*]. The presumed duplication and diversification of the tomato Pto gene family led to the generation of alternative recognition capabilities of the encoded proteins [26,27]. Finally, it has been proposed that the flax M rust resistance locus, which carries tandemly arrayed specificities, evolved from a rare duplication of an ancestral M gene [28]. Repeated DNA flanking the locus may have enhanced subsequent duplication through unequal crossing over events. These results indicate that gene duplication is a major force in R gene evolution.

In some cases, recombination between diverged family members occurs at highly conserved stretches of nucleotides. For instance, a large proportion of recombination events at the Xa21 locus were localized to a highly conserved domain in the 5' coding domain, resulting in new promoter/gene combinations [24*]. Similarly, the recombination exchange site in the M mutants can be localized to a 45 bp region that is invariant between LRR repeats [29*]. Such recombination events can lead to gross structural changes.

In addition to swapping of large gene regions, recombination can lead to fine structural changes within a gene. The repetitive structure of LRR coding regions could facilitate intragenic (and intergenic) gene recombination leading to expansion and contraction of the LRR number, as demonstrated in mutants of M and Rpp5. Whereas the wild-type M gene contains two DNA repeats encoding LRRs, spontaneous mutants contain a single repeat [29*]. The mutant alleles with a single LRR repeat may have been generated by an unequal exchange between the first repeat in one M gene and the second repeat in its homolog [29*]. A fast-neutron generated susceptible Rpp5 mutant contains an intragenic duplication of four complete LRRs. This duplication may have arisen from an unequal crossing over event between two sequences of identity in the LRRs [30*].

Rapid sequence exchange among tandemly repeated gene families generally leads to sequence homogenization between members [6**]. How can variability, therefore be maintained in R gene family members? To address this question, Parniske et al. [6**] sequenced three haplotypes at the Cf-4/Cf-9 locus. Comparison of intergenic regions revealed a high degree of sequence rearrangements, whereas in the coding regions a patchwork of sequence similarities was observed [6**]. The observed variable sequence patches could result either from successive rounds of reciprocal recombination or from gene conversion events. In a homozygous background, the Cf-9 gene was found to be very stable. In contrast, the meiotic stability of Cf-9 was dramatically reduced in a Cf-4/Cf-9 transheterozygous background. Parniske et al. [6**] propose that the polymorphism of the intergenic regions suppresses unequal recombination in homozygotes and sister chromatids, thereby preventing sequence homogenization of the gene family. In this situation, recombination between regions of high homology within a coding region may actually contribute to the maintenance of a useful combination of R gene specificities. In a Cf-4/Cf-9 transheterozygous background, homologous sequences aligned unequally are used as recombination templates. Such unequal recombination alters the number of gene family members as well as the composition of the clusters, resulting in increased variation within the population.

Lesion-mimic mutants

Recombination at R loci can also lead to the generation of lesion-mimic mutants which display a phenotype similar to the hypersensitive response controlled by R genes, but in the absence of pathogen. This observation led to the hypothesis that similar types of genes are involved in both phenotypes and that the lesion-mimic mutant genes may be derived from R gene loci [2]. The recovery of four rust resistance Rpl alleles with lesion-mimic phenotypes is the most direct evidence to date that at least some of the lesion-mimic mutants are variants of race-specific R genes [31]. Flanking marker analysis indicated that at least two of the four mutants were derived from crossover events.

The barley powdery mildew resistance gene mlo and the lsd (lesion stimulating disease) and add2 (accelerated cell death) genes from Arabidopsis provide other examples of genes displaying a lesion-mimic mutant phenotype together with defense responses associated with disease resistance [32–34]. In these cases, however, no genes conferring race-specific resistance have yet been mapped to these loci. Moreover, these lesion-mimic genes encode proteins with structures distinct from other cloned R genes, indicating that not all lesion mimic mutants have a direct evolutionary link to R genes [32,35].
Adaptive selection of pathogen recognition domains

Characterization of nucleotide substitution patterns in R gene families has provided insight into the function and evolution of particular coding domains. For the investigation of function, the ratio of nucleotide substitutions that lead to amino acid replacements (nonsynonymous substitutions, dn) and nucleotide substitutions that do not alter amino acids ( synonymous substitutions, ds) is particularly informative. In most protein-coding genes, the dn/ds ratio is less than one; this observation is consistent with functional constraint against amino acid replacements [36]. Conversely, a dn/ds ratio significantly greater than one indicates that adaptive selection events have fueled divergence between genes [37,38]. Evidence of adaptive selection is rare but appears to be most common in gene regions encoding surface antigens of parasites or viruses [39]. It is expected that regions that bind ligand will be subject to stronger adaptive selection than regions that play a structural role. For example, the antigen recognition site of alleles at the class I MHC loci in human and mouse displays a dn/ds ratio greater than one, indicating that the antigen recognition site is subject to strong adaptive selection events, whereas structural regions of the protein are not [37].

Analysis of 11 Cf gene family members revealed that the predicted solvent-exposed residues of the β-strand/β-turn region of the LRR domain exhibit increased dn/ds ratios relative to other residues in the LRR domain, suggesting that solvent-exposed residues play a role in ligand binding [6**]. Similarly, a comparison of nucleotide substitutions in the LRR coding regions of Xa21 and gene family member Xa21D revealed that, although Xa21 and Xa21D share 99.1% sequence identity, nonsynonymous substitutions occur significantly more frequently than do synonymous substitutions in the LRR; this result is consistent with the LRR's putative role in ligand binding [25*]. These results indicate that the LRR domain, which governs race-specific pathogen recognition, is subject to adaptive evolution. Diversity at the LRR domain would provide an evolutionary advantage for recognizing, binding, and defending against a broad array of pathogens.

Diversification of R gene family members by transposon-like elements

The human MHC class 2 region is among the most polymorphic part of the human genome. Multiple repetitive sequences representing more than 20 different families have been characterized in the MHC region [40]. Part of the interspecific and intraspecific variation observed in the MHC is caused by different integration patterns of retroelements. Comparative studies of different human haplotypes and primate species revealed that retroelement insertions have contributed to genome plasticity of the MHC during primate evolution. Retroelements also contribute to recombination and genomic instability by serving as sites for recombination and translocation events [40].

In plants, it has long been hypothesized that transposable elements (TEs, or transposons) play a role in the reconstruction of genomes in response to environmental stresses such as tissue culture, irradiation or pathogen infection [41,42]. In partial support of this hypothesis, Pouteau et al. [43] demonstrated that the transcription of the tobacco retrotransposon Tnt1 is induced by a broad spectrum of microbial and fungal elicitors. TE insertion into and excision from regulatory and coding regions can change the coding capacity and expression patterns of the gene [44-46].

There is no evidence yet for the generation of new specificity at R gene loci as a result of the insertion and excision of a TE. It has been shown, however, that TE-induced gene alterations can cause R gene inactivation. For example, in the case of the maize fungal resistance gene Hml, which confers resistance to Cochliobolus carbonum race 1, a 315 bp insertion (designated ΔHBr) was found in a mutant allele of this gene [46]. Moreover, the insertion of a transposon (a 256 bp element named Drone) disrupted the Hml-conferred resistance in an inbred line of maize and, as a result, led to the genesis of the leaf spot and ear rok disease of maize in 1938 [47]. In flax, two mutants of the L6 gene for rust resistance carry small (300 bp) insertion elements which inactivate the gene [48].

In rice, transposon-like elements appear to be a major source of variability of the Xa21-gene family members. Seventeen transposon-like elements grouped into 11 families, including three families of miniature inverted repeat repeat TEs (MITEs), five novel elements, Dt-like elements, a CAGTA-like element and a retrotransposable element are present at the Xa21 locus [24*,49,50]. Integration of two of these elements into coding sequences creates open reading frames (ORFs) that encode truncated proteins. At least one of these truncated proteins can confer an attenuated resistance with Xa21 specificity [25*]. TE insertion into 5' and 3' flanking regulatory regions was also observed. Many of the elements seem to have been active over the entire evolutionary period of the Xa21 gene family members [24*].

It is tempting to speculate that TEs contribute to the evolution of R gene diversity. Movement of these transposons in response to pathogen-induced stress would provide genetic plasticity with a possible selective advantage. As the insertion of TEs at the Hml, L6 and Xa21 loci results in loss of function or impaired function, however, the question of whether TE movement can confer a selective advantage remains open. In addition, more sequence information is needed to determine if TEs are more abundant or more active at R gene loci than in other regions of the genome.
Conclusions
Common themes in the evolution of R genes and gene families are emerging, on the basis of sequence analysis of cloned R genes. First, duplication and subsequent divergence of a progenitor R gene can amplify or create additional clusters of a gene family. Second, unequal recombination at intergenic regions between family members creates additional copy number variability within the population. Third, recombination at highly conserved regions in intragenic regions allows for the formation of novel gene combinations. Fourth, adaptive evolution of LRR domains allows for rapid generation of altered recognition specificities. Finally, movement of TEs may result in further allelic diversity.

Acknowledgements
The author is grateful to PA Anderson, JL Dangl, SH Hubbert, G Johal, DA Jones, JDG Jones, J McDowell, RW Michelmore, B Meyers, T Richter and W-V Song for helpful discussions and critical reading of the manuscript.

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


This work indicates that the LRRs can serve as a recombination template leading to variation in the presumed ligand recognition domain. The wild-type M gene contains two DNA repeats encoding LRRs, whereas spontaneous mutants contain a single repeat.


A fast-neutron generated Rpp5 mutant contains an intragenic duplication of four complete LRRs. Another example that the LRRs can serve as a recombination template leading to variation in the presumed ligand recognition domain.


