Defenses against most specialized plant pathogens are often initiated by a plant disease resistance gene. Plant genomes encode several classes of genes that can function as resistance genes. Many of the mechanisms that drive the molecular evolution of these genes are now becoming clear. The processes that contribute to the diversity of R genes include tandem and segmental gene duplications, recombination, unequal crossing-over, point mutations, and diversifying selection. Diversity within populations is maintained by balancing selection. Analyses of whole-genome sequences have and will continue to provide new insight into the dynamics of resistance gene evolution.

Addresses
Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19714, USA

Corresponding author: Meyers, Blake C (meyers@dbi.udel.edu)

Current Opinion in Plant Biology 2005, 8:129–134
This review comes from a themed issue on Genome studies and molecular genetics
Edited by Susan Wessler and James C Carrington
Available online 29th January 2005
1369-5266/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.

Introduction
Plant disease resistance genes (R genes) encode proteins (R proteins) that represent the first line of defense against infection by many biotrophic pathogens. These pathogens are usually highly specialized for specific host plants, and the interaction at the molecular level is often complex because of the co-evolution of the host and pathogen. Our understanding of the molecular mechanisms that underlie these interactions has dramatically improved in recent years. As a result, the genes that are crucial to the plant–pathogen interaction have formed the basis of numerous and ongoing evolutionary analyses. Natural selection has had a powerful influence on both the pathogen-produced avirulence or ‘effector’ molecules, which are crucial to the infection process, and host R genes, which are critically important to plant defense. Recent studies demonstrate how selection has shaped single R genes or gene families as well as gene frequencies in populations. Advances in the evolutionary analysis of R genes have benefited from sequence analysis, comparative genomics, phylogenetic analysis, large-scale genome sequencing, and population studies. In particular, the past few years have seen remarkable progress in understanding the mechanisms of R gene evolution in Arabidopsis, but notable findings have also emerged from studies in other plant species. In this review, we focus on recent studies describing the evolution of plant R genes.

R gene classes
Many R genes have now been cloned from a wide variety of plant species. One of the best recent reviews of R gene structure and function is that by Dangl and Jones [1]. It is now clear that most R genes encode proteins that have a putative amino-terminal signaling domain, a nucleotide-binding site (NBS) and a series of carboxy-terminal leucine-rich repeats (LRRs). These ‘NBS-LRR’ proteins have been divided into two major classes: those with an amino-terminal TIR (Toll/interleukin receptor) domain (which are known as TIR-NBS-LRR or TNL proteins) and those that encode an amino-terminal coiled-coiled motif (CC-NBS-LRR or CNL proteins). The details of the molecular functions of these protein domains and their interacting partners are still being established. However, the consistent identification of this class of proteins across diverse plant species demonstrates that the NBS-LRR genes are a pillar of plant defenses.

Several classes of plant R genes have been identified in addition to the seemingly pervasive NBS-LRR class of R genes. The powdery mildew resistance gene RPW8 is an interesting case; this Arabidopsis protein is small, has only an amino-terminal transmembrane domain and a CC domain, and confers an unusually broad spectrum of resistance [2]. The rice Xa21 gene encodes another type of R protein, a receptor-like kinase (RLK) that contains both LRR and kinase domains [3]. A second rice RLK that is involved in disease resistance, Xa26, has now been cloned, suggesting that other RLKs may recognize pathogens, at least in rice [4]. Yet another class of R genes, those that encode extracellular LRRs with a transmembrane domain (i.e. receptor-like proteins [RLPs]), is typified by the tomato genes conferring resistance to Cladosporium fulvum (Cf) [5,6]. In 2004, it was demonstrated that RLPs are also used in other plant species for pathogen recognition; the Arabidopsis RPP27 and the apple HcrVf2 R genes were both shown to encode RLPs [7,8]. Previously, most classes of R proteins other than NBS-LRRs were represented by only one gene or locus, raising the question of whether pathogen recognition by these proteins types was exceptional. Genes that encode RLKs and RLPs are abundantly represented in plant genomes and have well-characterized roles in plant development. So, the identification of defense functions for these proteins in...
multiple plant species suggests a wider role for these proteins in plant defenses. The diversity of species from which RLK and RLP $R$ genes have been identified suggests that the resistance function either evolved by convergence or originated in a common ancestor of these species. Evolutionary analyses of other classes of $R$ genes, such as that including the tomato kinase $Pto$, have also been described [9,10], but the majority of $R$ genes encode NBS-LRRs, and evolutionary studies have primarily focused on this class.

**Genome-wide evolutionary analyses of $R$ genes**

**Genomic sequence analyses**

Several recent publications have utilized the complete sequence of *Arabidopsis* to infer the evolutionary forces acting on the NBS-LRR protein family. Studies have focused on this family of proteins because its only known function is in disease resistance. The *Arabidopsis* genomic sequence contains 149 NBS-LRR genes and 58 shorter related genes [11**]. Classification based on protein domains, intron positions, sequence conservation, and genome distribution has been used to define specific subgroups of CNL and TNL proteins. Although the TNL family is nearly twice the size of the CNL family in *Arabidopsis*, the data suggest that TNLs are more homogeneous and have amplified more recently in this genome than have CNLs [11**]. Sequence comparisons among subgroups of *Arabidopsis* NBS-LRRs provided evidence that selection has acted to diversify LRR sequences, as has been shown for numerous clusters or families of $R$ genes [12]. Similar genetic studies have identified and characterized 59 RLPs and more than 600 RLKs from *Arabidopsis* [7,13]. However, as described above, few of these genes are known to function as $R$ genes.

Rice presents the next genome for which an extensive analysis of NBS-LRRs will be possible. Several groups have published early results obtained using the incomplete rice sequence [14,15]. More than 500 NBS-coding sequences have been characterized in rice, nearly all of which encode CNLs and none of which encode TNLs [14–16]. This NBS-LRR gene family in rice is quite diverse. However, analysis of *Arabidopsis* NBS-LRRs demonstrated that an accurate final analysis requires manual re-annotation and the verification of gene predictions, and this can only be performed on finished sequence [11**]. For evolutionary analyses in any species, the availability of a second, related genomic sequence would help immeasurably. For rice, comparisons of the rice *indica* and *japonica* sequences might be informative [17,18], although the sequences of more distantly related species would be more useful.

Genomic analyses have identified the TIR-containing sequences as a particularly interesting group of proteins. Two new families of TIR-containing proteins, which are encoded by more than 50 genes, were also identified in the *Arabidopsis* genome [16]. The TIR-X (TX) family lacks both the NBS and LRRs that are characteristic of the R proteins whereas, when compared to TNL proteins, the TIR-NBS (TN) proteins lack only the LRR domains [16]. TX and TN proteins are encoded in the genomes of conifers and grasses, and two TN proteins are extremely well-conserved in *Arabidopsis* and rice, suggesting that these are ancient protein families [16]. No TNL-encoding genes have been identified in cereal genomes, although they are found in gymnosperms, suggesting that the grasses might have lost this type of gene [14,16]. Intriguingly, the *Arabidopsis* $R$ gene RPP2A has recently been demonstrated to encode a protein that is similar to a fusion of TN and TNL proteins [19]. Many questions remain about the evolution of TIR proteins in plants, not the least of which is whether the presence of TIR-based defenses in both plants and animals is due to convergence or conservation (see below).

**Clustering of $R$ genes and the impact of genomic duplication events**

Genetically defined clusters of $R$ genes are well known, and molecular studies have demonstrated that this clustering usually results from tandem duplications of paralogous sequences [11**,20–22]. This clustering is a well-known phenomenon observed at many $R$ gene loci [23]. The many NBS-LRRs encoded in the *Arabidopsis* and rice genomes are found in numerous clusters, and expansion within these clusters is predicted to be a consequence of tandem duplications resulting from unequal crossing over [11**,15,21,24*]. Clusters of closely related and co-localized $R$ genes frequently exchange sequences, but there is no evidence of sequence exchange among related genes that are located in separate clusters [11**,24*]. Analyses of the *Arabidopsis* genome indicate that numerous small-scale genomic duplications have copied or translocated one or several NBS-LRR genes from these clusters to distal and probably random locations in the genome (‘ectopic duplications’) [11**,24*,25*]. Heterogeneous genomic clusters comprised of non-paralogous NBS-LRR sequences have probably formed by chance due to random rearrangements [11**].

Segmental genomic duplications have probably caused a substantial increase in the number of *Arabidopsis* NBS-LRR sequences [11**,24*,26]. A recent review by Leister [25*] nicely compares reports that have described the involvement of tandem and segmental duplications in *Arabidopsis* NBS-LRR evolution. Soon after most genomic segmental duplication events, many duplicate NBS-LRRs might have been lost, possibly because of selection [26]. At some loci, tandem duplications have expanded gene families and the duplicated sequences have diverged through accumulated mutations, increasing
the complexity of $R$ gene sequences [24*,25*]. Although NBS-LRRs show rates of duplication and mutation that are similar to those of other gene families [11**], natural selection might differentially shape the composition of this gene family.

**Population-level insights into $R$ gene evolution**

$R$ gene evolution must be considered in the context of naturally variable populations. Population studies have demonstrated that balancing or frequency-dependent selection maintains $R$ gene loci in a polymorphic state within a population. Selection is balanced by the positive impact of the enhanced fitness of the host in the presence of the pathogen and by a negative impact of the $R$ gene on host fitness in the absence of the pathogen. Balancing selection contrasts with the so-called ‘arms race’ model in which host and pathogen are continually ratcheting up the effectiveness of defensive and offensive proteins [27,28]. The ‘arms race’ hypothesis might adequately describe the evolution of a resistance gene recognizing a single pathogen ‘avirulence’ gene that is uniquely essential for infection of a homogeneous plant population. However, this does not adequately reflect the true complexity of populations, and pathogens express and secrete a diverse set of proteins, the combination of which is probably important for infection [29]. The diversity and dynamics of the natural host and pathogen populations, and the interactions of these populations, make it unlikely that any single $R$ gene or allele will be driven to fixation. For example, populations of *Arabidopsis* segregate for a functional and a null allele at the *RPM1* and *RPS5* loci, whereas the *RPS2* locus segregates for resistant and susceptible alleles [30–32]. Analyses of sequence diversity in these populations indicate that resistant and susceptible haplotypes have probably been maintained over an extensive period of time, consistent with balancing selection. It is possible that populations remain polymorphic for the presence of certain $R$ genes, despite the benefit of these genes in defense against pathogens, because of a fitness cost that is associated with the presence of functional alleles [33**]. However, most published studies have focused on *Arabidopsis* genes that confer resistance to *Pseudomonas syringae*, using strains of this pathogen isolated from crop species. An analysis of *Cf-2* homologs isolated from populations of a wild relative of tomato identified wide variation in these $R$ gene sequences, and the authors suggest that this polymorphism is favored by balancing selection [34]. Another recent study assessed the diversity of *RPP13* in natural populations of *Arabidopsis* and the pathogen *Peronospora para- sitica* [35*]. An analysis of 24 accessions of *Arabidopsis* showed that extremely high levels of polymorphism are maintained at the *RPP13* locus. This variation was attributed in part to extensive recombination [35*]. The rates of synonymous and non-synonymous substitutions were consistent with balancing selection that favors amino-acid variation in the LRR region of this gene [35*]. An interesting follow-up to this study resulted from the recent cloning of the *P. parasitica* avirulence gene *ATR13*. The *ATR13* gene is highly polymorphic in the pathogen population, consistent with balancing selection maintaining variation in both host and pathogen populations [36**]. Additional studies are needed, but it is clear that balancing selection maintains polymorphism at many $R$ gene loci.

One of the most intriguing and extensive analyses of $R$ genes across diverse accessions demonstrates the presence of two distinct types of genes at the *RGC2* locus of lettuce [37**]. Specific resistance in lettuce to the downy mildew pathogen is conferred by the *RGC2* family member *Dm3*, which is one of at least 32 NBS-LRR-encoding genes clustered at this locus. Kuang et al. [37**) amplified and sequenced 126 distinct *RGC2* fragments from seven lettuce genotypes. These sequences apparently represent two distinct types of genes, differentiated by numerous measures of evolutionary rates [37**]. Type I genes were characterized by a higher frequency of sequence exchanges, resulting in more chimeric genes. Hence, individual Type I genes were observed across different accessions less frequently than were Type II genes. Orthologs of Type II genes were found in different lettuce accessions, possibly maintained under the influence of purifying selection. Overall, the *RGC2* data suggest that the Type II genes evolve more slowly than Type I genes, reflecting different rates of evolution and selective pressures [37**]. As the lettuce *RGC2* cluster is exceptional in its size and complexity and no comparable loci exist in *Arabidopsis*, it will be interesting to see if the *RGC2* duality is corroborated in other clusters and plant species. One possible mechanism for the creation of two types of genes could be that genes that are located centrally in a cluster might evolve differently than genes in more terminal locations. This characteristic was described for the tomato *Cf-9* cluster and was also noted for the lettuce *RGC2* cluster [37**,38].

**Cross-species comparisons**

Relatively few studies have analyzed patterns of $R$ gene evolution across species, in part because of the difficulty in identifying and determining the function of truly orthologous sequences. One such analysis focused on the $R$ gene *RPW8*, which confers broad-spectrum resistance to powdery mildew, has recently been published [39*]. Analysis of the syntenic *RPW8* family in four *Brassica* and *Arabidopsis* species demonstrated duplication, deletion, and divergence within the gene family, themes that are common to the evolution of all $R$ gene families. The *RPW8* gene is particularly unusual because, unlike other cloned $R$ genes, it functions as a transgene in an unrelated genus, suggesting the conservation of the interacting components and signaling systems across plant families [40]. Most $R$ genes might not function in divergent species because of a requirement for
species-specific interacting proteins that have co-evolved with the R gene.

A second intriguing study resulted from the cloning of genes from divergent plant species that recognize the same bacterial avirulence protein. Ashfield et al. [41,42**] cloned Rpg1-b from soybean and compared this gene to the functionally analogous CNL gene RPM1 of Arabidopsis. Both genes confer resistance to Pseudomonas syringae expressing acrB. The level of amino-acid sequence identity between the encoded RPM1 and Rpg1-b proteins is low and they are not orthologous; because both plant genes recognize the acrB specificity, these genes must have independently evolved nearly identical functions [42**]. The evolutionary implications of this first identified case of the convergent evolution of an R gene have been reviewed by McDowell [43].

Ancestral inferences

Genes that encode NBS-LRRs have been identified in gymnosperms and non-vascular plants [44,45]. Related proteins are also involved in innate immunity in mammals. For example, the similarity between the NBS of plant R proteins and that of the mammalian apoptotic response protein Apaf-1 was identified several years ago [46]. It is also interesting that NBS and LRR domains are both present in the mammalian Nod family of immunity-related proteins [47] and in a family of more than 14 pyrin-containing Apaf-1-like proteins [48]. In addition, the TIR domain has been shown to mediate protein–protein interactions in the animal innate immune system [49,50], although TIR-encoding genes, such as the TX, TY and TNL genes, are more common in plant genomes than in animal genomes. The involvement of both TIR-containing and NBS-LRR proteins in animal and plant defense systems suggests a common and ancient origin, but additional analyses of the extant progenitors of higher plants are needed to determine the origin of plant defenses and of R genes.

Conclusions

Within the past few years, studies have shown the important contributions that tandem and segmental gene duplications, recombination, mutation and natural selection have made to the evolution and diversity of plant R genes. In the next few years, evolutionary studies of R genes are likely to take advantage of additional plant genome sequences, population biology, and functional studies of paralogs and orthologs. Numerous important questions remain to be addressed about the ancestral origin and evolution of plant R genes. How extensive among plant families is the notable absence of TNLs that has been observed in grasses? Why do dicots utilize two types of NBS-LRR R proteins, TNLs and CNLs, and are they entirely functionally equivalent? How extensive is the redundancy of the signaling systems for each of the classes of R proteins and where do the pathways converge? Is the apparent relationship between plant and animal immune responses a result of convergent evolution or did these defense systems share a common origin? Combined with focused experimentation, the expanding availability of plant genomic resources in model and crop species and in non-vascular plants will create novel opportunities to develop and test hypotheses about R gene evolution.

Acknowledgements

We are grateful to Ken Nobuhara and John McDowell for helpful comments on the manuscript. Funding from the University of Delaware Research Foundation (UDRF) and a National Institute of Health (NIH) Centers of Biomedical Research Excellence (COBRE) award (P20 RR015588) supports research on disease resistance in the Meyers laboratory.

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

NBS-LRR proteins. This paper describes the results of this genome-wide analysis and describes the evolutionary implications of these data.


24. On the basis of a genome-wide analysis of duplication events, the authors propose an evolutionary model that relates processes of genome evolution to mechanisms of evolution for the large diverse NBS-LRR family. The data suggest that segmental duplications and rearrangements have duplicated NBS-LRR sequences.


26. Several recent papers have analyzed the gene family encoding NBS-LRR proteins in the context of the Arabidopsis genome. This review compares and summarizes several of these papers, and the author describes several of the important evolutionary questions that remain unanswered regarding the influence of tandem and segmental duplications on R gene evolution.


33. Using a careful experimental design, the authors were able to measure the fitness penalty incurred by Arabidopsis expressing the RPPI1 R gene. Their data support the hypothesis that the frequency of an R gene in a population reflects a balance between selection by the pathogen and the fitness cost of expressing the R gene (i.e., ‘balancing selection’).


37. The cluster of NBS-LRR genes at the RGC2 locus of lettuce is the most complex described to date. Analyses of 126 LRR fragments from seven genotypes and 107 RGC2 sequences from 40 wild lettuce accessions identified two distinct types of RGC2 genes. Type I genes were characterized by evidence of frequent sequence exchange and a limited representation across accessions. ‘Type II’ genes apparently evolve much more slowly and are more conserved across accessions. It remains to be seen if these two types of genes are a general feature of R genes or complex clusters of R genes, or if this pattern is unique to the RGC2 family of lettuce.


40. RPWS8 is the only R gene known of its type. The authors investigated the evolutionary history of this gene by examining the sequence polymorphism in the syntenic RPWS8 loci of four Brassica species. They provide evidence that RPWS1.1 and RPWS8.2 evolved from a series of duplications of HR3, the probable progenitor gene of the RPWS8 family. The authors suggest that balancing selection maintains polymorphisms at the RPWS8 locus.


42. Ashfield T, Ong LE, Nobuta K, Schneider CM, Innes RW: Convergent evolution of disease resistance gene specificity in two flowering plant families. *Plant Cell* 2004, 16:309-318. In their respective genomes, the soybean gene *Rpg1*-b and the Arabidopsis gene *RPM1* both mediate recognition of the same bacterial effector protein (AvrB). These two plant genes both encode CNL proteins but demonstrate only limited sequence similarity and are apparently not orthologous. This is the first evidence of the convergent evolution of *R* genes.


