REVIEW

Evolutionary Dynamics of Plant *R*-Genes

Joy Bergelson,* Martin Kreitman, Eli A. Stahl, Dacheng Tian

Plant *R*-genes involved in gene-for-gene interactions with pathogens are expected to undergo coevolutionary arms races in which plant specificity and pathogen virulence continually adapt in response to each other. Lending support to this idea, the solvent-exposed amino acid residues of leucine-rich repeats, a region of *R*-genes involved in recognizing pathogens, often evolve at unusually fast rates. But within-species polymorphism is also common in *R*-genes, implying that the adaptive substitution process is not simply one of successive selective sweeps. Here we document these features in available data and discuss them in light of the evolutionary dynamics they likely reflect.

Genetic variation for disease resistance is characteristic of almost all species. In both mammals and plants, gene families involved in pathogen recognition pathways—the major histocompatibility complex (MHC) and plant Resistance (*R*) genes—contain loci that segregate for large numbers of alleles, some of which are highly divergent one from another. The antiquity of alleles in both humans (*1*, *2*) and plants (*3*) (but not *Drosophila*) (*4*) provides evidence for their evolutionary maintenance by some form of balancing selection.

In plant *R*-genes, polymorphism is often associated with loci that are present as tandem arrays of multiple copies. Because of promiscuous genetic exchange, paralogs within these clusters exhibit complex evolutionary relationships (*5*). Polymorphism in gene copy number is common in clusters and is also present in single-copy loci as the presence or absence of a locus. Copy number dynamics, intergenic exchange, and allelic diversity are all likely to be evolutionary responses to the same selective pressures for disease resistance. *R*-gene evolution, therefore, has both a vertical component across generations and a horizontal component throughout the genome, and each is likely to be shaped by natural selection for resistance.

The presence of ancient and many segregating alleles at *R*-gene loci is perplexing because disease resistance is thought to involve an evolutionary arms race between host and pathogen (*6*, *7*). A classic arms race is one that entails a series of selective sweeps as novel *R*-gene alleles, capable of recognizing pathogenicity determinants [called avirulence (*Avr*) factors] that previously avoided detection in a plant population, spread to high frequency. Support for these evolutionary dynamics centers on the common observation that amino acids evolve at a faster rate in functionally important regions of *R*-gene proteins than the corresponding rate of synonymous change (*8*–*12*). But according to the population genetics theory of selective sweeps, the rapid turnover of new *R*-gene specificity should cause a reduction in the age and number of alleles at a locus.

The longevity and high allelic diversity of some *R*-loci are inconsistent with a classic arms race and instead suggest a microevolutionary mechanism that promotes the maintenance of stable polymorphism (*3*). This review focuses on *R*-genes involved in "gene-for-gene" interactions (*13*) and, in particular, the class of *R*-genes containing leucine-rich repeat (LRR) regions. Our goal is to highlight features of *R*-gene variation and evolution that will be important in modeling the evolutionary dynamics of disease resistance. The striking parallels between mammalian MHC and plant *R*-gene variation suggest that understanding the evolutionary dynamics of disease resistance in plants will have applicability in other organisms, including humans.

Adaptive Divergence

Evidence for adaptive evolution is now commonplace in comparative studies of *R*-gene sequences. With few exceptions, studies draw results from the comparison of paralogs on a single chromosome. Although members of *R*-gene clusters often display evidence for intergenic exchange, paralogs can also be considerably diverged one from another, suggesting that exchange is no longer an active process contributing to the generation of new allelic variation.

Detecting adaptive evolution involves comparing amino acid substitution rates (K_a) to synonymous substitution rates (K_s) in the same gene. Under the assumption that synonymous changes approximate the neutral rate of molecular evolution, values of the ratio K_a:K_s greater than 1 provide evidence for positive selection for amino acid substitution. Detection of positive selection often depends on considering only those sites predicted to be important in recognition. Perhaps the most influential studies of this kind focused on class I and II MHC genes (*14*, *15*) and revealed positive selection as a force driving antigen recognition site evolution.

In plant *R*-gene products, several studies pinpoint LRR domains as the major determinants of recognition specificity for *Avr* factors (*16*). LRR regions are receptor domains for specific recognition of pathogen elicitors (*17*) and may be involved in direct protein-protein interactions with *Avr* gene products of the pathogen (*18*). On the basis of a model of LRR protein structure, solvent-exposed residues framed by aliphatic residues are predicted to be the amino acids involved in making these direct contacts (*17*). The framed, solvent-exposed residues often exhibit strikingly fast rates of evolution, but other regions within the LRRs can also be seen to evolve by positive selection (*11*, *19*), and specificity can reside outside of the LRR (*20*, *21*). Amino acids in the LRR may also influence the interaction with host factors, thereby modulating resistance through a mechanism other than recognition (*22*). Experimental investigation of functional differences among *R*-gene alleles has yet to be exploited fully as a tool for deciphering *R*-gene structure and function.

Adaptive divergence among *R*-gene paralogs has been investigated in tomato (*8*), rice (*23*), lettuce (*11*), and *Arabidopsis* (*9*, *10*, *12*). Without exception, complex loci reveal high rates of amino acid replacement changes in the exposed residues of domain 2 (the framed region) of the LRR (Table 1), almost always being more than twice the rate of synonymous substitution. Adaptive evolution in the LRR is consistent with an evolutionary arms race in that, under this model, pathogens should impose selection to continually alter recognition specificity.

Rates of evolution of single-copy *R*-loci can be determined from orthologous comparisons between species. Estimates have been obtained for three genes, *Rpm1*, *Rps2*, and *Rps5*, and all involve comparisons between *Arabidopsis thaliana* and its congener, *A. lyrata*. For each of these *R*-loci, amino acid replacement changes have accumulated considerably more slowly than synonymous changes (Table 1). In interpreting these results, it is important to realize that although large values of K_a:K_s provide strong evidence for adaptive evo-

Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: jbergels@midway.uchicago.edu

lution, small values do not strongly indicate its absence. K_a:K_s ratios represent the confluence of constraint, genetic drift, and adaptive evolution; the lower K_a:K_s ratios found in *Rpm1*, *Rps2*, and *Rps5* are compatible with adaptive evolution, albeit at a slower rate than that seen for complex loci.

Origin of New Alleles

Does intergenic exchange in multicopy clusters facilitate adaptive host response to pathogen pressure? In principle, the clustering of related *R*-genes can increase the opportunity for genetic exchange among paralogs that could act as reservoirs of mutational variation (*16*). Evidence supporting this idea comes from experimental investigations of both spontaneous mutants (*24*, *25*) and chimeric alleles (*20*) that confer novel specificity and also from differences in rates of evolution in single- versus multicopy *R*-genes (described above). But whether *R*-genes in complex loci are subject to different selection pressures than single *R*-genes is debated (*16*, *26*). Furthermore, little is known about baseline rates of cluster origination, expansion and contraction, and genetic exchange among *R*-genes within clusters.

To begin exploring these issues, we analyzed the *A. thaliana* genome sequence (*27*). *R*-loci are physically arranged in the *Arabidopsis* genome as 49 single *R*-loci and 32 clusters of 2 to 12 *R*-genes. Evolutionary analysis of 182 known and putative *R*-genes generated 20 clusters of related genes, 15 of which contain LRR regions (*28*). We additionally identified two sets of physically dispersed single *R*-gene loci (*28*). These data allowed us to ask whether clustered *R*-loci exhibit faster evolution than single *R*-genes (Fig. 1).

Positive selection appears to be common in *R*-genes belonging to evolutionary clusters; we find at least one pairwise comparison with K_a:K_s greater than 1 for the exposed residues of domain 2 in 11 of 17 sets. Furthermore, evidence for positive selection is seen at all cluster sizes, including isolated *R*-genes, suggesting that selection acts similarly regardless of locus complexity. Nevertheless, cluster size and average rates of adaptive evolution are weakly, positively correlated (P = 0.051; one-tailed test), but the correlation explains only a small proportion of the overall variance in rates (R^2 = 0.17).

More striking is the observation that closely related *R*-genes are uncommon in the *Arabidopsis* genome: Divergence averages K_s = 0.46 across sets, and only three sets (I-22.3, V-16.4, and IV-8.7 = *Rpp5*) exhibit any pairwise divergence with $K_s < 0.1$. As already noted, *R*-gene paralogs may be largely independently evolving entities, and, if so, they would be compatible with a birth/death model for their evolution (*29*). The rarity of evolutionary clusters with young *R*-genes indicates low rates of *R*-gene duplication, loss, and genetic exchange. This suggests that, where observed (e.g., *Rpp5*), natural selection may be the driving force favoring (otherwise) rare recombinants. Our observation that clusters with the most closely related paralogs show the highest K_a:K_s ratios suggests selection for novel alleles generated through duplication or recombination.

Allelic Polymorphism

R-gene polymorphism is an important component of variation for resistance to pathogens. Does this allelic variability represent transient polymorphism arising during the adaptive spread of novel resistance alleles, or does it represent evolutionarily stable polymorphism? This question can be answered by investigating the genealogies of resistance and susceptibility alleles. Under a selective sweep scenario, the alleles segregating at a locus must be descendants of the allele that most recently swept to high frequency. If the last selective sweep occurred recently, as is likely for rapidly evolving *R*-genes, then alleles segregating at a locus should be very closely related to one another and should be nearly identical in sequence. In contrast, long-lived polymorphism for resistance and susceptibility produces alleles with more ancient common ancestries, and regions tightly linked to the site(s) under selection will show relatively greater nucleotide divergence (*30*).

Levels of nucleotide polymorphism at synonymous sites, or in noncoding regions surrounding an *R*-gene, can be used to test these alternatives because accumulated changes here reflect the ages of alleles. Studies designed to determine ages of alleles have been published for only two loci, *Rps2* and *Rpm1*, both of which confer resistance in *A. thaliana* to *Pseudomonas syringae* pathovars. A polymorphism study of *Rps2* (*31*) found that resistance alleles were genetically similar whereas susceptibility alleles could be widely divergent from each other and from the resistance alleles. The deepest node in the gene tree separates resistance and susceptibility alleles, and these alleles differ by 29 synonymous changes and 11 amino acid replacement changes. Overall, these alleles differ by about 3% at synonymous sites, a value only slightly greater than that seen at other loci in this species. The shape of the gene tree, however, led the authors to posit a balanced polymorphism at the locus.

A stronger case for balancing selection is seen at *Rpm1* (*3*), a polymorphism for the presence or absence of the entire locus. Analysis of the junction region flanking this deletion reveals about 10% divergence between

Table 1. Adaptive evolution among paralogs and between orthologs. Nonallelic comparisons of *R*-gene loci. K_a:K_s is indicated for domain 2 (the framed region) of the LRR, and the complexity indicates the number of paralogs in each accession. Comparisons of *Rpp8* and *Rps5* involve unpublished data of the authors (*39*). *Rps2* combines published *A. thaliana* sequence (*31*) and unpublished *A. lyrata* sequence of Mauricio (*40*).

Plant locus	K_a:K_s in domain 2 of the LRR	Complexity	Nature of the comparison	Reference
Arabidopsis				
Rpp8	2.37	2/1	Paralogs from two accessions with differing copy number	(*10*)
Rpp5	2.13	10/8	Paralogs from two accessions with differing copy number	(*12*)
Rpp1	2.71	3	Paralogs from one accession	(*9*)
Rps2	0.25	1	Orthologs in *A. thaliana* and *A. lyrata*	(*40*)
Rps5	0.36	1	Orthologs in *A. thaliana* and *A. Lyrata*	(*39*)
Rpm1	0	1	Orthologs in *A. thaliana* and *A. lyrata*	(*3*)
Lettuce				
Dm3	2.06	24	Paralogs from one accession (nine members analyzed)	(*11*)
Tomato				
Cf-4/9	1.32	5/5/1	Paralogs in three accessions introgressed from three species	(*8*)
I2C-1 versus *I2C-2*	1.17	7	Paralogs from one accession	(*11*)
Mi copy 1 versus *2*	1.93	3	Paralogs from one accession	(*11*)
Rice				
Xa21 copies B, D, F	2.11	6	Paralogs from one accession	(*11*)

resistant and susceptible lineages, suggesting that the origin of the polymorphism dates to around the speciation event separating *A. thaliana* and *A. lyrata*. If a resistance allele is selectively deleterious in the absence of the pathogen it recognizes (*32*), then it is possible for a deletion (i.e., loss of function mutation) to be a balanced polymorphism (*3*). A similar pattern of divergence in the region flanking the insertion/deletion of *Rps5* in *Arabidopsis* has also been observed (*33*).

Polymorphism data for five additional *R*-gene loci are summarized in Table 2. Considerable differences are seen in the average pairwise synonymous divergence between alleles. *Rpp13*, a single-copy locus in *A. thaliana* conferring resistance to *Peronospora parasitica*, shows over 9% divergence at synonymous sites in the LRR region (but not the rest of the gene). Similar divergence is seen at each of the three loci comprising *Rpp1*. Alleles at all four of these loci, therefore, are candidates for balanced polymorphism.

Moderate divergence is found among alleles of the Flax *L* locus and among alleles of the *Rpp8* locus in *Arabidopsis*. Selection for variation is suggested by the fact that both loci are segregating for many (functionally distinct in the case of *L*) alleles. *Rps4* is distinguished by its near absence of polymorphism between two resistance alleles and a susceptible allele. Only a single amino acid polymorphism (and no synonymous differences) is present in the LRR region, whereas six synonymous differences separate resistance and susceptibility alleles in the 5′ TIR and NBS domains. Ten additional amino acid replacement mutations are spread throughout the rest of the gene, but only two of these distinguish resistance and susceptibility alleles. The relative lack of divergence between functionally distinct alleles suggests that they have descended recently from a common ancestor, and this may be an indication of a recent selective sweep.

Overall, *R*-gene alleles show a wide range of ages, with some loci harboring old alleles that may be the product of balancing selection (*Rpm1*, *Rps2*, *Rpp1*, and *Rpp13*) and others showing more modest levels of divergence. Of these latter alleles, *Rpp8* and *L* segregate for a large number of alleles, a pattern inconsistent with an arms race. Allelic diversity at *Rps4* has not been surveyed. The polymorphism data, therefore, indicate that a simple arms race model involving repeated selective sweeps may apply to, at most, a small complement of *R*-loci.

Unusual Relation Between Polymorphism and Adaptive Divergence

As is evident in Table 2, allelic divergence (within species) in amino acid sequence can be considerable, especially in the LRR region, and this divergence can be associated with high K_a:K_s ratios between pairs of alleles. Thus, the general finding of adaptively driven divergence among paralogs is also applicable to variants segregating within a locus.

What is most revealing about the divergence of alleles from the perspective of evolutionary dynamics is that these adaptive variants coexist with other alleles. Overall, there is a strong tendency for loci whose alleles have the largest K_a:K_s ratio (i.e., the most rapid adaptive evolution) to have the youngest alleles, as indicated by the smaller synonymous divergence in the non-LRR regions of the alleles. One dramatic example is provided by the Flax *L* locus, where 13 alleles representing 12 different functional specificities differ one from another by an average of 40 amino acid replacements in domain 2 of the LRR. However, they differ little at synonymous sites (full LRR K_s^* = 0.029), which indicates that these alleles arose from common ancestral alleles in the relatively recent past. At the other extreme, *Rpp13* and *Rpp1* display less marked K_a:K_s ratios but appear to have considerably older alleles, with synonymous divergence (K_s for the entire LRR) among alleles of 9 to 18%, more reminiscent of the extraordinary age of MHC alleles in humans.

Mutational Mechanisms

The possibility of elevated rates of mutation in domain 2 of the LRR has been suggested (*8*, *19*), and if this was occurring, it would bias estimates of the age of alleles. We do not believe this to be the case. First, K_a and K_s calculations overestimate the rate of synonymous substitutions and underestimate the rate of nonsynonymous substitutions whenever K_a is greater than K_s (*34*). Second, the number of synonymous mutations in domain 2 is not significantly greater than that seen in the other domains (Table 2). Third, we find no evidence for

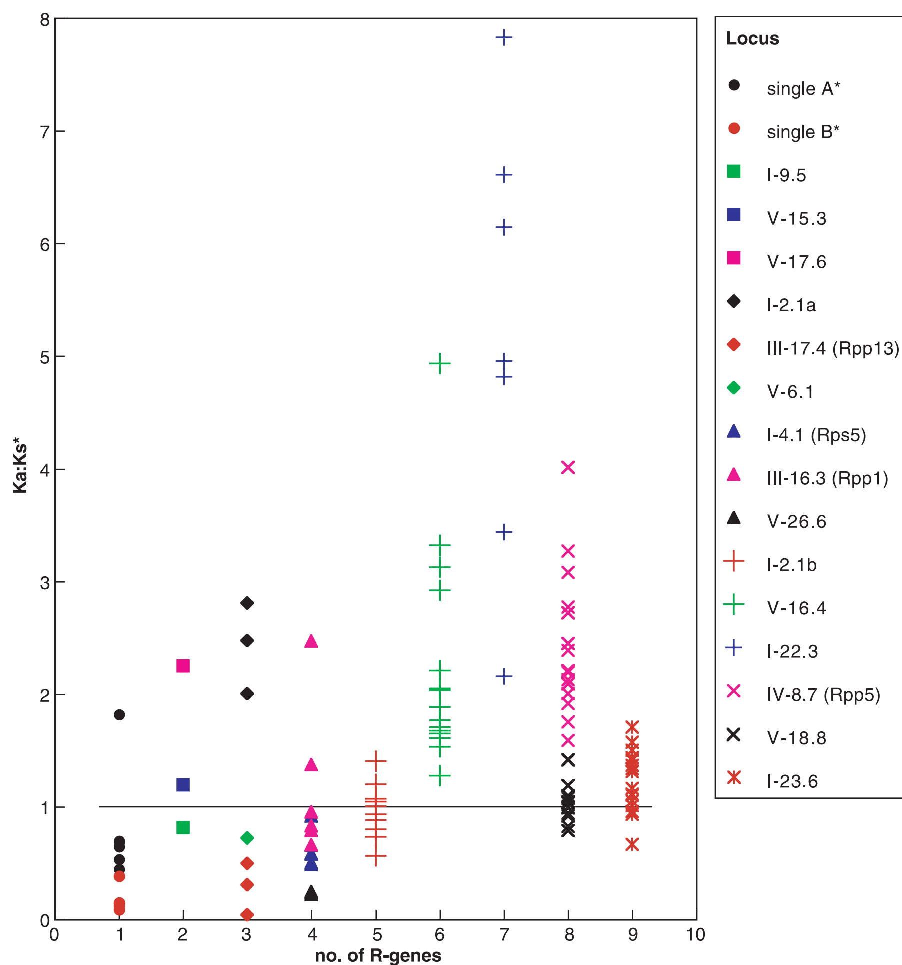

Fig. 1. Evolutionary analysis of *Arabidopsis* known and putative *R*-genes. The ratio K_a:K_s* is plotted against cluster size for 15 LRR-containing complex *R*-loci and two sets of related single *R*-genes (*28*). K_a:K_s*, calculated as in Table 2, is shown for all pairs of *R*-genes at a locus with 0.01 < K_s < 1. Complex loci are named by chromosome position in Mb (known *R*-locus if any). Cluster sizes for complex loci reflect numbers of *R*-genes that may be available for genetic exchange. The line indicates K_a:K_s = 1.

Table 2. Polymorphism among alleles of disease resistance genes. Numbers of silent and replacement (Rep.) changes and K_a:K_s ratios. LRR domain 2 corresponds to the framed region of the protein (*17*). Only exposed residues are included in the analyses of domains 1 to 3 (*19*), and domain 4 includes structural residues throughout the LRR. The rest of gene excludes the LRR. K_s*, which is K_s calculated for the entire LRR region, provides better estimates than that for each domain. Synonymous mutation rates do not vary significantly among domains. K_a:K_s calculations excluded codons mutated at all three positions and included multiple hits correction. Supplemental information is available at *Science* Online (*38*).

Gene	No. of alleles	LRR domain 1			LRR domain 2			LRR domain 3			LRR structure residues			Full LRR		The rest of gene	
		Silent	Rep.	K_a:K_s*	Silent	Rep.	K_a:K_s*	Silent	Rep.	K_a:K_s*	Silent	Rep.	K_a:K_s*	K_s*	K_a:K_s*	K_s	K_a:K_s*
Rpp13	3	0.67	3.67	0.63	4.03	22.9	2.29	3.83	26.2	2.00	5.67	15.3	1.25	0.092	1.60	0.045	0.439
Rpp1A	2	2	0	0	1	15	2.31	3.17	6.83	0.86	2	0	0.86	0.099	0.75	0.108	0.759
Rpp1B	2	3	7	0.13	2	26	2.94	2.17	6.83	0.58	5	2	0.18	0.094	0.83	0.071	0.816
Rpp1C	2	3	6	0.55	2	14	1.49	4	5	0.28	7.5	5.5	0.34	0.182	0.48	0.074	0.875
RPP8	6	0.87	4	0.86	3.1	18.2	3.54	1.7	12.8	1.06	2.2	1.37	0.29	0.035	1.50	0.032	0.668
RPH8B	6	1.4	3.2	0.83	2.64	20.8	4.07	1.4	12.7	1.34	1.13	4.67	1.07	0.029	2.25	0.044	0.423
rpp8	6	1.66	3.05	0.43	3.22	26.2	3.51	4.43	17.9	0.96	2.68	3.29	0.40	0.053	1.34	0.038	0.727
Flax *L*	13	1.64	3.35	0.48	3.14	40.6	4.66	5.7	24.5	1.34	3.32	7.03	0.72	0.029	1.82	0.025	0.988
Rps2	7	0.47	0	0	1.42	1.33	0.50	0.48	2.67	0.50	1.43	0.48	0.13	0.016	0.27	0.013	0.066
Rps4	3	0	0	–	0	0	–	0	0.33	–	0	0	–	0	–	0.007	0.140

elevated rates of divergence between species in the LRR domains. It is therefore likely that excess synonymous mutations in the LRR indicate a relatively greater genealogical antiquity of this region. This point is reinforced by the data for *Rpm1* (*3*) and *Rps5*, for which the high mutational divergence is centered on the sequences flanking the insertion/deletion site rather than on the LRR.

Evolutionary Mechanisms

The most striking feature to emerge in the available data is the similarity in the patterns of evolved differences seen among alleles at individual loci and between genes belonging to evolutionary clusters. In particular, we are struck by the presence of *R*-gene alleles and paralogs, representing a very wide range of evolutionary ages, undergoing rapid adaptive evolution. Furthermore, rates of adaptive evolution appear greatest between closely related *R*-genes, suggesting that genetic exchange has contributed to the production of new adaptive alleles. Clearly, selection plays a profound role in *R*-gene dynamics. However, a classic arms race involving a succession of adaptive variants may be a poor metaphor for *R*-gene dynamics because alleles are not young and loci are not monomorphic, as predicted by this model. This raises the question of how polymorphisms are maintained in the face of adaptive evolution, an issue that has received attention with respect to MHC in a series of reviews (*35*) that explore the potential roles of frequency-dependent selection and overdominance. *Arabidopsis*'s high selfing rate (*36*) suggests that, at least for this species, overdominance cannot be a potent evolutionary force. Instead, it is likely that frequency-dependent selection favoring novel alleles when rare is responsible for the pattern we see at *R*-gene loci.

The spread of a novel resistance allele in the host plant should open the door for pathogens carrying an ancestral *Avr* gene to outbreak, because this novel *R*-allele would have reduced the frequency of the ancestral *R*-allele in populations. Such frequency-dependent selection requires persistence of pathogens carrying the ancestral *Avr* gene in a refuge or isolated population as a novel resistance allele spreads. This situation may lead to cycling (*3*, *37*). Alternatively, fixation may occur if the pathogen is extremely virulent, the pathogen refuge is not sufficiently effective, or environmental stochasticity is strong. Demographic and ecological details are thus likely to influence the outcome of these interactions and may explain the observed variation in the ages of alleles among *R*-genes. The current picture of ubiquitous polymorphism may not be general, however, because the existing data are biased by the use of polymorphism to identify and clone these genes. These issues are likely to be quickly resolved as additional molecular population genetic and evolutionary data become available. What will then be needed most is ecological work to better understand short-term disease dynamics and theoretical work, as there is an almost complete absence of models exploring the age of alleles under different scenarios of adaptive evolution.

References and Notes

1. D. A. Lawlor, F. E. Ward, P. D. Ennis, A. P. Jackson, P. Parham, *Nature* **335**, 268 (1988).
2. W. E. Mayer *et al.*, *EMBO J.* **7**, 2765 (1988).
3. E. A. Stahl, G. Dwyer, R. Mauricio, M. Kreitman, J. Bergelson, *Nature* **400**, 667 (1999).
4. A. G. Clark, L. Wang, *Genetics* **147**, 713 (1997).
5. N. D. Young, *Curr. Opin. Plant Biol.* **3**, 285 (2000).
6. L. Van Valen, *Evol. Theory* **1**, 1 (1971).
7. K. Clay, P. X. Kover, *Annu. Rev. Phytopathol.* **34**, 29 (1996).
8. M. Parniske *et al.*, *Cell* **91**, 821 (1997).
9. M. A. Botella *et al.*, *Plant Cell* **10**, 1847 (1998).
10. J. M. McDowell *et al.*, *Plant Cell* **10**, 1861 (1998).
11. B. C. Meyers, K. A. Shen, R. Rohani, B. S. Gaut, R. W. Michelmore, *Plant Cell* **10**, 1833 (1998).
12. L. Noel *et al.*, *Plant Cell* **11**, 2099 (1999).
13. H. H. Flor, *Adv. Genet.* **8**, 29 (1956).
14. A. L. Hughes, M. Nei, *Nature* **335**, 167 (1988).
15. ———, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 958 (1989).
16. J. Ellis, P. Dodds, T. Pryor, *Curr. Opin. Plant Biol.* **3**, 278 (2000).
17. D. A. Jones, J. D. G. Jones, *Adv. Bot. Res. Adv. Plant Pathol.* **24**, 89 (1997).
18. R. T. Leister, F. Katagiri, *Plant J.* **22**, 345 (2000).
19. P. D. Bittner-Eddy, I. R. Crute, E. B. Holub, J. L. Beynon, *Plant J.* **21**, 177 (2000).
20. J. G. Ellis, G. J. Lawrence, J. E. Luck, P. N. Dodds, *Plant Cell* **11**, 495 (1999).
21. J. E. Luck, G. J. Lawrence, P. N. Dodds, K. W. Shepherd, J. G. Ellis, *Plant Cell* **12**, 1367 (2000).
22. D. Banerjee, X. Zhang, A. F. Bent, *Genetics* **158**, 439 (2001).
23. G. L. Wang *et al.*, *Plant Cell* **10**, 765 (1998).
24. M. A. Sudupak, J. L. Bennetzen, S. H. Hulbert, *Genetics* **133**, 119 (1993).
25. S. H. Hulbert, *Annu. Rev. Phytopathol.* **35**, 293 (1997).
26. E. A. Stahl, J. G. Bishop, *Curr. Opin. Plant Biol.* **3**, 299 (2000).
27. The *Arabidopsis* Genome Initiative, *Nature* **408**, 796 (2000).
28. We analyzed the 182 *A. thaliana* NBS-containing *R*-gene homologs in the NIBLRRS database (http://niblrrs.ucdavis.edu) and defined clusters of evolutionarily related *R*-genes on the basis of position and local DNA sequence similarity (fasta expectation < 0.1). Cluster sizes (frequencies) are 2 (13), 3 (6), 4 (4), 5 (1), 6 (1), 7 (2), 8 (2), 9 (1), 11 (1), and 12 (1). Of these complex clusters, 20 had at least 50% nucleotide similarity and were thus sufficiently related for statistically reliable DNA sequence analysis. Four sets of single *R*-genes also met the criterion of 50% nucleotide similarity and were more closely related to each other than to any clustered *R*-genes. HMMer Pfam analyses defined LRR regions and LRR domains (see Table 2) in 15 of the complex *R*-loci and 2 of the single *R*-gene sets. Supplemental information is available at *Science* Online (*38*).
29. R. W. Michelmore, B. C. Meyers, *Genome Res.* **8**, 1113 (1998).
30. R. R. Hudson, N. L. Kaplan, *Genetics* **120**, 831 (1988).
31. A. L. Caicedo, B. A. Schaal, B. N. Kunkel, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 302 (1999).

32. M. R. Grant *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15843 (1998).
33. Unpublished data of authors (*39*), comparing about 2 kb of flanking sequence in 23 *A. thaliana* accessions.
34. To calculate changes in codons with more than one position substituted, current algorithms for calculat ing K_s and K_a assume equal rates of synonymous and replacement changes. This assumption leads to the overestimation of synonymous changes and underestimation of replacement changes when the replacement rate actually exceeds the synonymous rate (J. Comeron, M. Kreitman, unpublished results).
35. W.-H. Li, *Molecular Evolution* (Sinauer Associates, Sunderland, MA, 1997).
36. R. J. Abbott, M. F. Gomes, *Heredity* **62**, 411 (1989).
37. R. M. May, R. M. Anderson, *Proc. R. Soc. London B* **219**, 281 (1983).
38. Supplemental material is available on *Science* Online at www.sciencemag.org/cgi/content/full/292/5525/2281/DC1.
39. J. Bergelson, M. Kreitman, E. A. Stahl, D. Tian, unpublished data.
40. R. Mauricio, unpublished data.
41. We gratefully acknowledge support from the NIH to J.B. and M.K., a Packard Fellowship to J.B., and a Sloan/Department of Energy fellowship to E.A.S. K. Thornton provided computer programs.

REVIEW

Common and Contrasting Themes of Plant and Animal Diseases

Brian J. Staskawicz,[1]* Mary Beth Mudgett,[1] Jeffrey L. Dangl,[2] Jorge E. Galan[3]

Recent studies in bacterial pathogenesis reveal common and contrasting mechanisms of pathogen virulence and host resistance in plant and animal diseases. This review presents recent developments in the study of plant and animal pathogenesis, with respect to bacterial colonization and the delivery of effector proteins to the host. Furthermore, host defense responses in both plants and animals are discussed in relation to mechanisms of pathogen recognition and defense signaling. Future studies will greatly add to our understanding of the molecular events defining host-pathogen interactions.

The ability of pathogenic microorganisms to harm both animal and plant hosts has been documented since the initial demonstration in the 1870s that microbes were causal agents of disease. Since the initial discoveries by Koch (*1*) that *Bacillus anthracis* caused anthrax and by Burrill (*2*) that *Erwinia amylovora* caused fire blight in pears, our knowledge base has expanded enormously. Today, the genomes of the most important animal and plant pathogens have been or will be sequenced, and the molecular basis of pathogenicity is beginning to be deciphered. Furthermore, several model plant and animal host genomes are fully sequenced. Basic discoveries made in the post-genomic era will fuel our quest for developing new strategies for disease control.

In the past 5 years, pivotal observations have revealed that bacterial pathogens share common strategies to infect and colonize plant and animal hosts. One is the ability to deliver effector proteins into their respective host cells to mimic, suppress, or modulate host defense signaling pathways and to enhance pathogen fitness. On the host side, plants and animals have evolved sophisticated surveillance mechanisms to recognize various bacterial pathogens. Interestingly, plants recognize distinct effectors from pathogenic bacteria, whereas animals recognize conserved "molecular patterns," such as those derived from lipopolysaccharide (LPS) or peptidoglycan. The discovery that surveillance proteins in diverse hosts share common protein signatures that perform similar functions invites speculation as to how these resistance mechanisms evolved.

This rapidly expanding field is obviously a large topic, and all aspects cannot be considered here. We will mainly focus on themes of plant pathogenesis. However, when appropriate, we will discuss unique and shared strategies used by microbial pathogens to infect animal hosts. Moreover, we will compare our emerging knowledge of pathogen surveillance mechanisms used by plant and animal hosts.

[1]Department of Plant and Microbial Biology, University of California at Berkeley, Berkeley, CA 94720, USA. [2]Department of Biology and Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, NC 27599, USA. [3]Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, CT 06536, USA.

*To whom correspondence should addressed. E-mail: stask@nature.berkeley.edu

Initial Interactions of Bacterial Pathogens with Host Plant Cells

The initial interactions of bacteria with their plant hosts are critical in determining the final outcome of infection. Curiously, the majority of infections do not lead to overt disease. Microorganisms are usually repelled by plant defense mechanisms. However, in some cases, interactions of microbial pathogens with their host lead to the overt harm to the presumptive host and to pathology. Factors intrinsic to both the pathogen and the plant determine the final outcome of the encounter.

The epiphytic (saprophytic) life stage of phytopathogenic bacteria often precedes entry into the host plant and the onset of pathogenicity. For example, phytopathogenic bacteria in the genera *Pseudomonas* and *Xanthomonas* can colonize leaf surfaces of plants and reach dense bacterial populations (10^7 colony-forming units per square centimeter) without causing disease. Under the appropriate environmental conditions, bacteria enter leaf mesophyll tissue through natural stomatal openings, hydothodes, or wounds, thus making their first contact with internal host cells. Phytopathogenic bacteria multiply in the intercellular spaces (apoplast) of plant cells and remain extracellular. This is in contrast to many animal bacterial pathogens that gain entry into their host cells and then multiply intracellularly.

Mechanism for Plant Cell Infection: Conservation of Type III Secretion System

To grow in the apoplast, phytopathogenic bacteria sense their environment and induce genes required for host infection. A primary locus induced in Gram-negative phytopathogenic bacteria during this phase is the *Hrp* locus (*3*). The *Hrp* locus is composed of a cluster of genes that encodes the bacterial type III machinery that is involved in the secretion and translocation of effector proteins to the plant cell. Mutations in *Hrp* genes affect both the induction of localized plant disease resistance (the hypersensitive response) and bacterial pathogenicity. This mutant phenotype and the subsequent demonstration that Hrp structural proteins and type III effectors are transcriptionally coregulated (*4*) provided important evidence that effectors not only caused disease but were the components of the pathogen recognized by the host. Although the general physiology of low sugar and low pH in the apoplast is known to induce the assembly of the Hrp type III apparatus in phytopathogenic bacteria, a specific plant-derived signal has yet to be identified.

The demonstration that phytopathogenic bacteria use genes (i.e., *Hrp*) that are remarkably similar to genes encoding the type III secretion system in animal pathogenic bacteria provided an immediate conceptual framework to explain the molecular mechanisms