Molecular evolution of plant immune system genes

Peter Tiffin and David A. Moeller

Department of Plant Biology, 250 Biosciences, University of Minnesota St Paul, MN 55108, USA

Molecular population genetic studies are providing new perspectives on the evolution of genes that confer resistance to pathogens and herbivores. Here, we compare the evolutionary history of different components of the defense response (detection, signaling and response) and of genes with parallel function in plants and *Drosophila*. A review of the literature indicates that the dominant form of selection acting on defense genes (balancing, positive and purifying) differs among components of defense. Sampling of particular classes of genes and genes from non-model organisms, however, remains limited. Future studies combining molecular evolutionary analyses with ecological genetic and functional analyses should better reveal how natural selection has shaped defense gene evolution.

Introduction

Herbivores and pathogens can have strong effects on plant fitness, regulate plant population sizes, and cause considerable economic damage in managed ecosystems [1]. Understanding the evolution of defense-related traits, including a diverse array of morphological structures, physiological responses, secondary metabolites, RNAs and proteins, provides insight into the role of herbivore- and pathogen-imposed selection on plants. This information could help to guide the development of durably resistant crop varieties and more sustainable pest management strategies.

Our knowledge of the evolution of plant defense traits comes primarily from ecological genetic investigations, which examine the process of natural selection and the genetic architecture of traits in contemporary populations. Whereas ecological genetic approaches provide insight into short-term evolution, molecular population genetics (see Glossary) provides a means for examining the long-term evolution of genes that contribute to defense and other ecologically important traits [2,3] (Box 1). Although these approaches are most often taken in isolation from one another, recent work has highlighted the potential for integrating molecular and field studies to understand the role of natural selection in defense gene evolution.

Here we review recent studies that have characterized the evolutionary history of plant defense genes. Because the interpretation of molecular evolutionary analyses is strengthened by knowledge of gene function, most studies have focused on components of defense that are genetically

best characterized. In particular, analyses have focused on genes involved in pathogen detection and the initiation of a defense response, and genes encoding protein-based defenses that are part of that response, primarily from *Arabidopsis thaliana* and close relatives of maize (Box 2). In an effort to identify broad patterns in immune system evolution, we take a comparative approach to reviewing past studies. In particular, we compare the evolution of defense-related genes that differ in putative function, and also compare the evolutionary histories of plant defense genes with genes having parallel functions in *Drosophila*.

Detection genes

The vast majority of plant proteins involved in detection are encoded by NBS-LRR (nucleotide binding site-leucine rich repeat) resistance genes (R-genes), most of which are members of large multigene families [4]. Analyses of the rate at which nonsynonymous and synonymous mutations accumulate (dN/dS; Box 3) in R-gene family members have revealed evidence for positive selection having driven their divergence in all taxa for which tests have been conducted, including *A. thaliana* [5], *Lactuca sativa* [6], *Linum usitatissimum* (flax) [7], *Oryza sativa* [8] and the Solanaceae [9]. However, positive selection might not be a driving force in all R-gene subfamilies and certainly cannot

---

Glossary

Balancing selection: A form of natural selection that maintains genetic variation at an individual locus. Unlike positive selection, balancing selection increases levels of nucleotide variation at linked sites. A variety of mechanisms can result in elevated levels of polymorphism including over-dominance (heterozygote advantage), frequency-dependent selection, and temporally or spatially variable selection.

Ecological genetics: A branch of evolutionary biology focused on understanding the process of adaptive evolution through the study of natural selection and inheritance in wild populations. Ecological genetic studies employ direct observation or experimentation in present-day populations and focus on ecologically important traits that could influence lifetime fitness.

Molecular population genetics: A branch of evolutionary biology focused on understanding the evolutionary and demographic processes that shape genetic variation at the nucleotide level. Since its inception, there has been considerable focus on identifying genes that have been the targets of natural selection and the form of selection shaping nucleotide variation.

Positive selection: A form of natural selection that increases the frequency of alleles carrying an advantageous mutation. Positive selection reduces nucleotide variation at nonsynonymous and synonymous sites linked to the selected site. When positive selection drives a single allele to fixation and nucleotide variation is eliminated, this is referred to as a 'selective sweep'. Positive selection is also commonly referred to as 'directional selection'.

Purifying selection: A form of natural selection that eliminates or reduces the frequency of deleterious mutations. Purifying selection decreases nonsynonymous nucleotide variation and synonymous variation at linked sites. This type of selection is also referred to as 'background selection' or 'negative selection'.

Corresponding author: Tiffin, P. (ptiffin@umn.edu)
Available online 2 October 2006.

Box 1. Molecular evolutionary versus ecological genetic analyses

Our understanding of how herbivores and pathogens shape host evolution comes largely from ecological genetic studies in contemporary populations. These approaches are powerful for identifying patterns of selection acting on defense traits and genetic constraints that limit an evolutionary response to selection [68]. Ecological genetic approaches, however, are limited to providing insight into short-term evolution. By contrast, molecular population genetic approaches provide insight into evolutionary processes that operated in the distant past. Because the effects of selection on DNA sequence polymorphism can accumulate across many generations, molecular evolutionary analyses can detect evidence of selection even when selection has been too weak or intermittent to identify using ecological genetic approaches.

Molecular population genetic and ecological genetic approaches also differ in their ability to identify agents of selection. Experimental manipulation can be used to test directly the role of specific enemies or physiological costs in shaping the evolution of defense or other traits of ecological interest (see e.g. ref. [57]). Molecular analyses provide no means for identifying selective agents but can lead to evolutionary hypotheses that can be tested using manipulative approaches (see e.g. ref. [19]). In some cases, however, it can be difficult to test hypotheses developed to explain molecular data. For example, the evolution of defense might cause changes in pathogen and herbivore host range or performance; past selective agents might be no longer affected by a specific defense or even found on a focal host species. Similarly, pathogen and herbivore populations might fluctuate dramatically through time resulting in infrequent episodes of strong selection that are unlikely to be captured during short-term experiments. Because evolution in response to positive selection is likely to alter herbivore and pathogen pressure and community composition, experimental investigations of the selective forces responsible for a molecular signature of positive selection can be particularly challenging. Linking molecular and ecological approaches might be easier when examining balanced polymorphisms, which can be associated with more stable interactions.

A third difference is that the interpretation of molecular evolutionary analyses is strengthened by knowledge of gene function, whereas ecological genetic approaches examine phenotypic variation and require no knowledge about the molecular genetic basis of trait variation. For this reason, ecological genetic analyses are often conducted in non-model systems and on traits with complex genetic architecture; whereas the majority of molecular evolutionary analyses have focussed on model systems, primarily close relatives of *Arabidopsis thaliana* and *Zea mays* ssp. *mays* [69]. The increasing availability of genomic resources should enable molecular evolutionary analyses to be applied more commonly to non-model organisms.

have acted on all loci; powerful codon-based tests [10] have revealed evidence for positive selection having contributed to the divergence of members of only 13 of 22 NBS-LRR subfamilies in *A. thaliana* [11].

Investigations of intralocus nucleotide diversity also reveal evidence for past selection, but the signature is most frequently one of balancing selection. Six *A. thaliana* R-genes (*RPM1* [12], *RPS2* [13,14], *RPS5* [15], and three additional presence/absence loci [16] (*RPM1* and *RPS5* are also presence/absence loci, i.e. loci segregating full-length and deleted alleles) harbor an excess of polymorphism associated with two distinct haplotypes, consistent with selectively maintained balanced polymorphisms. For the three loci with assigned functions, *RPM1*, *RPS2* and *RPS5*, the two haplotypes confer resistance or susceptibility to the generalist pathogen *Pseudomonas syringae*. Selection also

seems to be the cause of an excess of amino acid replacement diversity, although not two distinct allelic classes, at a seventh *A. thaliana* R-gene, *RPP13*. The excess diversity at *RPP13* is probably the result of tradeoffs in the ability of genotypes to confer resistance to different genotypes of the pathogen *Peronospora parasitica* [17]. Excess amino acid replacement polymorphism is also found at the *Lycopersicon peruvianum* gene, *Pto* (L.E. Rose *et al.*, unpublished), and the *Solanum pimpinellifolium* gene, *Cf-2* [18], which confer resistance to the bacterial pathogen, *P. syringae* pv. *tomato* and strains of the fungus *Cladosporium fulvum*, respectively. The interpretation of the *Solanum* data, however, is complicated by the duplication and deletion of LRR-coding units and sequence exchange between genes.

An important caveat in interpreting patterns of nucleotide diversity is that population structure or demographic history, including recent population bottlenecks or population expansion, can produce patterns of variation similar to those caused by balancing or positive selection. For this reason, interpretation of molecular evolutionary analyses is often speculative and the case for selection is much stronger if coupled with functional analyses in contemporary populations. To test whether selection is responsible for the maintenance of the balanced polymorphism at *RPM1*, as suggested by molecular data [12], Tian *et al.* [19] examined the fitness of plants homozygous for either *RPM1* susceptible or resistance alleles in a replicated field experiment. This experiment revealed that plants carrying *RPM1* resistance alleles had 9% lower fitness than plants carrying susceptible alleles when grown in the absence of pathogens. These results bolster the interpretation of the molecular data and provide empirical support for theoretical models predicting that fitness costs can result in the long-term maintenance of genetic variation for defense traits [20]. From an evolutionary perspective, this experiment is exemplary because the researchers examined the fitness consequences of different alleles in the field and not just biochemical or physiological function.

Reconciling results from analyses of interlocus divergence and intralocus diversity

How can results from analyses based on intralocus diversity, which reveal evidence for selectively maintained polymorphisms at R-loci, be reconciled with results from analyses of interlocus divergence, which suggest positive selection has been a driving force? One possibility is that some NBS-LRR proteins activate defense after direct interactions with pathogen effector proteins whereas others seem to activate defense after modification of proteins 'guarded' by NBS-LRR R-proteins [21]. Dodds *et al.* [22] recently suggested that ancient polymorphisms might be maintained only at R-loci encoding proteins that are activated indirectly (e.g. *RPM1*, *RPS2* and *RPM5*), whereas selection results in excess amino acid polymorphisms, but not distinct allelic classes at R-proteins activated by direct interaction with pathogen elicitors (e.g. RPP13). They also present the intriguing suggestion that different evolutionary histories and modes of action are associated with

---

Box 2. Detection and signaling pathways in *Arabidopsis* and *Drosophila*

Although the genetic basis of plant defenses is complex, many genes involved in detection of enemy attack, signaling and the immune response have been identified and characterized, particularly in the model system *A. thaliana*. Enemy detection can occur through direct interactions between enemy elicitors and transmembrane or cytoplasmic NBS-LRR R-proteins. Many NBS-LRR proteins, however, seem to trigger immune responses after pathogen elicitors alter host proteins that are 'guarded' by R-proteins [21]. Following detection, different R-proteins trigger immune responses through one of several different signaling pathways. One signaling pathway activated by direct binding of pathogen molecules by membrane-bound NBS-LRR involves a mitogen-activated protein (MAP) kinase cascade and activation of a WRKY transcription factor [70] (Figure Ia). A second signaling pathway activated by some cytoplasmic R-proteins involves EDS1, PAD4, salicylic acid (SA), NPR1 and WRKY proteins [71] (Figure Ia). The details of these and other R-protein-mediated signaling cascades are still being actively investigated. Plant defenses can also be induced by the products generated when plant enzymes (e.g. β-1,3-glucanase [42]) degrade fungal cell walls, in addition to ethylene and jasmonic acid signaling cascades after attack by herbivores and some pathogens [72]. Regardless of the signaling pathway, detection of infection or attack can result in localized cell death, release of reactive oxygenase species, and upregulation of pathogenesis-related (PR) proteins, including chitinases, β-1,3-glucanases, protease inhibitors and lipid transfer proteins, many of which have biocidal activity [73]. PR and other immunity proteins are also upregulated as part of the basal defense response, in response to abiotic stress, and are expressed constitutively in some tissues [73].

As shown in the simplified schematics in Figure Ia, the general structure of pathways involved in pathogen detection, signaling and response in plants is similar to the Toll- and Imd-mediated immune response in *Drosophila* [29] (Figure Ib). The similarities between *Arabidopsis* and *Drosophila* signaling seem to be due to convergent evolution rather than because the pathways have a common origin [74].
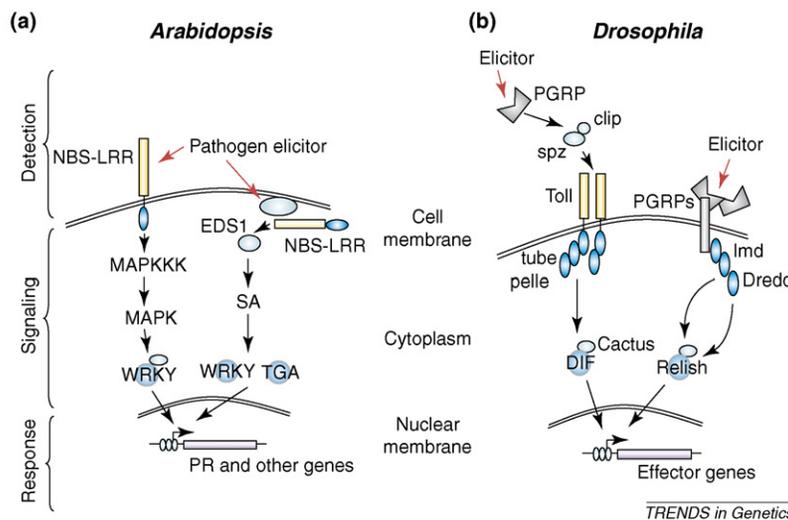

Figure I. Overview of detection and signaling pathways that result in the upregulation of defense-response genes in **(a)** *Arabidopsis* and **(b)** *Drosophila*. For further details of these pathways, see Refs [21,29,71,72].

---

different classes of pathogens; they hypothesize that indirectly activated R-proteins protect against generalist pathogens whereas directly activated R-proteins defend hosts against specialist pathogens.

A second possible reason that intralocus and interlocus analyses might reveal different selective histories is that the two approaches have been applied to R-genes in different genomic environments. The majority of R-genes are located in large multigene clusters [23,24], but the majority of the R-genes subjected to intralocus analyses are located either away from others or in small gene clusters [16,23]. Isolated R-genes or those in small clusters might have different evolutionary dynamics than genes in large complex clusters [3,5], perhaps because they are less susceptible to rapid turnover through gene conversion and deletion. More extensive sampling of intralocus diversity, particularly for genes located in large clusters, is needed to test this idea. Sampling diversity at R-gene clusters could be difficult, however, given potentially rapid rates of contraction, expansion, rearrangement and gene conversion that can occur at these complex loci [18,25].

A third possibility is that interlocus analyses are detecting selective events older than those detected by intralocus analyses. A newly duplicated R-gene can be maintained within a genome only if it rapidly evolves a novel and advantageous function (e.g. an ability to recognize a previously unrecognized pathogen). Positive selection will then drive the rapid increase in the frequency of the allele that confers this new function. Once a new function is established, however, the pattern of selection could shift and selection could maintain distinct haplotypes with different fitness costs or detection specificities. Both positive selection driving divergence after duplication and selection maintaining allelic polymorphism would increase genomic diversity. Diversity in defense genes is predicted to be selectively advantageous for hosts that are attacked by diverse and rapidly evolving pathogens [26].

Finally, it is possible that the discrepancy in results could arise because balancing selection maintains distinct allelic classes at a locus over long time periods, whereas intermittent bouts of positive selection favor new variants within one or both allelic classes. For example, resistant and susceptible alleles could be maintained as a result of

---

Box 3. Some methods for inferring natural selection from DNA sequence data

A more comprehensive review of these and other analyses can be found in Nielsen [75]. The application of molecular population genetics to the study of adaptive evolution in plants is discussed in Wright and Gaut [69].

Tests based on comparative data
These use single orthologous sequences from each of two or more related species or paralogous sequences from duplicated loci. Comparative tests examine whether positive selection has contributed to evolutionary divergence but cannot detect selection that maintains polymorphism.

dN/dS tests
In the absence of selection, substitutions are expected to accumulate at amino acid replacement and synonymous sites at equal rates ($dN/dS = 1$). Under purifying (negative) selection $dN/dS < 1$, whereas under positive selection $dN/dS > 1$. When applied to an entire gene, this is a conservative test because a large number of codons must be affected by positive selection in order for dN/dS to exceed 1. Recently developed likelihood-based analyses account for variation in dN/dS among sites and identify specific amino acids that have been the targets of selection [10]. These codon-based tests are more powerful than traditional dN/dS tests, but power depends upon the depth of phylogenetic sampling and the consistency of selection. Likelihood approaches can also be used to test for variation in selection among evolutionary lineages.

Tests based on patterns of intra-specific variation
These rely upon a sample of sequences from within a population (or species). These can be used in combination with information on levels of divergence from orthologous sequences in related taxa.

Frequency distribution of polymorphisms within a population
This set of tests is based on comparing the frequency of mutations in a sample with that expected under a standard neutral model, which assumes random mating, constant population size and no population structure (e.g. Tajima's *D* [76], Fu and Li's *D* [77]). Deviations from neutral expectations can indicate the operation of natural selection. For example, a population sample is expected to contain an excess of rare mutations following a recent selective sweep or under purifying selection, an excess of high-frequency mutations under positive selection, and an excess of mutations segregating at intermediate frequencies under balancing selection. Several modifications of these tests use an outgroup sequence to take into account whether mutations are derived in the focal taxon (e.g. Fay and Wu's *H* [78]). Interpretation of the results of these analyses is greatly complicated by the fact that demographic

history and selection can produce similar patterns of polymorphism. Comparing patterns of polymorphism at genes of interest to reference loci is one approach used to lessen problems associated with demography.

Intra-specific polymorphism and inter-specific divergence
Two of the most commonly used tests for non-neutral evolution examine the relationship between levels of intra-specific variation and inter-specific divergence; under a neutral model both intra-specific variation and inter-specific divergence are functions of mutation rate and expected to be positively correlated. The Hudson–Kreitman–Aguade (HKA) [79] test requires multiple loci to examine whether the ratio of polymorphism to divergence varies among genes to a greater extent than expected under a neutral equilibrium model. The MacDonald–Kreitman (MK) test extends the HKA test by separating mutations based on whether they are nonsynonymous versus synonymous [80]. Positive selection is expected to increase evolutionary rates at nonsynonymous (replacement) sites compared with synonymous sites resulting in an increase in the ratio of inter-specific fixed differences relative to intra-specific polymorphism at nonsynonymous compared with synonymous sites.

Linkage disequilibrium
Natural selection also influences levels of linkage disequilibrium (LD) and haplotype structure. For example, strong episodes of positive selection are expected to increase LD and reduce haplotype diversity. Much like tests based on the frequency spectrum of polymorphisms, inferences from tests based on LD and haplotype structure [81,82] are complicated by demography, and population structure can mimic the effects of natural selection. In addition, the signature of selection detected by these tests can be erased quickly by mutation and recombination [83].

Empirically derived distributions of population genetic parameters
Because assumptions of the neutral model are often violated in nature, tests based on both intra-specific diversity and inter-specific divergence can suffer from an inability to distinguish the effects of selection from population subdivision or demographic history (e.g. population bottlenecks). New approaches that should reduce this problem involve comparing test statistics from individual loci with empirically derived distributions based on genome-wide surveys of nucleotide polymorphism [45,84]. In these approaches, loci in the tails of the distributions are candidates for having evolved in response to selection. The power of these approaches is twofold: they avoid reliance on the assumptions of the neutral equilibrium model and they enable identification of targets of selection even when there is no *a priori* reason to investigate them.

---

fitness costs of resistance or virulence, repeated epidemics and fluctuating population sizes, or metapopulation dynamics [20,27,28]. At the same time, selection could favor new alleles within the resistant class after pathogens evolve counter-adaptations that enable them to evade detection. Although this scenario has not been modelled formally, one might expect that it would result in intralocus patterns of diversity consistent with a balanced polymorphism and interlocus patterns consistent with positive selection.

Detection in Drosophila
Insect humoral innate immune responses (Box 2) are activated after the detection of invading microorganisms by peptidoglycan recognition proteins (PGRPs) [29]. Analyses of intra-specific diversity at PGRP-encoding genes from *Drosophila melanogaster* and *D. simulans* have revealed no evidence for balancing or positive selection [30,31]. By contrast, two *D. simulans* genes, *Toll* and *Tehao*

[31,32], encoding transmembrane receptors that activate intracellular signaling and an immune response after binding host-derived ligands, have patterns of diversity consistent with positive selection; both genes have an apparent excess of fixed nonsynonymous differences that differentiate *D. simulans* from *D. melanogaster* sequences – a significant MacDonald–Kreitman (MK) test [31] (see Box 3). Moreover, *Tehao* in North American populations of *D. simulans* exhibits extremely great levels of linkage disequilibrium, possibly as a result of adaptation to the novel, non-African environment [32].

Evidence for positive selection on *Drosophila* transmembrane receptors suggests that pathogens have evolved mechanisms that interfere with receptor binding of host-derived ligands. By contrast, the apparent history of purifying selection on *Drosophila* PGRPs suggests that pathogens have either not evolved suppressors of these molecules, or if they have, the host proteins have not evolved in response to the presence of these suppressors.

Interestingly, the majority of plant detection genes whose intra-specific diversity has been investigated seem to have been the target of balancing selection, unlike *Drosophila* detection genes.

Intracellular signaling genes

In both plants and insects, pathogen detection leads to the initiation of a signal cascade that culminates in a defense response (Box 2). Several plant genes involved in these cascades have been identified [33], but none has been subject to molecular evolutionary analysis. Interestingly, some insect defense signaling genes show strong evidence of having evolved in response to positive selection. For example, elevated levels of amino acid fixation relative to polymorphism in the *D. simulans* transcription factor gene *Relish* [31,34], and elevated dN/dS ratios among *Relish* orthologs in Australian termite species [35] are both indicative of positive selection. Although genes involved in a signaling pathway are often thought to be evolutionarily conserved, these results suggest that pathogens might evolve increased virulence through mechanisms that disrupt signaling downstream of detection; that is, pathogens might counter host defenses by killing the messenger rather than evading surveillance or circumventing immune responses. In fact, several pathogen proteins that suppress host immune responses by disrupting signaling have been identified [36]. Studies examining the evolution of plant signaling genes are needed to determine if these are often the targets of selection, as they seem to be in some insect species.

Genes encoding proteins that inhibit pathogen and herbivore growth

The other major class of plant defense genes that have been subject to molecular evolutionary analyses are those encoding proteins that can directly inhibit enemy growth and fitness. To date, inter-specific divergence of chitinases in *Arabis* [37] and Poaceae taxa [38], β-1,3-endoglucanases in *Glycine* and a set of other dicots [39], and polygalacturonase inhibitor proteins (PGIPs) among legumes [40] and other dicots [41] have been investigated. These analyses have used codon-specific tests of dN/dS and all have detected evidence of positive selection having driven the evolution of some codons in one or more lineages. The case for selection having driven the divergence of the PGIPs was bolstered by a comparison of results from mutagenesis experiments and molecular evolutionary analyses. In particular, codons that mutagenesis experiments identified to affect the ability of a PGIP to bind pathogen ligands were identified by codon-specific tests as targets of selection [40]. Because, β-1,3-endoglucanases, PGIPs and possibly chitinases also have a role in pathogen detection [42,43], the apparent selection that has acted on these genes could be due to their role in detection rather than inhibition of pathogen growth.

Most of the genes encoding proteins directly involved in inhibiting growth and fitness seem to be single-copy or members of small gene families, and the divergence of gene family members within a genome has not been investigated. An exception is the *A. thaliana* defensin gene family, which comprises >300 members, many located in tandemly duplicated arrays. Positive selection seems to have had a role in the divergence of some defensin gene family members [44]. The similarities between the genomic structure of NBS-LRR and defensin gene families could offer an opportunity to investigate the degree to which the evolutionary dynamics of R-genes are affected by their function in detection versus gene family size and genomic arrangement.

Unlike analyses of intralocus nucleotide diversity at detection genes, the majority of which have produced evidence for selectively maintained polymorphism, only seven of >40 intralocus analyses of genes encoding proteins that directly inhibit infection or growth have detected

Table 1. Selective history inferred from patterns of intralocus nucleotide diversity segregating at plant immunity genes involved in detection and inhibition of enemy infection and growth

Defense type and taxon	Molecular signature of past selection					Refs
	Balancing	Weak balancing	Purifying or neutral	Weak positive	Positive	
Detection						
Arabidopsis thaliana	RPM1, RPS2, RPS5, RPP13, R-p/a1, R-p/a2, R-p/a4					[12–17]
Solanum pimpinellifolium		Cf-2 (5′ end)		Cf-2 (3′ end)		[18]
Inhibition of enemy infection and growth						
A. thaliana			ATTI1	CHIA, CHIB, CHIA	ATTI2	[60–62]
Arabis gemmifera						[63]
Populus tremula				Ti3, Ti4, Ti5, Ci1	Ti1, Ti2	[54]
Zea diploperennis		pr1, rip1	chiB, chil, hag, mir1, wip1[a], plt1, plt2, pr5, pr6, rip2	prms, zlp	chiA, mpi, hm2	[48,53]
Z. mays ssp. *parviglumis*[b]			chiA, chiB, chil, hag, mir1, plt1, plt2, pr1, pr5, pr6, prms, rip1, rip2, wip1, zlp, hm1		hm2	[48,53,64]
Z. perennis			mpi, wip1		hm2	[48]
Biosynthetic enzymes						
A. thaliana	MAM2		MAML, MAM1		TGG1	[56,59]

[a]Similar patterns of diversity in *Z. luxurians*, *Z. mays* ssp. *mays* and *Z. mays* ssp. *mexicana*.
[b]Multigene comparison of defense with non-defense genes reveals an apparent excess replacement polymorphism in defense genes [53].

strong evidence for selection (Table 1). A genome-wide scan of nucleotide diversity coupled with empirically generated distributions of molecular evolution test statistics also identified a chitinase gene as a candidate for having been a target of positive selection during maize domestication [45].

A particularly interesting example is *Zea hm2*, which confers partial resistance against *Cochliobolus carbonum* and is a duplicate of the first cloned R-gene, *hm1* [46] (*hm1* and *hm2* are not, however, functionally or structurally similar to NBS-LRR R-genes involved in detection). Three species of *Zea* have patterns of nucleotide diversity at *hm2* that indicate recent selection, yet the patterns are distinctly different. *Zea mays* ssp. *parviglumis* has one haplotype at great frequency [47], *Z. diploperennis* shows evidence of a recent selective sweep, and *Z. perennis* has three distinct haplotypes [48]. These patterns seem consistent with multi-allele co-evolutionary models that predict complex cycling of alleles through time [49]. If this explanation is correct then we would expect that the *hm2* alleles present in these species would encode enzymes that differ in their ability to inhibit the growth of different pathogen genotypes. Testing this model, however, could be difficult because the pathogen genotypes against which some of the *hm2* alleles are most effective might either no longer exist or now occur at low frequencies. Regardless, parallels with multi-allele models are only evident because multiple closely related species were examined and seem to be at different phases of cycling; results from any single species would be consistent with a simpler co-evolutionary model (e.g. arms-race [20] for *Z. diploperennis*).

Analyses of intra-specific diversity for *D. melanogaster* antimicrobial protein-encoding genes are similar to those from plants in that most individual genes (including genes encoding attacins [50], cecropins [51], defensin, drosocin and metchnikowin [52]) reveal little evidence for strong selection. The two exceptions, *andropin* and a *diptericin*, harbor an excess of rare polymorphism, suggestive of recent selective sweeps [51]. Similar to what has been found for functionally similar plant proteins, no single-gene analysis reveals evidence for selectively maintained polymorphism.

Because molecular evolutionary tests of non-neutral evolution have limited statistical power, it is possible that the failure of most intralocus analyses to detect evidence of selection is a result of the short lengths of genes encoding protein-based defenses. One approach that increases statistical power and helps to account for complications arising from demographic history is to compare diversity at groups of functionally related defense genes with diversity at groups of non-defense genes [32]. This approach has been applied to pathogen-inhibiting proteins in *Z. mays* ssp. *parviglumis* [53] and *D. simulans* [52], and both analyses detected a significant excess of amino acid replacement polymorphism. Excess replacement polymorphism has also been reported for a protease inhibitor gene, *TI1,* in *Populus tremula* [54], *RPP13* in *A. thaliana* [17] and *Pto* in *L. peruvianum* (L.E. Rose *et al.*, unpublished). These elevated levels of replacement polymorphism seem inconsistent with neutral equilibrium expectations but also defy simple evolutionary scenarios such as selective

sweeps or balanced polymorphisms. Elevated replacement polymorphism, however, could be consistent with negative frequency-dependent selection, where alleles are favored only when rare. Alternatively, studies employing species-wide collections of individuals, such as the *parviglumis* study, could be sampling across local populations in which different alleles have been favored by selection. Nevertheless, the effects that incomplete sweeps, multi-allele frequency-dependent selection or geographically variable selection have on patterns of nucleotide diversity have not been modeled formally (see Stahl *et al.* [12] for a model predicting the effects of two-allele frequency-dependent selection on nucleotide polymorphism).

Other defense-related genes

Although little studied by molecular population geneticists, many defenses are products of biochemical pathways – secondary metabolites – that are either preformed or induced following infection or attack. One exception involves glucosinolate biosynthesis and hydrolysis genes in the Brassicaceae. A major quantitative trait locus (QTL) for glucosinolate production lies within the *GS-Elong* region, which contains *MAM1, MAM2* and *MAM-L* [55]. Nucleotide polymorphism in one biosynthesis-related gene, *MAM2*, seems to be maintained by balancing selection, whereas other closely linked genes (*MAM1, MAM-L, MYB37*) have patterns of variation consistent with neutral expectations [56]. In support of the molecular evolutionary analyses, ecological studies have detected stabilizing selection acting on glucosinolate concentrations [57], and functional studies have indicated that variation at *MAM2* affects herbivore resistance [56]

Comparison of *MAM2* with its duplicate, *MAM1*, however, revealed an elevated rate of nonsynonymous nucleotide substitutions, suggesting positive selection following duplication. *MAM1* seems to have acquired a novel biochemical function following duplication – positive selection is probably associated with this neofunctionalization [58]. Although functional studies support claims of past selection on *MAM1*, such a relationship was not apparent for another gene involved in glucosinolate biosynthesis, *TGG1*. *TGG1* harbors low nucleotide diversity and an excess of high-frequency derived variants, a pattern consistent with recent positive selection [59]. *TGG1* alleles did not, however, correspond to variation in myrosinase activity found among *A. thaliana* accessions, leading Stranger and Mitchell-Olds [59] to conclude that the pattern of molecular diversity at this gene could reflect the demographic history of *A. thaliana* rather than selection. The increasing availability of empirically derived distributions of population genetic parameters should provide an opportunity to discriminate better between selective and demographic explanations for patterns of molecular diversity.

Conclusions and future prospects

Recent molecular evolutionary analyses have been used to characterize the evolutionary processes that shape nucleotide variation in genes involved in protecting plants against herbivores and pathogens. These analyses reveal apparently diverse selective histories including rapidly evolving protein-based defenses (e.g. chitinases), ancient

Table 2. Suggested sampling schemes and rationale for molecular evolutionary analyses that will advance our understanding of the evolution of plant immunity genes

Sampling	Rationale
Intralocus diversity for more *A. thaliana* R-genes	Compare the selective history of R-genes located in different genomic environments: isolated, in small clusters, and in large complex clusters
Intralocus diversity at R-genes from a more diverse assemblage of species	Determine generality of patterns emerging from *A. thaliana*. The high selfing rate and population structure of this species could have important influences on R-gene evolution and complicate interpretation of molecular data
All defense components (detection, signaling, response) from a single species	Direct comparisons of the evolutionary history of different types of defense, without complications arising from comparisons across species
Multiple genes from multiple closely related species	Identify evolutionary lineages in which enemies have been particularly important selective agents
Genes involved in defense against different enemies (fungi, bacteria, herbivores, viruses)	Identify the importance of selection imposed by different classes of enemies
Coordinated sampling of host defense and enemy targets	Examine molecular co-evolution [40,41,65,66]
Diversity within multiple geographically distinct subpopulations	Investigate the importance of geographically variable selective pressures (geographically variable co-evolution [67]) in shaping molecular diversity [2]

polymorphisms maintained by selection (e.g. *RPS2*) and genes in which only purifying selection has operated. Although the number of molecular population genetic studies remains limited, some patterns have begun to emerge. For example, balancing selection seems to maintain polymorphism at many plant detection genes. By contrast, genes encoding protein-based defenses induced following enemy detection tend to have more complex evolutionary histories that sometimes involve strong episodes of directional selection.

There are many promising directions for future molecular evolutionary investigations of plant immunity (Table 2). For example, the evolution of plant genes encoding proteins involved in immune system signaling cascades has not been examined. *Drosophila* signaling genes, however, seem to be targets of strong selection. Characterizing the evolutionary history of signaling genes will increase our understanding of how plants counter enemy invasion and how enemies evade plant immune responses.

Another promising direction is the coupling of molecular population genetics with ecological genetic and functional investigations. Although the application of molecular population genetic analyses has cast new light on the evolution of plant defense genes, it is often difficult to distinguish a signature of natural selection from that of demographic events (e.g. population expansion or population bottlenecks) leading to speculative interpretations of patterns of nucleotide variation. Functional and ecological genetic studies can strongly bolster tests of adaptive hypotheses. Functional analyses can offer insight into whether changes at the nucleotide or protein level lead to altered patterns of expression or enzymatic activity, as can be expected under positive selection. Beyond function, examining the fitness consequences of changes at the molecular level under natural field conditions should prove most powerful for elucidating the mechanisms of adaptive evolution. Overall, broadening molecular evolutionary studies to more classes of genes involved in the defense response and including additional plant species (especially non-model organisms) will enable a more rigorous evaluation of general patterns to the evolutionary processes that shape plant defense systems.

References
1 Rausher, M.D. (2001) Co-evolution and plant resistance to natural enemies. *Nature* 411, 857–864
2 de Meaux, J. and Mitchell-Olds, T. (2003) Evolution of plant resistance at the molecular level: ecological context of species interactions. *Heredity* 91, 345–352
3 Stahl, E.A. and Bishop, J.G. (2000) Plant-pathogen arms races at the molecular level. *Curr. Opin. Plant Biol.* 3, 299–304
4 Michelmore, R.W. and Meyers, B.C. (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* 8, 1113–1130
5 Bergelson, J. *et al.* (2001) Evolutionary dynamics of plant R-genes. *Science* 292, 2281–2285
6 Meyers, B.C. *et al.* (1998) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *Plant Cell* 11, 1833–1846
7 Dodds, P.N. *et al.* (2000) Genetic analysis and evolution of plant disease resistance genes. In *Molecular Plant Pathology* (Dickinson, M. and Beynon, J., eds), pp. 88–107, Sheffield Academic Press
8 Wang, G-L. *et al.* (1998) *Xa21D* encodes a receptor-like molecule with a leucine rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. *Plant Cell* 10, 765–779
9 Couch, B.C. *et al.* (2006) Pervasive purifying selection characterizes the evolution of *I2* homologues. *Mol. Plant Microbe Interact.* 19, 288–303
10 Yang, Z. and Bielawski, J.P. (2000) Statistical methods for detecting molecular evolution. *Trends Ecol. Evol.* 15, 496–503
11 Mondragon-Palomino, M. *et al.* (2002) Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res.* 12, 1305–1315
12 Stahl, E.A. *et al.* (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400, 667–671
13 Caicedo, A.L. *et al.* (1999) Diversity and molecular evolution of the *Rps2* resistance gene in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 302–306
14 Mauricio, R. *et al.* (2003) Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis*. *Genetics* 163, 735–746
15 Tian, D. *et al.* (2002) Signature of balancing selection in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11525–11530
16 Shen, J. *et al.* (2006) Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics* 172, 1243–1250
17 Rose, L.E. *et al.* (2004) The maintenance of extreme amino acid diversity at the disease resistance gene, *RPP13*, in *Arabidopsis thaliana*. *Genetics* 166, 1517–1527

18 Caicedo, A.L. and Schaal, B.A. (2004) Heterogeneous evolutionary processes affect R gene diversity in natural populations of *Solanum pimpinellifolium*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17444–17449

19 Tian, D. *et al.* (2003) Pleiotropic cost of R-gene mediated resistance in *Arabidopsis thaliana*. *Nature* 423, 74–77

20 Frank, S.A. (1992) Models of plant pathogen coevolution. *Trends Genet.* 8, 213–219

21 Chisholm, S.T. *et al.* (2006) Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 124, 803–814

22 Dodds, P.N. *et al.* (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8888–8893

23 Richly, E. *et al.* (2002) Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. *Mol. Biol. Evol.* 19, 76–84

24 Meyers, B. *et al.* (2005) Evolving disease resistance genes. *Curr. Opin. Plant Biol.* 8, 129–134

25 Mondragon-Palomino, M. and Gaut, B.S. (2005) Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 22, 2444–2456

26 Haldane, J.B.S. (1949) Disease and evolution. *Ric. Sci. Suppl. A.* 19, 68–76

27 Salathe, M. *et al.* (2005) Neutral drift and polymorphism in gene-for-gene systems. *Ecol. Lett.* 8, 925–932

28 Thrall, P.H. and Burdon, J.J. (2002) Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal. *Plant Pathol.* 51, 169–184

29 Leclerc, V. and Reichhart, J-M. (2004) The immune response of *Drosophila melanogaster*. *Immunol. Rev.* 198, 59–71

30 Jiggins, F.M. and Hurst, G.D.D. (2003) The evolution of parasite recognition genes in the innate immune system: purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. *J. Mol. Evol.* 57, 598–605

31 Schlenke, T.A. and Begun, D.J. (2003) Natural selection drives *Drosophila* immune system evolution. *Genetics* 164, 1471–1480

32 Schlenke, T.A. and Begun, D.J. (2005) Linkage disequilibrium and recent selection at three immunity receptor loci in *Drosophila simulans*. *Genetics* 169, 2013–2022

33 Nimchuk, Z. *et al.* (2003) Recognition and response in the plant immune system. *Annu. Rev. Genet.* 37, 579–609

34 Begun, D.J. and Whitley, P. (2000) Adaptive evolution of relish, a Drosophila NF-κB/IκB protein. *Genetics* 154, 1231–1238

35 Bulmer, M.S. and Crozier, R.H. (2006) Variation in positive selection in termite GNBPs and Relish. *Mol. Biol. Evol.* 23, 317–326

36 Mudgett, M.B. (2005) New insights to the function of phytopathogenic bacterial type III effectors in plants. *Annu. Rev. Plant Biol.* 56, 509–531

37 Bishop, J.G. *et al.* (2000) Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5322–5327

38 Tiffin, P. (2004) Comparative evolutionary histories of chitinase genes in the genus *Zea* and the family Poaceae. *Genetics* 167, 1331–1340

39 Bishop, J.G. *et al.* (2005) Selection on glycine β-1,3-endoglucanase genes differentially inhibited by a *Phytophthora* glucanase inhibitor protein. *Genetics* 169, 1009–1019

40 Bishop, J.G. (2005) Directed mutagenesis confirms the functional importance of positively selected sites in polygalacturonase inhibitor protein. *Mol. Biol. Evol.* 22, 1531–1534

41 Stotz, H.U. *et al.* (2000) Identification of target amino acids that affect interactions of fungal polygalacturonases and their plant inhibitors. *Physiol. Mol. Plant Pathol.* 56, 117–130

42 Rose, J.K.C. *et al.* (2002) Molecular cloning and characterization of glucanase inhibitor proteins: coevolution of a counter-defense mechanism in plant pathogens. *Plant Cell* 14, 1329–1345

43 Cote, F. *et al.* (1998) Oligosaccharide elicitors in host-pathogen interactions: generation, perception, and signal transduction. *Subcell. Biochem* 29, 385–432

44 Silverstein, K.A.T. *et al.* (2005) Genome organization of more than 300 defensin-like genes in *Arabidopsis*. *Plant Physiol.* 138, 600–610

45 Wright, S.I. *et al.* (2005) The effects of artificial selection on the maize genome. *Science* 308, 1310–1314

46 Johal, G.S. and Briggs, S.P. (1992) Reductase activity encoded by the *HM1* disease resistance gene in maize. *Science* 258, 985–987

47 Zhang, L. *et al.* (2002) Population genetics of duplicated disease-defense genes, *hm1* and *hm2*, in maize (*Zea mays* ssp. *mays* L.) and its wild ancestor (*Zea mays* ssp. *parviglumis*). *Genetics* 162, 851–860

48 Tiffin, P. *et al.* (2004) Population genetic evidence for rapid changes in intraspecific diversity and allelic cycling of a specialist defense gene in *Zea*. *Genetics* 168, 425–434

49 Seger, J. (1988) Dynamics of some simple host-parasite models with more than two genotypes in each species. *Philos. T. Roy. Soc. B* 319, 541–555

50 Lazzaro, B.P. and Clark, A.G. (2001) Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the *attacin* genes of *Drosophila melanogaster*. *Genetics* 159, 659–671

51 Clark, A.G. and Wang, L. (1997) Molecular population genetics of *Drosophila* immune system genes. *Genetics* 147, 713–724

52 Lazzaro, B.P. and Clark, A.G. (2003) Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. *Mol. Biol. Evol.* 20, 914–923

53 Moeller, D.A. and Tiffin, P. (2005) Genetic diversity and the evolutionary history of plant immunity genes in two species of *Zea*. *Mol. Biol. Evol.* 22, 2480–2490

54 Ingvarsson, P.K. (2005) Molecular population genetics of herbivore-induced protease inhibitor genes in European aspen (*Populus tremula* L., Salicaceae). *Mol. Biol. Evol.* 22, 1802–1812

55 Kliebenstein, D. *et al.* (2002) Comparative analysis of quantitative trait loci controlling glucosinolates, myrosinase, and insect resistance in *Arabidopsis thaliana*. *Genetics* 161, 325–332

56 Kroymann, J. *et al.* (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14587–14592

57 Mauricio, R. and Rausher, M.D. (1997) Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense. *Evolution* 51, 1435–1444

58 Benderoth, M. *et al.* (2006) Positive selection driving diversification in plant secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 103, 9118–9123

59 Stranger, B. and Mitchell-Olds, T. (2005) Nucleotide variation at the myrosinase-encoding locus, *TGG1*, and quantitative myrosinase enzyme activity variation in *Arabidopsis thaliana*. *Mol. Ecol.* 14, 295–309

60 Kawabe, A. *et al.* (1997) Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. *Mol. Biol. Evol.* 14, 1303–1315

61 Kawabe, A. and Miyashita, N.T. (1999) DNA variation in the basic chitinase locus (*ChiB*) region of the wild plant *Arabidopsis thaliana*. *Genetics* 153, 1445–1453

62 Clauss, M.J. and Mitchell-Olds, T. (2003) Population genetics of tandem trypsin inhibitor genes in *Arabidopsis* species with contrasting ecology and life history. *Mol. Ecol.* 12, 1287–1299

63 Kawabe, A. and Miyashita, N.T. (2002) DNA variation in the acidic chitinase locus (*ChiA*) region in *Arabis gemmifera* and its related species. *Genes Genet. Syst.* 77, 167–175

64 Tiffin, P. and Gaut, B.S. (2001) Molecular evolution of the wound-induced serine protease inhibitor *wip1* in *Zea* and related genera. *Mol. Biol. Evol.* 18, 2092–2101

65 Allen, R.L. *et al.* (2004) Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. *Science* 306, 1957–1960

66 Rose, L.E. *et al.* (2005) Natural variation in the *Pto* pathogen resistance gene within species of wild tomato (*Lycopersicon*). I. Functional analysis of *Pto* alleles. *Genetics* 171, 345–357

67 Thompson, J.N. (2005) *The Geographic Mosaic of Coevolution.* University of Chicago Press

68 Rausher, M.D. (1996) Genetic analysis of coevolution between plants and their natural enemies. *Trends Genet.* 12, 212–217

69 Wright, S.I. and Gaut, B.S. (2004) Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.* 22, 506–519

70 Asai, T. *et al.* (2002) MAP kinase signaling cascade in *Arabidopsis* innate immunity. *Nature* 415, 977–983

71 Durrant, W.E. and Dong, X. (2004) Systemic acquired resistance. *Annu. Rev. Phytopathol.* 42, 185–209

72 Bostock, R.M. (2005) Signal crosstalk and induced resistance: straddling the line between cost and benefit. *Annu. Rev. Phytopathol.* 43, 545–580

73 van Loon, L.C. *et al.* (2006) Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.* 44, 135–162

74 Ausubel, F.M. (2005) Are innate immune signaling pathways in plants and animals conserved? *Nat. Immunol.* 6, 973–979

75 Nielsen, R. (2005) Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197–218

76 Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595

77 Fu, Y. and Li, W. (1993) Statistical tests of neutrality of mutations. *Genetics* 133, 693–709

78 Fay, J.C. and Wu, C-I. (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413

79 Hudson, R.R. *et al.* (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116, 153–159

80 McDonald, J.H. and Kreitman, M. (1991) Adaptive protein evolution at the *Adh1* locus in *Drosophila*. *Nature* 351, 652–654

81 Kim, Y. and Nielsen, R. (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167, 1513–1524

82 Kelly, J.K. (1997) A test of neutrality based on interlocus associations. *Genetics* 146, 1197–1206

83 Przeworski, M. (2002) The signature of positive selection at randomly chosen loci. *Genetics* 160, 1179–1189

84 Nordborg, M. *et al.* (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 3, 1289–1299