

THE ARMS RACE IS ANCIENT HISTORY IN *ARABIDOPSIS*, THE WILDFLOWER

Eric B. Holub

Plant pathology was born after the nineteenth-century potato famine, and since then insightful genetic experiments have contributed to the great progress in our understanding of disease control. Our current view of plant resistance focuses on numerous polymorphic resistance loci, which contain genes known as *R* genes. The complete sequence of the *Arabidopsis thaliana* genome provides a framework for exploring the 'big bang' of *R* genes that occurred and how *R* genes evolved in plants from their associations with microorganisms, and for improving strategies for more sustainable deployment of disease resistance in crops.

BIOTROPIC

A type of parasite that derives its energy from the living cells of its host.

NECROTROPIC

A type of parasite that derives its energy from dead cells that are part of a live host.

Current knowledge of how disease resistance has evolved in plants has been greatly influenced by research into the association of plants with highly specialized BIOTROPIC pathogens, such as the causal agents of powdery mildew, downy mildew, blight and rust in a wide diversity of annual crops. These pathogens have adapted to infect and reproduce in a narrow range of host species, often resulting in disease resistance with a high degree of genetic variability within the suitable host species (often referred to as genotype or race-specific resistance); by contrast, the resistance in a vast majority of alternative plant species seems to be genetically non-variable (referred to henceforth as species-level resistance). For example, the subspecies of the fungus that causes powdery mildew in wheat (*Erysiphe graminis*) does not naturally cause disease in barley, and vice versa.

During the past decade, many genes that determine genotype-specific resistance (so-called *R* genes) have been characterized on the molecular level from natural variation in crops and in wild species, most notably *Arabidopsis thaliana*. Artificial mutagenesis has been used effectively to generate further genetic variation, which has begun to reveal underlying genes that are required for the resistance conferred by *R* genes. This has allowed researchers to propose models for the process of defence signalling, a subject that has been reviewed extensively elsewhere¹⁻⁸. It is important to note that the

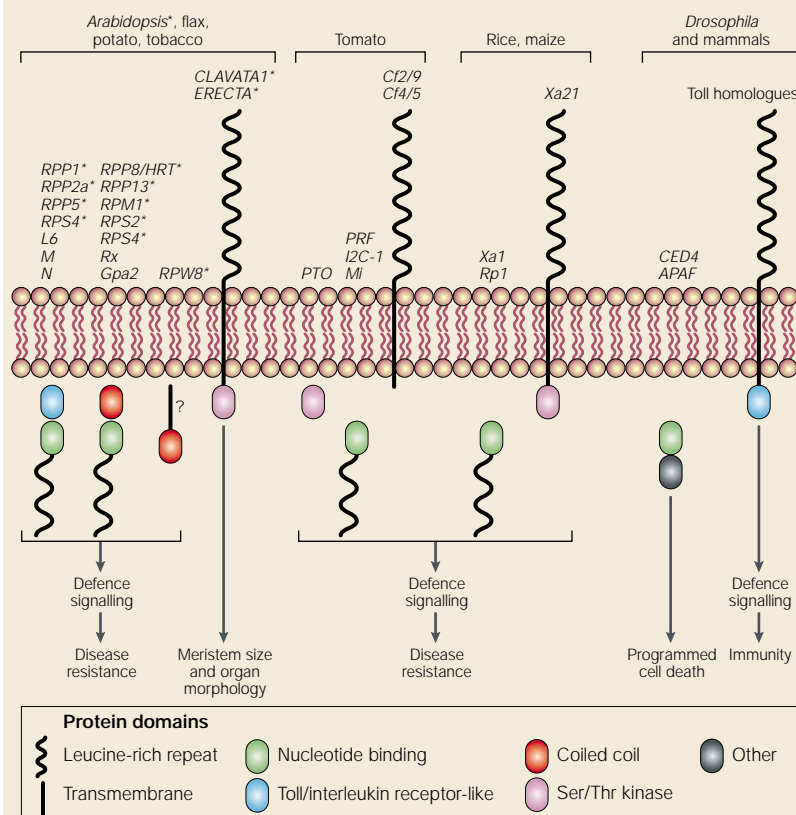
relevance of these recent molecular advances for understanding resistance involving plants and parasites with alternative lifestyles (for example, woody perennials and NECROTROPIC fungi) is poorly understood.

This article uses the current knowledge of disease resistance in *Arabidopsis* to discuss what is known about the functional and structural features of *R* genes and how recent DNA sequence information from this wildflower has advanced our understanding of how such genes might have evolved in higher plants. I discuss issues surrounding the 'arms race' debate on how polymorphism in *R* genes is thought to have arisen and is maintained, and propose that contemporary plant species in natural ecosystems depend more on recycling relatively old genetic variation than on continually generating new resistance specificities.

The infamous potato blight Late blight of potato (caused by the 'plant destroyer' *Phytophthora infestans*) was central in an epidemic that devastated this staple crop during the middle of the nineteenth century, and, as a consequence, significantly altered the demographics of Europe and North America⁹⁻¹⁰. The pathogen was introduced and began to proliferate throughout western Europe in a manner similar to the recent crisis of 'foot and mouth' disease in European livestock. At that time, however,

Species Level Resistance Research Group, Plant Genetics and Biotechnology Department, Horticulture Research International, Wellesbourne CV35 9EF, UK.
e-mail: eric.holub@hri.ac.uk

Box 1 | Structural features of *R* genes in plants



The largest structural class of proteins that are encoded by resistance (*R*) genes is characterized by two domains: the leucine-rich-repeat (LRR) domain and a nucleotide-binding (NB) site. The LRRs vary in number (typically ranging from 10–40 repeats) and each repeat can vary in its motif of ~24 amino acids¹⁵. The NB site is a more highly conserved protein motif, common in all organisms. NB motifs occur in many structural conformations and are presumed to bind ATP/GTP and/or phosphate at a highly conserved 'P-loop'^{22,28}.

NB–LRR genes constitute one of the largest and most diverse gene families in plants. In the relatively small genome of *Arabidopsis thaliana* (~120 Mb), more than 150 such genes have been annotated in the 'completed' DNA sequence that was announced in December 2000 by the *Arabidopsis* Genome Initiative³³. At least another ten genes have been predicted by Kato *et al.*⁵⁹ in one of the gaps on chromosome 1 that still remains to be sequenced⁶⁰. There are two major NB–LRR CLADES, which are generally distinguished by the presence or absence of an amino-terminal Toll/interleukin receptor-like (TIR) domain, described previously in genes in *Drosophila melanogaster* and mammals^{20,21}. Meyers *et al.*²⁵ refer to these clades as TIR-class and non-TIR-class genes; whereas Pan *et al.*²⁸ refer to them as group I and group II, respectively (throughout this article I will use the former designation). Non-TIR genes either lack a third domain or have an amino-terminal coiled-coil (CC) domain, which might be involved in protein dimerization.

Other structural classes of *R* genes in plants also exist. For example, *Cf2/9* and *Cf4/5*, which belong to a family of fungal

resistance genes in tomato, encode membrane-bound LRR proteins without an NB site¹⁵; the powdery mildew resistance gene *RPW8* in *Arabidopsis* encodes an unusual small protein with a possible transmembrane region and a CC domain⁶¹; the bacterial resistance gene *PTO* in tomato encodes a serine–threonine kinase⁶²; and the bacterial resistance gene *Xa21* in rice combines a Ser/Thr kinase with an LRR⁶³. These genes seem to represent variations on a theme in the evolutionary shuffling of domains (NB, LRR, CC, TIR, Ser/Thr kinase) to achieve a wide variety of functional capability in plants. Such domains are not used exclusively for defence, as illustrated by the *CLAVATA1* and *ERECTA*^{64,65} genes in *Arabidopsis*. These encode highly conserved proteins, structurally similar to the rice *Xa21* gene, that are involved in organ development. Interestingly, evidence indicates that the origin of NB, LRR and TIR domains might be ancient, and precede the divergence of plants and animals. The conserved motifs of the NB site are found in regulators of cell death in animals such as *CED4* (also known as *APAF*)²²; and Toll-like proteins involved as receptors for the innate immune response in animals, combine a TIR and an LRR domain, but not an NB site.

APAF, apoptotic protease activating factor; *Cf*, *Cladosporium fulvum* resistance gene; *CED4*, cell death 4; *Gpa*, *Globodera pallida* (potato cyst nematode) resistance gene; *HRT*, turnip crinkle virus resistance gene; *I2C-1*, *Fusarium oxysporum* resistance gene; *L6* and *M*, *Melampsora lini* resistance genes in flax; *Mi*, resistance to *Meloidogyne incognita*; *N*, tobacco mosaic virus resistance gene in tobacco; *PRF*, gene required for *PTO* function; *PTO*, *Pseudomonas syringae* pv. tomato resistance gene; *Rp*, *Puccinia sorghi* resistance gene; *RPM*, resistance to powdery mildew; *RPP*, resistance to *Peronospora parasitica*; *RPS*, resistance to *Pseudomonas syringae*; *RPW*, powdery mildew resistance gene; *Rx*, resistance to potato virus X; *Xa*, *Xanthomonas campestris* pv. *oryzae* resistance gene; *Xa21*; resistance to the bacterial pathogen *Xanthomonas oryzae* pv. *oryzae*.

CLADE
An organismal lineage comprising an ancestor and all its descendants.

microorganisms were not widely accepted as causes of disease in plants or animals. Potatoes were imported to West Flanders from North and South America in the hope of rejuvenating supplies of seed tubers that appeared to be suffering throughout Europe from a progressive degeneration or loss of vigour. Pollution or a change in the weather was generally thought to be the cause of this malady, but it was probably due to an increase in chronic fungal and viral diseases. Unfortunately, late blight had already begun to spread in eastern coastal areas of North America a few years before the exportation of tubers to Europe. As the pathogen began to spread, advice from amateur mycologists in Europe suggesting that affected plants should be destroyed to

contain the new disease was not heeded. Another decade passed before *P. infestans* was taxonomically identified and widely accepted as the cause of late blight.

Nonetheless, the stage was set for emergence of a new agricultural science — plant pathology — with practical goals to classify pathogens, to understand the survival and transmission of those pathogens in crops, and to devise control strategies that would minimise diseases in crop production.

The 'gene for gene' legacy
Natural variation in disease resistance was among the earliest examples of Mendelian trait to be genetically described in plants, and has since provided a rich

resource for crop improvement¹¹. One result of this discovery was that much of the variation used for breeding resistant crops has been monogenic. Such resistance is often highly selective in its ability to control only disease caused by specific genotypes of the pathogen, and is therefore often disparaged for its role in 'boom and bust' cycles of disease control in crops. These cycles have involved sequential deployment of CULTIVARS with different resistance specificities that have each in turn succumbed within a few years of monoculture to the emergence of new virulent genotypes of the pathogen. However, monogenic resistance can be highly effective, particularly if strategies for deploying such genes can be improved.

In the mid-1900s, Harold Flor¹² provided a cornerstone for contemporary research of monogenic disease resistance with his combined genetic investigations of both a plant host (flax) and a fungal rust parasite (*Melampsora lini*). He concluded from his experiments that a 'gene for gene' relationship occurs in matching pairs between resistance (*R*) genes in the host and avirulence (*Avr*) genes in the pathogen. He observed that the host was resistant as long as it contained at least one *R* gene that matched a corresponding *Avr* gene in the infecting pathogen.

Most importantly, Flor showed the power of genetics in plant pathology, and provided a unifying model that opened the door for later molecular analyses. For example, increasing evidence indicates that *Avr* genes might often encode determinants required for causing disease in a susceptible host^{13,14}. 'Avirulence' therefore is an appropriate name for these virulence determinants when expressed in a resistant host, because *R* genes often encode receptor-like proteins^{15,16,17}, which presumably have evolved to enable the plants to detect and respond defensively to the matching *Avr* protein (BOX 1). Similar genetic approaches have been adopted in investigations of innate resistance in animals¹⁸. Intriguingly, animals and plants have common defence strategies¹⁹, which extend to structural similarities in their respective receptors for detecting pathogens^{15,20–22}.

The wildflower *Arabidopsis thaliana*

Over the past decade, this wild relative of horticultural brassicas (for example, cabbage, broccoli, cauliflower) and oilseed rape has been instrumental in dissecting monogenic disease resistance at the molecular level. Evidence from *Arabidopsis* has helped to eliminate the misconception that monogenic resistance is an artefact of selective breeding in crops; such resistance is as common in this wild species as it is in crops, and *R* genes that govern resistance in *Arabidopsis* are structurally similar to *R* genes in crops (BOX 1). Ineffective deployment of disease resistance through monoculture is therefore the most likely cause of boom and bust. Mutational analyses of *Arabidopsis* have also strongly supported the view that so-called monogenic resistance is typically the concerted expression of multiple genes in a cascade of defence responses. The *R* gene is a single polymorphic locus that constitutes the pathogenic-specific defence component of a multigenic process (at least in the examples that have been molecularly investi-

gated). Presumably, the other components of the defence response are functionally conserved or redundant in the same species (multiple defence responses elicited by the same *R* gene, for example) and are therefore not genetically variable.

Comparative sequence analyses of *R* genes^{16,17,23–30}, largely underpinned by examples from *Arabidopsis*, focus attention on the unresolved debate over how *R*-gene variation has been generated. *R* genes have had a tremendous capacity to proliferate and diversify in plants, allowing the host to keep pace with pathogens that are themselves diversifying to overcome prevalent sources of resistance in the host. Evidence that genetic variation is maintained is also substantial. However, investigations of CROP PATHOSYSTEMS (an inherently closed system because evolution of genetic variation is severely restrained in the host) has not resolved whether natural plant populations continue to keep pace with their pathogens by evolving further novel disease resistance or instead rely on recycling of existing genetic variation. Determining the relative age of known *R* genes is central to clarifying this debate.

Proliferation of *R* genes in plants

Disease resistance genes have been molecularly characterized from diverse lineages of plants, including cereals (rice, barley and maize), DICOTYLEDONOUS crops (flax, tomato, tobacco and lettuce), and the wildflower *Arabidopsis*. The first disease resistance gene to be characterized, *Hm1* from maize³¹, is the only example for which the mode of resistance is known. This gene encodes a reductase enzyme that inactivates a toxin that is secreted into plant cells by the fungal pathogen *Cochliobolus carbonum*. The toxin is thought to interfere with transcription of host defence genes and allows the fungus to grow unrestricted.

Widespread reliance on evolving enzymes to protect against every new variant of a potential pathogen would be a cumbersome means to counter changes in a highly variable pathogen. Defence mechanisms involving more complex signalling responses seem to have provided plants with a prolific and presumably more flexible means to expand their capability for resistance. This is indicated by evidence that most *R* genes in plants encode receptor-like proteins, which probably allow parasite perception as the first step in plant defence^{15,16,18,26}. Further evidence comes from the common observation from crops that *R* genes are often clustered in the genome^{11,26}. An allelic series of different resistance specificities at a single gene locus have also been described, such as 13 alleles at the *L* locus in flax that differ in specificity of rust resistance¹⁷.

In *Arabidopsis* alone, more than 30 resistance specificities have been identified using widely divergent pathogens³². For example, phenotypic variation of resistance to downy mildew (*Peronospora parasitica*), caused by a close relative of *P. infestans*, is shown in FIG. 1. Four bacterial specificities (*RPM1*, *RPS2*, *RPS4* and *RPS5*) and four downy mildew specificities (*RPP1*, *RPP5*, *RPP8* and *RPP13*) have been described at a molecular level (TABLE 1); these are alleles of the largest structural class of *R* gene,

CULTIVAR

A cultivated variety (genetic strain) of a domesticated crop plant.

PATHOSYSTEM

An ecological subsystem defined by the phenomenon of parasitism. A plant pathosystem might include one or more host plant species along with the various parasites (insects, fungi, bacteria and so on) that use the hosts.

DICOTYLEDONS

Members of a subphylum of angiosperms that have two seed leaves (cotyledons) in the embryo.

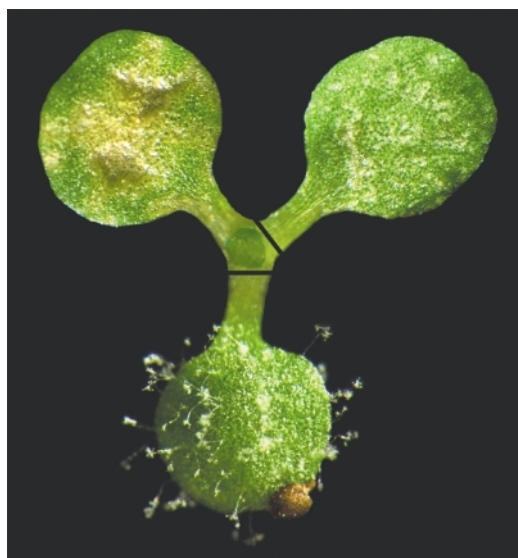


Figure 1 | Phenotypes caused by downy mildew (*Peronospora parasitica*) in *Arabidopsis*. *Arabidopsis thaliana* cotyledons show a spectrum of phenotypes one week after inoculation with different isolates of *Peronospora parasitica*, a close relative of *Phytophthora infestans*, the potato late blight pathogen. Bottom: full susceptibility of the accession Wassilewskija to the isolate Emwa1. The parasite grows unrestricted throughout the host and sporulates heavily as it emerges from every STOMATAL opening. Note the tree-like structures, which bear an abundance of spores. Top right: discrete host cell death response by the Landsberg *erecta* ACCESSION to the isolate Emc05. Tissue colonized by the pathogen, which appears as necrotic flecks, is barely visible on the host surface. This phenotype is conferred by the downy mildew resistance gene, *RPP8*. Top left: expansive necrotic response shown by the accession Wassilewskija to the isolate Emoy2. This phenotype is conferred by the gene *RPP1* and is readily visible to the unaided eye. *RPP8* resistance only permits the parasite to penetrate through the epidermis and into the first MESOPHYLL cell, which rapidly commits suicide to contain the parasite and provides a signal for defence responses in neighbouring cells. By contrast, *RPP1* resistance permits the parasite to colonize much further into the host; this results in larger patches of necrosis that eventually contain the parasite and limit its reproduction.

STOMATA

Natural openings in the epidermis of a stem or leaf of a plant that are surrounded by specialized guard cells, and permit gas exchange with the air.

ACCESSION

A sample of a plant variety collected at a specific location and time.

MESOPHYLL

Leaf cells that lie beneath the outer epidermal cell layer.

PATHOVAR

Abbreviation for pathogen variety, a subspecific classification that indicates host origin and to some extent host specialization.

which is characterized by a highly conserved nucleotide-binding (NB) site and hyper-variable leucine rich repeat (LRR) domain (BOX 1). More than 150 NB–LRR genes have been revealed by the completed genome sequence of a single *Arabidopsis* accession (Columbia, the source for the completed genome sequence³³) (FIG. 2). Although the *Arabidopsis* pathogens are widely divergent, no known functional class of *R* gene or their genomic organization (for example, single or multiple gene) are predictive of resistance to a particular pathogen.

RPM1 and *RPS2*, which confer resistance to *Pseudomonas syringae*, were among the first plant *R* genes to be characterized molecularly, and both were mapped to single loci in the same accession, Columbia (Col-0)^{34,35,36}. A different crop pathogen was used to characterize each gene: an isolate from *Brassica oleracea* (*Pseudomonas syringae* PATHOVAR (pv.) *maculicola* carrying *avrRpm1*), and an isolate from tomato (*Pseudomonas*

syringae pv. *tomato* carrying *avrRpt2*), respectively. *RPM1* and *RPS2* were originally identified by genetic analysis and positional cloning of artificially induced *rpm1* or *rps2* mutants of Col-0 that were susceptible to each respective pathogen. Wild-type accessions that were naturally susceptible to either pathogen were eventually identified, as well as natural variants of *P. syringae* that lacked avirulence determinants and were therefore able to grow in Col-0. Both *RPM1* and *RPS2* belong to the CC–NB–LRR subclass of *R* genes. An important distinction between the two genes, however, is that accessions that naturally lack resistance to *avrRpt2* have wild-type *rps2* alleles of unknown function, whereas only deletions of *RPM1* have been described in wild *avrRpm1*-susceptible accessions. Interestingly, Caicedo *et al.*²³ describe different allelic classes of *rps2* in other accessions (represented in their publication by Wü-0 and Zu-0-7) that lack resistance to *avrRpt2*, indicating that alternative resistance specificities have evolved at this locus to unknown avirulence determinants.

The study of *RPS4* and *RPS5* has shown that bacterial resistance is conferred in *Arabidopsis* by members of tightly clustered NB–LRR gene families^{37,38} (FIGS 3, 4). Both genes were characterized using avirulence determinants (*avrRps4* and *avrPphb*, respectively) that were derived from legume pathogens of *P. syringae* (*P. syringae* pv. *pisi* from pea and *P. syringae* pv. *phaseolicola* from green bean, respectively), and had been characterized previously as *Avr* determinants in the legume hosts. *Arabidopsis* shows species-level resistance to these crop pathogens, so experimental isolates were derived by introducing the *avrRps4* or *avrPphb* genes into a Col-0 virulent pathogen (called a carrier isolate). Wild-type accessions and artificial mutants of *Arabidopsis* were identified that were susceptible to these isolates. This enabled molecular genetic analyses that revealed *RPS4* as a TIR–NB–LRR gene³⁷, and *RPS5* as a CC–NB–LRR gene³⁸. Using a similar approach, the avirulence determinant *avrB* from a soybean isolate of *P. syringae* was introduced into a Col-0 virulent carrier isolate and used to show that *RPM1* recognizes *avrB* as well as *avrRpm1*. This showed an intriguing example of dual specificity³⁵.

RPP8/HRT provided an important demonstration of allelic variation in resistance specificity to widely divergent pathogens. *RPP8* was characterized first in the Landsberg *erecta* accession (Ler-0) as a CC–NB–LRR gene that confers resistance to an *Arabidopsis* isolate of *Peronospora parasitica*³⁹. In this accession, *RPP8* is tightly linked to an *RPP8* homologue (designated *RPH8A*), which has no known function. By contrast, the susceptible accession Col-0 contains only a single gene (*rpp8-Col*) at the same locus, which seems to have been derived from unequal crossing-over between ancestral alleles of *RPP8-Ler* and *RPH8A* in a genotype similar to Ler-0. The result of the recombination event was loss of one copy and retention of an intragenic hybrid gene, which probably caused the loss of resistance specificity. However, functional resistance at a single gene locus, as found in Col-0, has also evolved in other accessions, as demonstrated by the *HRT* gene for turnip crinkle virus, recently characterized from the *Arabidopsis* Di-17

Table 1 | Pathogen-resistance genes that have been molecularly characterized in *Arabidopsis thaliana*

Pathogen*	Locus	Chromosome	Cluster ID [†]	Molecular structure [§]	Functional allele	Isolate(s) detected	Refs
<i>Pseudomonas syringae</i>	<i>RPM1</i>	III	Singlet (F17A9)	CC–NB–LRR	Col-0 (A57072)	AvrRpm1, AvrB	35
	<i>RPS2</i>	IV	Singlet (F20B18)	CC–NB–LRR	Col-0 (A54809)	AvrRpt2	34,36
	<i>RPS4</i>	V	J2 (K21C13–K9E15)	TIR–NB–LRR	Col-0 (CAB50708)	AvrRps4	37
	<i>RPS5</i>	I	A1 (T28K15–F5O11)	CC–NB–LRR	Col-0 (AAC26126)	AvrPphb	38
<i>Peronospora parasitica</i>	<i>RPP1</i>	III	F1 (T22K7–T18B22)	TIR–NB–LRR	WsA (AF098964)	Noco2, Emoy2, Maks9, Cala2	43
					WsB (AF098963)	Noco2, Emoy2, Maks9	
					WsC	Noco2	
	<i>RPP2a</i>	IV	H2 (F24J7)	TIR–NB–LRR	Col-0 (CAB78952)	Cala2	45
	<i>RPP5</i>	IV	H1 (A1FCA0)	TIR–NB–LRR	Ler-0 (AAF08790)	Noco2	46
	<i>RPP7</i>	I	B2 (F9K23–T30E16)	CC–NB–LRR	Col-0 (n.a.)	Hiks1	69
	<i>RPP8</i>	V	Doublet (MWF20)	CC–NB–LRR	Ler-0 (AF089710)	Emco5, Waco5	39
<i>RPP13</i>	III	F2 (F12A12–T6H20)	CC–NB–LRR	Nd-1 (AF209732)	Emco5, Maks9, Aswa1	42	
				Rld-2 (AF209731)	Wela3		
Turnip crinkle virus	<i>HRT</i>	V	Singlet (MWF20)	CC–NB–LRR	Di-17 (AF234174)	(not specified)	40
<i>Erysiphe</i> spp.	<i>RPW8</i>	III	Doublet (T20E23)	TM–CC	Mt (AF273059)	18 isolates from 4 spp.	61

* Examples include bacterial (*P. syringae*), fungal (*Erysiphe* spp.), OOMYCETE (*P. parasitica*, fungal-like descendent from yellow-green algae) and viral (turnip crinkle virus) pathogens. [†] Refers to clusters shown in FIG. 3, which identifies groups of three or more homologous genes in an interval of less than 200 kb. They are designated according to chromosome arms (A and B, top and bottom arms of chromosome 1; F, H and J are bottom arms of chromosome 3, 4 and 5, respectively). One and two gene loci are identified as singlets and doublets. The BAC (bacterial artificial chromosome) contig that spans the interval containing the genes is shown in parentheses. [§] CC, coiled-coil domain; TI, Toll/interleukin receptor-like domain; NB, nucleotide-binding site; LRR, leucine-rich repeat region; TM, transmembrane region. ^{||} *Arabidopsis* accession provided the allele: Col-0, Columbia; Ws-0, Wassilewskija; Ler-0, Landsberg *erecta*; Nd-1, Niederzenz; Rld-2, Rschew; Di, Dijon. GenBank accession number is shown in parentheses. ^{||} Not available. *RPP7* was mapped to a gap that remains in the *Arabidopsis* Genome Initiative sequencing of *Arabidopsis*. Kato *et al.*⁵⁹ predict that this interval contains a family of ~15 genes that are similar in sequence to *RPP8*.

accession⁴⁰. Interestingly, *RPP8* and *HRT* seems to differ in the signalling components of downstream defence responses: along with *RPP7*, *RPP8* was the first example of downy mildew resistance that does not depend on salicylate (a defence signalling component) in *Arabidopsis*, whereas *HRT* seems to be a salicylate-dependent example of virus resistance^{5,41}.

The *RPP13* gene provides an example of allelic variation in the specificity of resistance to a single pathogen, *P. parasitica*⁴². Alternative specificities of this CC–NB–LRR gene were characterized from accessions Niederzenz (Nd-1) and Rschew (Rld-2). The gene is a member of a triplet in Col-0 (this cluster is designated F2 in FIG. 3), which contains a *rpp13* allele that might confer resistance to an unknown avirulence determinant. It will be interesting to determine if *RPP13* has been evolving independently because it is separated from the other two copies by 17 genes, largely as the result of multiple duplications of an intervening glucuronosyl transferase-like gene (FIG. 4).

The *RPP1*, *RPP2* and *RPP5* loci provide further contrasting examples of TIR–NB–LRR downy mildew resistance genes that are members of gene clusters. Four *RPP1* gene members (*WsA*, *WsB*, *WsC* and *WsD*) were identified in the accession Wassilewskija (*Ws-0*), and three were shown to provide different resistance specificities (*WsD* was not characterized)⁴³ (TABLE 1). The completed *Arabidopsis* sequence indicates that Col-0 contains four genes at this locus that are dispersed among many other genes (FIG. 4); their function, however, is unknown. By contrast, genes at the *RPP2* and *RPP5* loci are mostly arranged in tandem arrays. *RPP2* resistance was mapped to a single locus in Col-0 (REF. 44) but, interestingly, only one TIR–NB–LRR gene (designated *RPP2a*) has so far been confirmed as being required, but insufficient, for resistance to the matching

P. parasitica isolate⁴⁵. Resistance was restored when this gene was used to transform a susceptible *rpp2a* mutant of Col-0, but did not confer resistance on its own in susceptible wild-type accessions. A second, closely linked gene that is also required to confer this resistance is therefore predicted to exist in Col-0.

RPP5 was the first downy mildew resistance gene to be characterized in *Arabidopsis*, and belongs to one of the largest NB–LRR clusters described in Col-0 (H1; FIG. 4). The functional allele was identified as one of ten copies in Ler-0 at the centromeric end of a 95-kb interval. This region shows a high degree of divergence from the corresponding region in Col-0, which only contains eight copies^{27,46}. The NB–LRR genes themselves seem to have diversified owing to several evolutionary forces, including unequal crossing-over (both intergenic, which has altered the number of NB–LRR copies, and intragenic, which has altered the number of LRR repeats within different genes), GENE CONVERSION, point mutation and retrotransposon activity. The high degree of polymorphism has made it impossible to distinguish orthologous copies (which occupy the same physical position within the cluster in different accessions) from flanking copies (which are surrounded by highly similar sequence).

The functional role of most *Arabidopsis* NB–LRR genes, other than those described here, is unknown; however, it is interesting that nearly half of the NB–LRR genes in *Arabidopsis* are located in one of thirteen clusters (defined here as three or more copies in a region of less than 200 kb) that are distributed throughout the genome of the Col-0 accession (FIG. 3). Most of these reside in megaclusters (20 cM regions) on the bottom arms of chromosomes 1 and 5 (previously described genetically as major resistance complexes, *MRC-B* and

OOMYCETE
A phylum of filamentous eukaryotic microorganisms that, although fungal-like, are more closely related to yellow-green algae.

GENE CONVERSION
The non-reciprocal transfer of information between homologous genes as a consequence of heteroduplex formation, followed by repair of mismatches in the heteroduplexes.

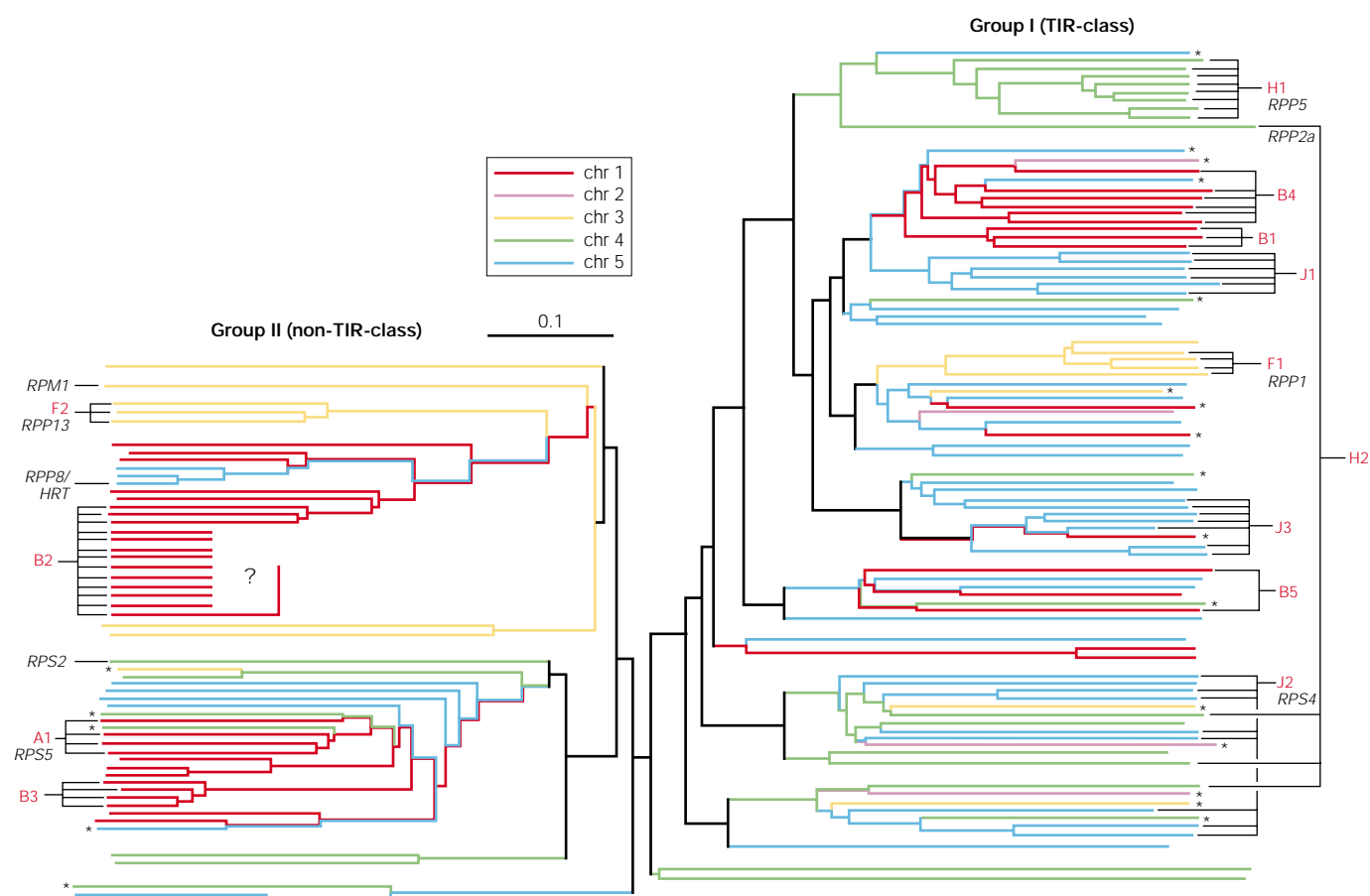


Figure 2 | **Unrooted phylogenetic tree of nearly 150 NB-LRR genes in the *Arabidopsis thaliana* accession Columbia.** The two main clades of NB-LRR gene (groups I and II) are shown here. The coloured lines indicate the chromosome from which each gene is derived, and clusters of genes in regions of less than 200 kb are indicated by brackets. Letters in red type refer to chromosome arms: A and B, top and bottom arms of chromosome 1; F, H and J, bottom arms of chromosomes 3, 4 and 5, respectively. Numbers in red type refer to clusters shown in FIG. 3. Known functional genes that confer resistance to the bacterium *Pseudomonas syringae* (*RPS* or *RPM* designations) or the eukaryotic oomycete *Peronospora parasitica* (*RPP*) are indicated in black type. Asterisks indicate genes for which the nearest related genes are located on a different chromosome. The question mark shown in the B2 cluster indicates additional genes predicted by Kato *et al.*⁵⁹, which lie in a gap that remains to be sequenced by the *Arabidopsis* Genome Initiative. Note the clustering on the same chromosome of phylogenetically related NB-LRR genes. Bar indicates unit of phylogenetic distance based on the proportion of amino-acid substitutions (Modified from data analysis provided in February 2001 by Meyers *et al.* (see link to [UC Davis](#)), which updates two previous publications^{21,24}.)

MRC-J (REF.47)). The remaining NB-LRR genes in *Arabidopsis* are scattered throughout the genome in simple (singlet or doublet) loci. Singlets and doublets also seem to be common within the NB-LRR gene clusters because 3 or more tandem copies (uninterrupted by other types of genes) were only found in 6 of the 13 clusters (B2, B4, B5, H1, H2 and J3) (FIG. 4). Surprisingly, only four genes in total were identified at separate loci on chromosome 2.

Diversification of NB-LRR genes in plants
 Michmore and Meyers²⁶ have proposed a model, adapted from research into the evolution of the vertebrate major histocompatibility complex (MHC) and immunoglobulin gene families⁴⁸, that postulates mechanisms for the proliferation of *R* genes in plants, and for their subsequent diversification either in allelic series at single gene loci or as reservoirs of variation in multiple gene loci. This two-phase 'birth and death' model assumes that the first duplication event (birth), which would have 'primed the

pump' for further *R*-gene evolution, has already occurred. This assumption seems to be valid in *Arabidopsis* because doublet NB-LRR loci and even-numbered doublets within multiple gene loci (FIG. 4) are common throughout the *Arabidopsis* genome. In addition, transposable elements — likely contributors to the initial doubling events — are commonly found at these loci. The number of *R* genes would have probably expanded within loci by unequal crossing-over between mispaired tandem copies during meiosis. Subsequent divergence (death) of these different copies can result from further unequal crossing-over (especially intragenically to create chimeric genes, as demonstrated above by *RPP8/HRT*), point mutation, gene conversion or transposon insertion. Evidence for all of these events is also illustrated at the *RPP5* locus (described above). By analysing data on *R*-gene loci in crops, Michmore and Meyers²⁶ conclude that neighbouring genes in the same cluster (PARALOGOUS genes) are typically more divergent than different alleles of the same gene in different haploid genotypes of the same plant

PARALOGUE

Homologous genes that originated by gene duplication.

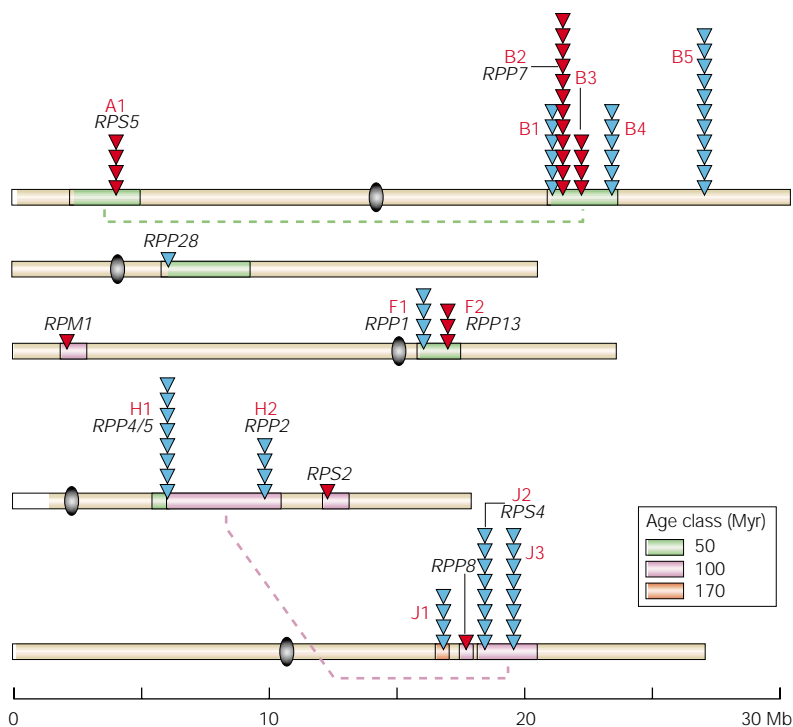


Figure 3 | Map of NB-LRR gene clusters and of resistance genes in the *Arabidopsis* genome. The chromosomal locations of NB-LRR gene clusters and of known functional genes that confer resistance to the bacterium *Pseudomonas syringae* (*RPS* or *RPM* designations) or the eukaryotic oomycete *Peronospora parasitica* (*RPP*) in *Arabidopsis thaliana* are shown. The clusters are defined by the presence of more than three genes in a region of less than 200 kb, and are named according to the chromosome arms on which they reside. Group I genes (which contain a TIR domain) are indicated by blue triangles, and group II genes (which do not contain a TIR domain) are indicated by red triangles. Black ovals indicate the approximate location of the centromere. Vision *et al.*⁵⁸ have estimated the age of genomic duplications, which are shown here by horizontal bars on the chromosomes, to 50 (green), 100 (purple) or 170 (orange) million years (Myr) ago; hatched lines beneath the chromosomes indicate related duplications. (Letter/number designations in red type refer to locations on chromosome arms described in FIG. 2.)

species, or even in different species (ORTHOLOGOUS genes) — a conclusion supported by recent evidence from the *N* locus for rust resistance in flax⁴⁹. However, examples from complex loci in *Arabidopsis* have not been investigated sufficiently; in sequencing members of *RPP5* gene family from only two haploid genotypes, Noël *et al.* were unable to distinguish reliably orthologous genes²⁷.

The common theme revealed from the molecular study of *R* genes in plants with an LRR domain has been the evidence for positive selection for diversification. Protein variation can be assessed by comparing base-pair changes in nucleotide sequence from numerous variants of the same gene (orthologues or paralogues) that either alter the encoded amino acid (non-synonymous substitutions) or leave the amino acid unaltered (synonymous substitutions). The ratio of non-synonymous (K_a) to synonymous (K_s) amino-acid changes provides a measure of diversifying selection; a K_a/K_s value less than 1 indicates conservative evolution of the sequence, whereas a value greater than 1 indicates positive selection to diversify⁵⁰.

Parniske and colleagues²⁹ were the first to use this comparative method to analyse sequence variation in *R* genes, in an examination of tandemly repeated genes at

the *Cf4/Cf9* locus from different subspecies of tomato. They, and others^{23,27,40,43,51}, have reached the conclusion that the LRR domain shows much higher levels of diversity, particularly at solvent-exposed faces in the repeats, than other domains within the genes. These changes have occurred in addition to changes in the number of LRR repeats. The combined effect indicates that variation in the LRR domain might be important for determining the specificity of a given *R* gene. However, recent evidence from flax rust indicates that the more highly conserved TIR domain can also determine resistance specificity⁵².

Further empirical data is required to assess whether the birth and death model provides a universal description of *R*-gene evolution, and whether diversifying selection inferred from K_a/K_s analyses of NB-LRR genes is a valid hypothesis. The birth and death model seems to be robust because it incorporates the idea that multiple gene loci can maintain genetic variation, mostly of relatively old origin, and can therefore serve as a reservoir of DNA sequence for generating novel specificities. However, it falls short of describing the spatial and temporal dynamics of existing genetic variation in local populations, as described below. The evidence for diversifying selection seems to be overwhelming; however, the selective forces have not been rigorously debated in the literature. Although pathogen(s) are intuitively an important selective force, physical interaction between an NB-LRR protein and a corresponding Avr protein has yet to be convincingly shown. It is also important to note that neither the age of given *R* alleles nor the rate of diversification can be inferred from these analyses.

Maintaining *R*-gene polymorphisms

The debate surrounding how polymorphisms are maintained in natural populations revolves around the issue of whether 'defeated' *R*-gene alleles (those variants to which pathogens have evolved resistance) are driven to extinction and are therefore transient in host populations, or instead simply become rare until they are recycled, increasing in frequency as the corresponding avirulence re-emerges in the pathogen population (BOX 2).

A biological arms race between host and parasite has been a useful concept for describing the evolution of disease resistance⁵³ (BOX 3). Stahl and colleagues³⁰ attempted to determine whether *R* genes are transient or recycled in plants using a small collection of wild *Arabidopsis* to investigate allelic variation at the single gene locus *RPM1*. The transient polymorphism model, according to which *R* alleles are replaced in each cycle by new ones, and that Stahl *et al.* referred to in a restrictive sense as an 'arms race', was rejected on the basis that *RPM1* is an old resistance specificity without alternative functional alleles. The apparent lack of functional alternatives at this locus might actually support a transient model if the contemporary functional allele represents an adaptive optimum. A major constraint in testing a transient model with only one example of a single gene locus is that the crucial evidence, namely extinct alleles, are not available. Conversely, Stahl *et al.* provide a strong argument in favour of recycling polymorphism, which they prefer to call 'trench warfare'. This

ORTHOLOGUE
Homologous genes that originated through speciation.

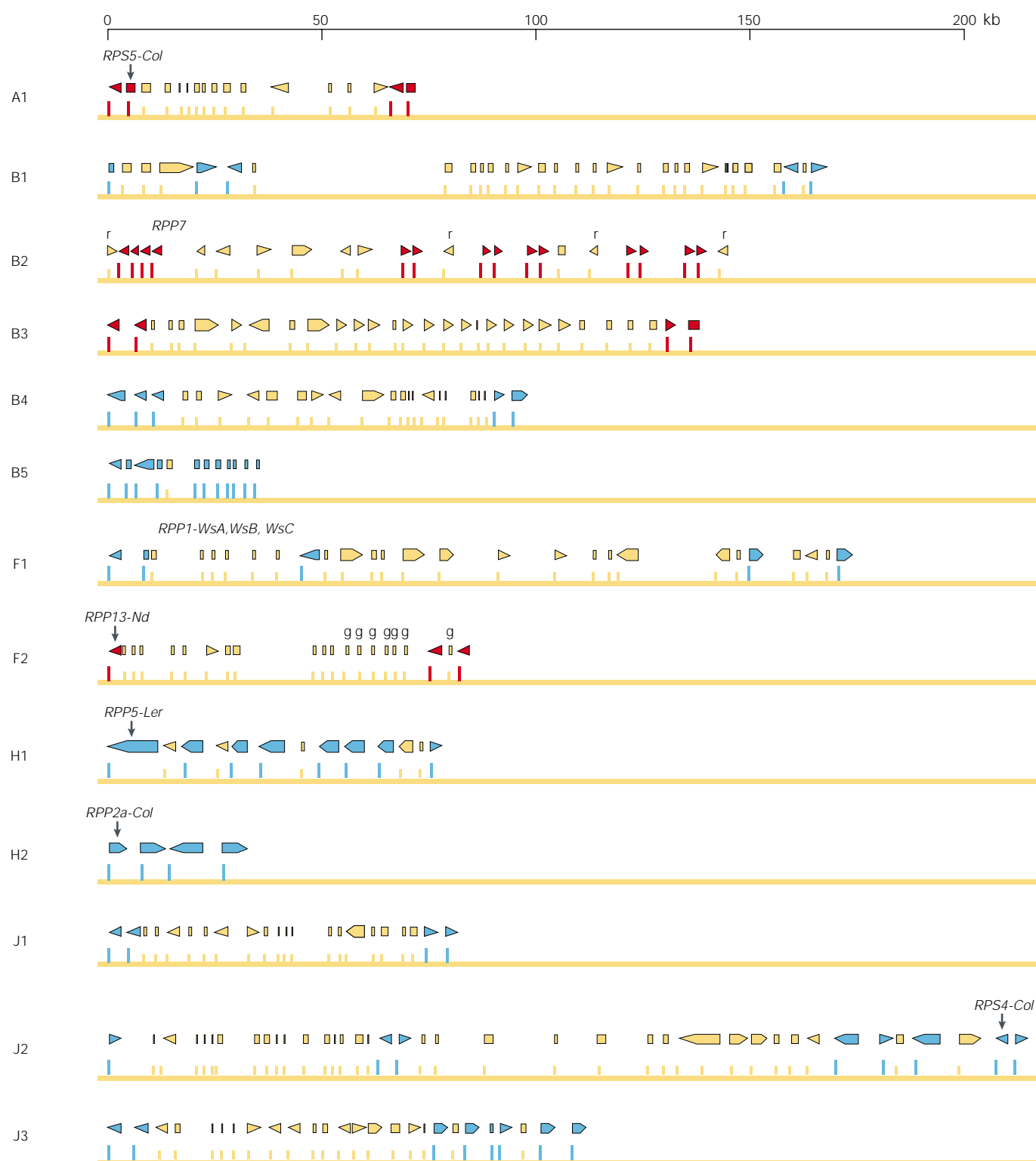


Figure 4 | Physical structure of regions of *Arabidopsis thaliana* that contain clusters of NB-LRR genes. The clusters of NB-LRR genes are defined here by more than three genes in a region of less than 200 kb. Group I genes (presence of TIR domain) are indicated with blue symbols, and group II genes (absence of TIR domain) are indicated with red symbols, with relative positions marked in corresponding colours on the scale below each cluster. Other genes are indicated with yellow symbols. The position of known functional genes that confer resistance either to the bacterium *Pseudomonas syringae* (*RPS* or *RPM* designations) or the eukaryotic oomycete *Peronospora parasitica* (*RPP*) is shown. In the B2 and F1 clusters, the corresponding physical location in Columbia has not yet been determined; the letters 'r' and 'g' indicate repetitive copies of ribosomal or glucuronosyl transferase-like genes, respectively. The points of the symbols indicate the direction of transcription. Ws, Wassilewskija; Nd, Niederzenz; Ler, Landsberg *erecta*; Col, Columbia. (Letter/number designations on the left refer to locations on chromosome arms described in FIG. 2.)

Box 2 | Evolution of *R* genes in local populations

The evolution of disease resistance genes in plants can be explained by population level models (reviewed in REF. 66), each of which is based on different assumptions regarding the origin and dynamics of the polymorphisms that exist at resistance (*R*)-gene loci.

According to the 'transient polymorphism model', old *R* alleles are sequentially driven to extinction within a local population by emerging virulent forms of the pathogen; however, these are replaced in each cycle by a new allele that has either diverged from an existing allele by mutation or recombination, or has immigrated from other populations.

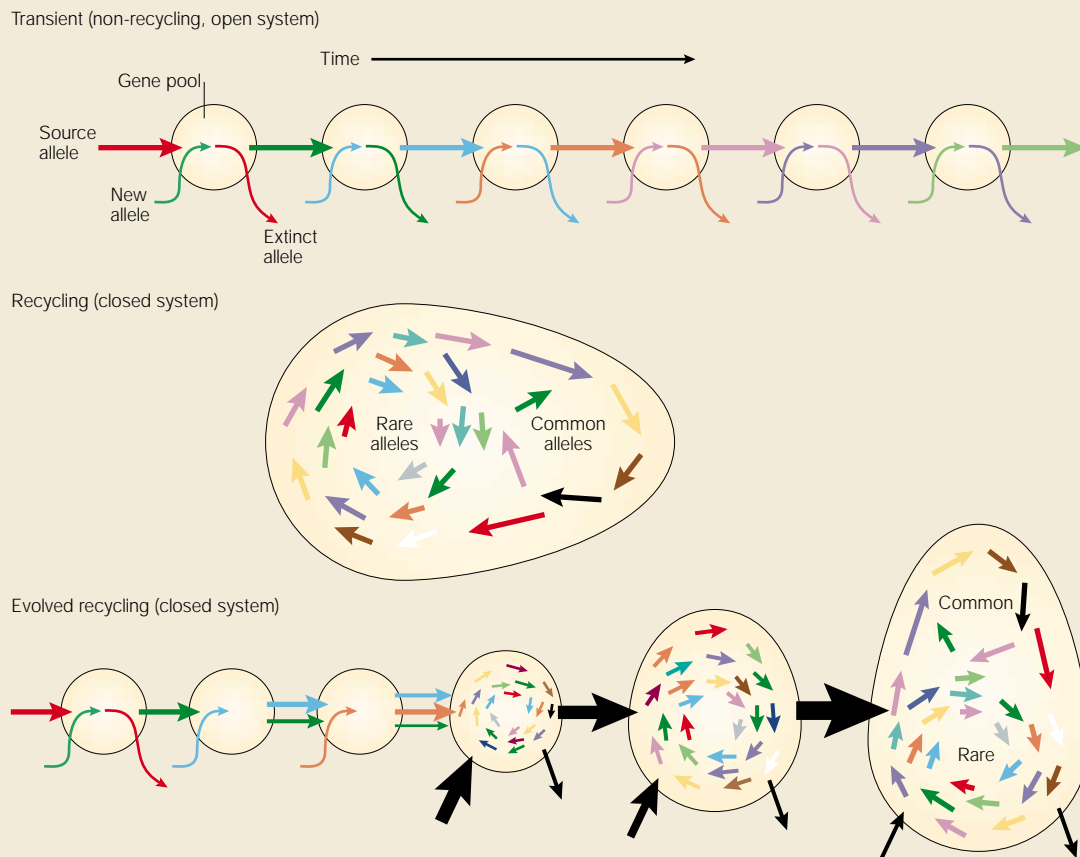
The 'recycling polymorphism' model assumes that many alleles are maintained in the same population, but that the system is artificially closed

to the introduction of new alleles or extinction of existing alleles. Rare alleles would be maintained in some organisms — at major histocompatibility complex loci in vertebrates, for example — by means of BALANCED POLYMORPHISM⁶⁷. Indeed, it has been recently indicated, in a report of rust resistance in maize⁶⁸, that heterozygote advantage occurs in plants. Although this seems to be an unlikely mechanism for maintaining *R*-gene polymorphism in a self-pollinating plant species, such as *Arabidopsis thaliana* (inbreeding tends to increase the proportion of homozygotes over heterozygotes), the low frequency of outcrossing that occurs in this species means that recombination of existing genetic variation should still have a significant effect on a MICROEVOLUTIONARY timescale.

Recycling assumes that: a pool of genetic variation has already been generated and is subsequently maintained; variation is recombined among individuals; and the variation undergoes FREQUENCY-DEPENDENT SELECTION. Alleles are common when they provide a principal fitness advantage by conferring effective disease resistance at a given time to the corresponding avirulent pathogen. Intense selection pressure on the pathogen will in turn cause new virulent forms to emerge and overcome the resistance prevalent in the host. Consequently, *R* alleles that protect against that virulent form will begin to increase in frequency as the 'defeated' alleles (*R*-gene variants to which the pathogen has become resistant) become rare. Most importantly, chance events rather than an emergent virulent pathogen will drive a rare allele to extinction.

These two models are not mutually exclusive, so the 'evolved recycling polymorphism' model is shown here as a possible synthesis. It proposes that the population progresses from transient cycles into recycling of variation as alleles begin to diverge rapidly and accumulate in the host population. Extensive proliferation of *R* genes is therefore a key element in the transition from a transient mode of evolution to a recycling mode.

Mathematical modelling of transient and balanced polymorphisms has been used to describe the population dynamics of *R* genes in plants (for an example, see REF. 66). Such modelling generally needs to assume random mating among individuals in a population to recombine alleles and does not account for important episodes such as polyploidization, which can greatly expand the potential genetic capacity for evolving new resistance specificities. The evolved recycling model can incorporate such an event; although this model might attempt to describe biology more accurately, it will be more difficult to describe mathematically.



BALANCED POLYMORPHISM
Selective advantage of heterozygous individuals.

MICROEVOLUTION
Refers to small-scale changes that usually occur at the level of the species, and on short timescales.

argument is supported by evidence from many other loci (for example, *RPP1*, *RPP8*, *RPP13*) in which alternative alleles are common, showing that polymorphism has been generated, has accumulated and is apparently maintained at these loci.

Population studies that incorporate current molecular knowledge of *R* genes will be essential for resolving

how such genes are maintained and used in a given plant species. Exciting opportunities were shown by a recent study of cultivated rice in China, which provided evidence that highly susceptible genotypes cannot only survive but can also be very productive in mixed planting with resistant genotypes⁵⁴. By extrapolation, defeated alleles would not necessarily be eliminated

from natural plant populations as long as they are not genetically linked to traits that are under intense selection pressure (which would make them more vulnerable to extinction). It will be interesting to determine whether genetic variation in *R* genes is predominantly maintained by combining many *R* genes in the same genotype, or by assorting functional genes into simple genotypes and maintaining them in heterogeneous populations. The latter has been suggested by two studies of wild groundsel (*Senecio vulgaris*) and flax^{55,56}. Detailed molecular genetic analyses have yet to be pursued in these pathosystems; however, functional NB–LRR genes already characterized in flax provide exciting examples for future studies of wild flax populations. *Arabidopsis* clearly has enormous potential as well, particularly with comparative investigations of its OUTCROSSING relative, *A. lyrata*.

Ancient origins of disease resistance
Cellular level investigations provide the most striking evidence for ancient origins of disease resistance, in which a common strategy for resisting an infecting parasite has been described in both animals and plants. The underlying defence mechanisms have several primal features, which are reflected in the existence of conserved protein motifs (BOX 1). Pathogen perception is coupled in both animals and plants with signalling mechanisms that involve common usage of molecules, such as nitric oxide and reactive oxygen species, to facilitate a controlled and rapid suicide of the first responding host cell¹¹. These defence signalling components have been assembled and compartmentalized differently in animals and plants, which may largely be determined by how the ancestral organisms evolved into mobile or sedentary, multicellular lifestyles.

Cellular autonomy seems to be a prerequisite for the sedentary lifestyle of plants. If every cell is equipped for rapid suicide when attacked, tight regulation must be required to reserve this capability for an appropriate response in attacked cells and contain it to avoid suicide of the whole organism. A signalling mechanism for such a response seems an obvious evolutionary solution for plants, along with proliferation of receptor-like *R* genes that can allow a specific response to a myriad of pathogens. It therefore seems plausible that evolution of *R* genes themselves will provide further evidence of ancient origins for disease resistance.

Phylogenetic analysis of NB–LRR (FIG. 2) clearly shows ancient ancestry reaching back to the emergence of higher plants^{25,28}. The coupling of NB–LRR genes to different defence responses provides an alternative means of containing parasites^{3,4,5,8}, and presumably was an important selective factor in ancient divergence of the two main NB–LRR clades. Interestingly, TIR–NB–LRR genes have not been identified so far in cereals, even though they represent two-thirds of the NB–LRR genes in *Arabidopsis*. As examples of this subclass have been found in pine, Meyers *et al.*²⁵ and Pan *et al.*²⁸ speculate that the two main NB–LRR clades are more ancient than the divergence of ANGIOSPERMS and GYMNOSPERMS, which occurred at least 200 million years

Box 3 | The biological ‘arms race’

The metaphor of a biological ‘arms race’ is a macroevolutionary description of how one species evolves in response to changes in another competing, predatory or parasitic species.

In an early discussion of this metaphor, Dawkins and Krebs⁵³ described numerous examples of an evolutionary escalation in ever more refined counter adaptations between two opponents, all of which had several common features: the parties racing against one another are lineages rather than individual organisms; the race therefore occurs over a historical timescale; and progressive changes rather than random changes are implied between the parties. (The distinction between progressive and random is used here to indicate the directionality or non-directionality, respectively, of phenotypic selection rather than the underlying mutational mechanisms, which in both cases will occur at random.) They also proposed that the race will evolve towards a reduction of the arms race (that is, the continued reliance on generating novel functional variation) rather than continued escalation. For instance, the race could end decisively with one species winning as it drives its opponent to extinction, or else by reaching an optimum adaptation that prevents the other from reaching its own optimum. Alternatively, the race might verge on stalemate either because both sides attain a mutual optimum in which genetic variation becomes fixed, or the race persists in an endless limit cycle in which the struggle continues by recycling of available genetic variation within populations of each opponent species. Examples that attempt to describe population, as opposed to species-level dynamics of *R* genes, are described in BOX 2.

(Myr) ago. This indicates that TIR–NB–LRR genes either have not proliferated or were lost during the evolution of MONOCOTYLEDONS, such as the cereals.

Stahl and colleagues³⁰ provided the first attempt to estimate the age of a functional *R* gene using a comparative analysis of DNA sequence variation in regions flanking the *RPM1* locus. They compared variation in sequence among accessions of two *Arabidopsis* species (*A. thaliana* and the outcrossing species *A. lyrata*), and concluded that the functional resistance allele and the null deletion allele have coexisted at this locus for ~10 Myr. This estimate coincides with the predicted divergence of *Brassica* and *Arabidopsis* lineages⁵⁷, in which deletions of *RPM1* seem to have occurred independently²⁴.

Vision and colleagues⁵⁸ have provided estimates for the age of large duplicated regions distributed throughout most of the *Arabidopsis* genome. These regions seem to have remained intact with respect to gene order for the estimated age of each region. It is interesting that all of the known functional NB–LRR genes are located in regions estimated to be at least 50 Myr old (FIG. 3). For instance, three of the single gene loci shown (*RPM1*, *RPS2* and *RPP8*) lie in regions that date to ~100 Myr ago — an important period for speciation in

FREQUENCY-DEPENDENT SELECTION

Describes the situation in which the selection pressure (in this case, imposed on the pathogen by different *R* genes alleles) varies with the proportion of individuals that carry a given allele or allelic combination in a population.

OUTCROSSING

(or outbreeding) Mating between unrelated individuals. *Arabidopsis lyrata* is self-incompatible and therefore breeds by crossing to other individuals in the species. By contrast, *A. thaliana* is self-compatible and therefore can self-pollinate.

ANGIOSPERMS

The phylum of flowering seed plants, including both dicotyledons and monocotyledons (for example, grasses and cereals).

GYMNOSPERMS

Non-flowering seed plants (for example, pine).

MONOCOTYLEDONS

Angiosperms that have one seed leaf (cotyledon) in the embryo.

the angiosperms. In the case of *RPM1*, the region is considerably older than that estimated by Stahl *et al.* *R*-gene clusters lie in other regions of similar age; duplication in clusters would certainly have occurred more recently, but one of the existing copies might represent an ancestral gene for each respective family.

Although the regional age estimates provided by Vision *et al.* are still highly speculative, they nonetheless provide a useful guide for choosing loci to further assess the relative age of *R* genes. For instance, *RPP2* and *RPP8* will be particularly interesting for such investigations. *RPP2* lies in a region ~100 Myr old, and includes four gene members that seem to be widely divergent within the TIR–NB–LRR subfamily (FIG. 4). By contrast, *RPP8* might provide an opportunity to estimate the age and perhaps determine the mechanism of an initial duplication event by comparing sequences from accessions with singlet or doublet loci. The four NB–LRR genes on chromosome 2, which seem to be widely divergent (FIG. 2), would also be an interesting example for age assessment regardless of whether functional alleles can be identified at any of the respective loci. It will be essential to include other plant genera in such studies as benchmarks for divergence. *A. lyrata* and *B. oleracea* could be used, as they diverged from *A. thaliana* ~5 and 10 Myr ago, respectively. POLYPHYLETIC origins of *R* genes could also be investigated by comparing closely related genes on different chromosomes (asterisks in FIG. 2).

Concluding remarks

Although the proliferation and diversification of NB–LRR genes is clearly evident, the proportion of these genes that contribute functional resistance specificities is unknown, and whether any participate in signalling processes other than disease resistance is also not clear. Efforts are being made to establish a functional role for numerous NB–LRR genes (for example, see the link to UC Davis). The wide distribution of known *R* genes in the *Arabidopsis* genome and the occurrence of at least one example in most of the multiple clusters of NB–LRR genes, supports the prediction that disease resistance will prove to be the primary role of the entire NB–LRR family.

The emergence of *R* genes, and of much of the genetic variation at these loci, seems to be relatively old, at least pre-dating the domestication of plants and perhaps even the divergence of many contemporary plant species. Proliferation and diversification therefore represent MACROEVOLUTIONARY aspects of disease resistance. Important factors, such as the identity of the pathogens that contributed to the proposed diversifying selection of functional *R* genes, will therefore mostly be beyond the reach of hard, empirical evidence. Determining the age of known functional genes, as shown by Stahl and colleagues³⁰, will certainly help to substantiate whether much of the genetic variation is, in fact, old. In future, the complete genome sequence of additional plant genomes (for example, rice and tomato), as well as more divergent relatives (for example, liverwort or green algal species), will be especially useful for tracing the origins of *R*-gene families such as the NB–LRR genes.

The maintenance of existing genetic variation at *R*-gene loci is a microevolutionary aspect of disease resistance, so empirical data should be more readily obtainable to test current hypotheses (BOX 2). The investigation of spatial and temporal changes in natural populations are essential to advance knowledge in this subject. The debate surrounding transient versus recycling polymorphisms as described above is deliberately simplified for the sake of discussion. However, the two models are not mutually exclusive, as I have attempted to illustrate in the evolved recycling model (BOX 2). The early stages in the evolution of disease resistance, when genetic variation was limited, must have progressed in a manner different from its subsequent evolution following partial or whole genome duplications (as evident even in a relatively small plant genome of *Arabidopsis*), as well as proliferation of genes within complex loci. This raises an important point: the historical context of events that are unique to a particular species and that have markedly altered the course of its evolution cannot be ignored in any future debate. Clear reminders of this can be seen in the unique characteristics inherent to each of the *R*-gene loci described above.

Arabidopsis affords tremendous opportunities to advance our understanding of *R*-gene evolution, both at the molecular level in refining our understanding of defence signalling and of mechanisms that drive the evolution of *R*-gene specificities, and at the population level in establishing how genetic variation is maintained and whether existing genetic variation is recycled.

As my profession is plant pathology rather than evolutionary or molecular biology, I wonder how this research in *Arabidopsis* will affect crop improvement. Studies of *RPM1*, *RPS4* and *RPS5*, for example, already show that *R*-gene-mediated resistance to crop pathogens does in fact exist in *Arabidopsis*, including evidence for components of species-level resistance in the case of *RPS4* and *RPS5*. Similarly, one could expect that other receptor-like genes contribute towards species-level resistance to downy mildew pathogens from crops, and perhaps even to *P. infestans*. Knowledge of how well the underlying defence mechanisms are conserved between *Arabidopsis* and crops is essential before transfer of species-level resistance into a suitable crop will be possible. Nevertheless, the prospect of old resistance specificities and a wildflower that is universally resistant to many important crop pathogens provides enormous scope for optimism.

Links

DATABASE LINKS *CF4* | *PTO* | *Xa21* | *CLAVATA1* | *ERECTA* | *APAF* | *RPM1* | *RPS2* | *RPS4* | *RPS5* | *RPP1* | *RPP5* | *RPP8* | *RPP13* | *avrRpt2* | *avrRps4* | *avrPphb* | *HRT* | *rpp8* | *RPP7* | *rpp13* | *RPP2* | *Cf4* | *Cf9*
 FURTHER INFORMATION *Phytophthora infestans* | Resurgence of the Irish potato famine fungus | Functional and comparative genomics of *R* genes in plants at UC Davis | The *Arabidopsis* Information Resource | Nottingham *Arabidopsis* stock centre | AIMS: *Arabidopsis* Information Management System | Eric Holub's lab

POLYPHYLETIC
 Descended from different ancestors.

MACROEVOLUTION
 Describes evolution that occurs at or above the level of species.

1. Dangi, J. L. & Jones, J. D. G. Plant pathogens and integrated defence responses to infection. *Nature* **411**, (in the press)
2. Durner, J. & Klessig, D. F. Nitric oxide as a signal in plants. *Curr. Opin. Plant Biol.* **2**, 369–374 (1999).
3. Glazebrook, J. Genes controlling expression of defense responses in *Arabidopsis*. *Curr. Opin. Plant Biol.* **2**, 280–286 (1999).
4. Hammond-Kosack, K. & Jones, J. D. G. In *Biochemistry and Molecular Biology of Plants*, (eds Buchanan, B., Gruissem, W. & Jones, R.) 1102–1155 (Am. Soc. Plant Physiol. 2000).
This chapter from a graduate-level textbook provides an excellent summary with many helpful illustrations of disease resistance in plants.
5. McDowell, J. M. *et al.* Downy mildew (*Peronospora parasitica*) resistance genes in *Arabidopsis* vary in functional requirements for *NDR1*, *EDS1*, *NPR1* and salicylic acid accumulation. *Plant J.* **22**, 523–529 (2000).
6. Morel, J. B. & Dangi, J. L. The hypersensitive response and the induction of cell death in plants. *Cell Death Differ.* **4**, 671–683 (1997).
7. Ryals, J. A. *et al.* Systemic acquired resistance. *Plant Cell* **8**, 1809–1819 (1996).
8. Shapiro, A. D. Using *Arabidopsis* mutants to delineate disease resistance signalling pathways. *Can. J. Plant Pathol.* **22**, 199–216 (2000).
9. Bourke, A. Potato blight in Europe in 1845: the scientific controversy. In *Phytophthora* (eds Lucas, J. A., Shattock, R. C., Shaw, D. S. & Cooke, L. R.) 12–24 (Cambridge Univ. Press, Cambridge, UK, 1991).
10. Large, E. C. *The Advance of the Fungi* (Dover Publications, New York, 1962).
11. Crute, I. R. & Pink, D. A. C. Genetics and utilization of pathogen resistance in plants. *Plant Cell* **8**, 2747–2755 (1996).
A useful contemporary review that provides a plant breeding perspective on recent advances in molecular characterization of disease resistance.
12. Flor, H. H. The current status of the gene for gene concept. *Annu. Rev. Phytopathol.* **9**, 275–296 (1971).
13. Collmer, A. Determinants of pathogenicity and avirulence in plant pathogenic bacteria. *Curr. Opin. Plant Biol.* **1**, 329–335 (1998).
14. Gabriel, D. W. Why do pathogens carry avirulence genes? *Physiol. Mol. Plant Pathol.* **55**, 205–214 (1999).
15. Jones, D. & Jones, J. The role of leucine-rich repeat proteins in plant defenses. *Adv. Bot. Res.* **24**, 89–167 (1997).
16. Staskawicz, B. J., Ausubel, F. M., Baker, B. J., Ellis, J. G. & Jones, J. D. G. Molecular genetics of plant disease resistance. *Science* **268**, 661–667 (1995).
Although numerous reviews of R genes have followed, this paper is historically important for being the first to describe the common structural features in R genes, which was surprising news at the time.
17. Ellis, J. *et al.* Advances in the molecular genetic analysis of the flax-flax rust interaction. *Annu. Rev. Phytopathol.* **35**, 271–291 (1997).
Another historically important review that updates Flor's classic 'gene for gene' description of disease resistance in flax with advances in molecular characterization of R genes.
18. Malo, D. & Skamene, E. Genetic control of host resistance to infection. *Trends Genet.* **10**, 365–370 (1994).
19. Gura, T. Roots of immunity. *New Sci.* **165**, 24–33 (2000).
20. Aderem, A. & Ulevitch, R. J. Toll-like receptors in the induction of the innate immune response. *Nature* **406**, 782–787 (2000).
21. Hoffman, J. A., Kafatis, F. C., Janeway, C. A. & Ezekowitz, R. A. B. Phylogenetic perspectives in innate immunity. *Science* **284**, 1313–1318 (1999).
22. van der Kleen, E. A. & Jones, J. D. G. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr. Biol.* **8**, 226–227 (1998).
23. Caicedo, A. *et al.* Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **96**, 302–306 (1999).
24. Grant, M. *et al.* Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **95**, 15843–15848 (1998).
25. Meyers, B. *et al.* Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* **20**, 317–332 (1999).
- Examines the motif structure of NB-LRR genes from several plant genera and presents an extensive phylogenetic tree of these genes, highlighting the divergence between TIR- and non-TIR subclasses and noting the apparent absence of TIR-containing genes in grasses.**
26. Michmore, R. W. & Meyers, B. C. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* **8**, 1113–1130 (1998).
An ambitious review that provides a useful conceptual framework for explaining how disease resistance genes have proliferated and diversified in plants.
27. Noël, L. *et al.* Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus in *Arabidopsis*. *Plant Cell* **11**, 2099–2111 (1999).
28. Pan, Q., Wendel, J. & Fluhr, R. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* **50**, 203–213 (2000).
This paper should be read in conjunction with reference 25, providing further insights into conserved motifs of the NB and coiled-coil domains.
29. Parniske, M. *et al.* Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. *Cell* **91**, 821–832 (1997).
30. Stahl, E. *et al.* Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**, 667–671 (1999).
The paper which has rekindled much needed debate about the evolution of disease resistance in plants.
31. Briggs, S. P. & Johal, G. S. Genetic patterns of plant host-parasite interactions. *Trends Genet.* **10**, 12–16 (1994).
32. Holub, E. In *The Gene-for-Gene Relationship in Plant Parasite Interactions* (eds Crute, I., Holub, E. & Burdon, J.) 5–26 (CAB International, Wallingford, UK, 1997).
33. The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
34. Bent, A. *et al.* *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance gene. *Science* **265**, 1856–1860 (1994).
35. Grant, M. *et al.* Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* **269**, 843–846 (1995).
36. Mindrinos, M. *et al.* The *Arabidopsis thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide binding site and leucine-rich repeats. *Cell* **78**, 1089–1099 (1994).
37. Gassman, W. *et al.* The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease resistance genes. *Plant J.* **20**, 265–277 (1999).
38. Warren, R. *et al.* A mutation within the leucine rich repeat domain of the *Arabidopsis* disease resistance gene *RPS5* partially suppresses multiple bacterial and downy mildew resistance. *Plant Cell* **10**, 1439–1452 (1998).
39. McDowell, J. M. *et al.* Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. *Plant Cell* **10**, 1861–1874 (1998).
40. Cooley, M. B. *et al.* Members of the *Arabidopsis HRT/RPP8* family of resistance genes confer resistance to both viral and oomycete pathogens. *Plant Cell* **12**, 663–676 (2000).
41. Kachroo, P. *et al.* Resistance to turnip crinkle virus in *Arabidopsis* is regulated by two host genes and is salicylic acid dependent but *NPR1*, ethylene, and jasmonate independent. *Plant Cell* **12**, 677–690 (2000).
42. Bittner-Eddy, P., Crute, I. R., Holub, E. B. & Beynon, J. L. *RPP13* is a simple locus in *Arabidopsis* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *Plant J.* **21**, 177–188 (2000).
43. Botella, M. *et al.* Three genes of the *Arabidopsis RPP1* complex locus recognize distinct *Peronospora parasitica* (downy mildew) avirulence determinants. *Plant Cell* **10**, 1847–1860 (1998).
44. Tor, M. *et al.* Map positions of three loci in *Arabidopsis thaliana* associated with isolate specific recognition of *Peronospora parasitica* (downy mildew). *Mol. Plant-Microbe Interact.* **7**, 214–222 (1994).
45. Sinapidou, E. Genetic and molecular dissection of the *RPP2* locus of the *Arabidopsis thaliana* accession Columbia that confers resistance to the *Peronospora parasitica* isolate Cala2. Thesis, Imperial College of Biotechnology, Science and Medicine at Wye, University of London, UK (2000).
46. Parker, J. E. *et al.* The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the Toll and interleukin-1 receptors with *N* and *L6*. *Plant Cell* **9**, 879–894 (1997).
47. Holub, E. B. & Beynon, J. L. Symbiology of mouse-ear cress (*Arabidopsis thaliana*) and oomycetes. *Adv. Botanical Res.* **24**, 227–273 (1997).
48. Nel, M., Gu, X. & Sitnikova, T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl Acad. Sci. USA* **94**, 7799–7806 (1997).
49. Dodds, P. N., Lawrence, G. J. & Ellis, J. G. Identification of the *N* rust resistance gene of flax and analysis of the role of intragenic sequence exchange in the evolution of the complex *N* locus. *Plant J.* (in the press).
50. Krietman, M. & Akashi, H. Molecular evidence for natural selection. *Annu. Rev. Ecol. Syst.* **26**, 403–422 (1995).
51. Wang, G.-L. *et al.* *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**, 1–15 (1998).
52. Luck, J. E. *et al.* Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. *Plant Cell* **12**, 1367–1377 (2000).
53. Dawkins, R. & Krebs, J. Arms races between and within species. *Proc. R. Soc. Lond.* **205**, 489–511 (1979).
54. Zhu, Y. *et al.* Genetic diversity and disease control in rice. *Nature* **406**, 718–722 (2000).
55. Bevan, J. R., Clarke, D. D. & Crute, I. R. Resistance to *Erysiphe-fischeri* in two populations of *Senecio vulgaris*. *Plant Pathol.* **42**, 636–646 (1993).
56. Burdon, J. J., Thrall, P. H. & Brown, A. H. D. Resistance and virulence structure in two *Linum marginale-Melampsora lini* host-pathogen metapopulations with different mating systems. *Evolution* **53**, 704–716 (1999).
57. Muller, J. Fossil pollen records of extant angiosperms. *Bot. Rev.* **47**, 1–142 (1981).
58. Vision, T. J., Brown, D. G. & Tanksley, S. D. The origins of genomic duplications in *Arabidopsis*. *Science* **290**, 2114–2117 (2000).
59. Kato, A. *et al.* Isolation and analysis of cDNA within a 300 kb *Arabidopsis thaliana* genomic region located around the 100 map unit of chromosome 1. *Gene* **239**, 309–316 (1999).
60. Theologis, A. *et al.* Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature* **408**, 816–820 (2000).
61. Xiao, S. *et al.* Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by *RPW8*. *Science* **291**, 118–120 (2001).
62. Martin, G. B. *et al.* Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432–1436 (1993).
63. Song, W.-Y. *et al.* A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa-21*. *Science* **270**, 1804–1806 (1995).
64. Clark, S. E., Williams, R. W. & Meyerowitz, E. M. The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* **89**, 575–585 (1997).
65. Torii, K. U. *et al.* The *Arabidopsis ERECTA* gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell* **8**, 735–746 (1996).
66. Leonard, K. Modelling gene frequency dynamics. In *The Gene-for-Gene Relationship in Plant Parasite Interactions* (eds Crute, I., Holub, E. & Burdon, J.) 211–230 (CAB International, Wallingford, UK, 1997).
67. Edward, S. V. & Hedrick, P. W. Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol. Evol.* **13**, 305–311 (1998).
68. Delaney, D. E., Webb, C. A. & Hubert, S. H. A novel rust resistance gene in maize showing overdominance. *Mol. Plant-Microbe Interact.* **11**, 242–245 (1998).
69. Cuzick, A. Genetic characterisation of four genes in *Arabidopsis* required for a non-salicylate-dependent source of downy mildew resistance. Thesis, Imperial College of Biotechnology, Science and Medicine at Wye, University of London, UK (2001).

Acknowledgments

I am grateful for constructive reviews of the manuscript from four anonymous referees and helpful discussions from J. Dangi, J. Ellis, J. Jones, J. Burdon, M. Tor and D. Barbara. The author's work is supported largely by grants from the UK Biotechnology and Biological Sciences Research Council.