

# A cost of disease resistance: paradigm or peculiarity?

James K.M. Brown

Department of Disease and Stress Biology, John Innes Centre, Colney, Norwich NR4 7UH, UK

**Disease is one of the main driving forces of biological evolution. Parasites cause natural selection for disease resistance in populations of their hosts. Why then are all organisms susceptible to some parasites? One explanation is that resistance to disease is costly, reducing the fitness of the host in the absence of disease. A recent article shows that such costs might have helped to maintain polymorphism at a resistance locus. Other work, however, has questioned whether the costs of resistance are indeed necessary to account for polymorphism in host–parasite interactions.**

Plants are good models for studies of co-evolution because, unlike animals, their interactions with some important parasites are controlled by a fairly simple genetic system. In the gene-for-gene relationship [1], plants are resistant to some genotypes of a parasite species but not to others (Box 1). The asymmetry of this system has long fascinated evolutionary biologists because a resistance (*R*) allele should always have an advantage over a susceptibility allele by protecting the plants that have the *R* allele against pathogens that possess the matching avirulence (*Avr*) gene. How then is polymorphism for resistance and susceptibility maintained in plant populations? One possibility is that *R*-genes might impose metabolic costs on plants. A recent study by Joy Bergelson and colleagues [2] provides the first evidence for such a cost of a gene-for-gene resistance. This cost, balanced against the benefit of the *R*-gene in plant defence, is sufficiently high to have helped to maintain polymorphism at this *R*-locus for millions of years.

## Costly resistance

Polymorphism has persisted at the *RPM1* locus of *Arabidopsis thaliana* for at least 10 million years [3]. *RPM1* confers resistance to genotypes of the pathogenic bacterium *Pseudomonas syringae* pv *maculicola*, which express one of two *A*-genes, *AvrRpm1* or *AvrB*. Tian *et al.* [2] investigated whether a cost of resistance might have contributed to the longevity of this polymorphism. They produced four pairs of NEAR-ISOGENIC LINES (see Glossary) that were identical except at a transgenic ectopic *RPM1* locus, where they had integrated resistance alleles (*RPM1*<sup>+</sup>) or susceptibility alleles (*rpm1*; written as *RPM1*<sup>−</sup> in Ref. [2]). In the *RPM1*<sup>+</sup> lines, a 3.84-kb insert included *RPM1* itself, its promoter and its terminator and was flanked by *lox* sites. The expression and function of *RPM1* was phenotypically normal in these plants. Wild plants that are susceptible to *AvrRpm1* and *AvrB* isolates

of *P. syringae* pv *maculicola* lack the complete coding region of *RPM1* [4]. Thus, Tian *et al.* mimicked this allele via *cre*-mediated excision of the 3.84-kb insert at the *lox* sites to generate plants that possessed *rpm1*.

The effect of *RPM1* on the fitness of these plants was tested in a field trial in the absence of *P. syringae* pv *maculicola*. *RPM1*<sup>+</sup> plants were consistently smaller than *rpm1* plants, with smaller shoots, and had lower reproductive fitness, with fewer SILIQUES (seed pods) per plant and seeds per silique, resulting in 9% fewer seeds per plant in total (Figure 1). This high cost of resistance was consistent across the four pairs of transgenic lines, indicating that the cost of resistance was indeed associated with *RPM1* and was not an artefact of the position at which the transgene was inserted.

A resistance cost of 9% is surprisingly high. Indeed, it is probably not typical of gene-for-gene resistances. There are estimated to be 163 gene-for-gene *R*-genes in *A. thaliana* [5]. Thus, if *R*-alleles incurring a cost of 9% segregated at all these loci, the genetic load on the plant population would be unsustainable [2]. Gene-for-gene resistances are also used widely in plant breeding but any *R*-gene with such a high cost would be rapidly eliminated from breeders' germplasm.

There is circumstantial evidence from recent research on POWDERY MILDEW of barley that *R*-genes other than *RPM1* might incur costs. Transcripts of two *R*-genes, *Mla6* and *Mla13*, are induced between 16 and 20 h after inoculation, but only if the fungus has the corresponding *Avr* gene [6]. The fact that these *R*-genes are inducible

## Glossary

**Cultivar:** a variety of a crop plant, produced by plant breeding in modern agriculture or by natural selection among diverse genotypes in traditional farming systems.

**Downy mildew:** a plant disease that appears as a downy growth on leaves, flowers and fruit, caused by oomycetes in the order Peronosporaceae.

**Meristem:** a localized region of active cell division and differentiation in plants.

**Metapopulation:** a 'population of populations', in which an organism exists at several discrete sites. In each generation, the organism might become extinct at some sites and might be re-colonized from other sites.

**Near-isogenic lines:** a group of plant lines that are genetically identical except around one locus, produced by genetic engineering or by repeated backcrossing to a recurrent parent.

**Oomycete:** a group of organisms with an outward appearance similar to fungi, but actually part of the Stramenopile group of protists, which includes diatoms and brown algae among its better-known members.

**Powdery mildew:** a plant disease important mainly in humid, temperate environments, in which the fungal spores appear as a powder on leaves, stems and flowers, caused by ascomycetes in the order Erysiphales.

**Rust:** a common disease of many plants, in which pustules of spores break through the surface of leaves and stems, caused by basidiomycete fungi in the order Uredinales.

**Silique:** the elongated pod containing several seeds, typical of plants in the family Brassicaceae, including *Arabidopsis thaliana*.

### Box 1. The gene-for-gene relationship

The gene-for-gene relationship controls the responses of plants to some important parasites (Table I). Effective resistance is elicited only if the presence of an avirulence (*Avr*) gene in the parasite is recognized by a matching resistance (*R*) gene in the plant. *Avr* does not elicit effective defence if the plant lacks the matching *R*-gene altogether, has an ineffective, susceptibility allele of the *R*-gene or has an allele of the *R*-gene that recognizes a different *Avr* gene. If the pathogen lacks the *Avr* gene or has an ineffective allele, the plant does not sense the presence of the pathogen, whether or not the *R*-gene is functional, and so defence is not elicited or is ineffective. Different parasite *Avr* genes are detected by different host *R*-genes or by different alleles of a multi-allelic *R*-gene. There need only be one matching *R*-*Avr* gene pair for resistance to be effective. Most *R*-genes are dominant or semi-dominant, as are *Avr* genes. Here the parasite is shown as haploid, as is the case for ascomycete fungi and bacteria, two of the most important groups of plant pathogens.

Table I. Specificity of the gene-for-gene resistance

Plant	Parasite	
	Avirulent ( <i>Avr</i> )	Virulent ( <i>avr</i> )
Resistant ( <i>Rl</i> - )	Resistance	Disease
Susceptible ( <i>rlr</i> )	Disease	Disease

suggests that constitutive expression might be costly to the plant. By contrast, the *L6* gene of flax for resistance to RUST is expressed constitutively even after inoculation with an avirulent pathogen isolate [7], which tentatively suggests that its expression might not be costly in evolutionary terms.

*R*-genes that control gene-for-gene interactions are by no means the only genes involved in the defence of the plant against disease. An *R*-protein is a receptor molecule that induces effective defences if it perceives the presence of a parasite elicitor (e.g. the matching *Avr* protein). Many other genes are therefore involved in plant disease resistance, some of which are indeed costly to the

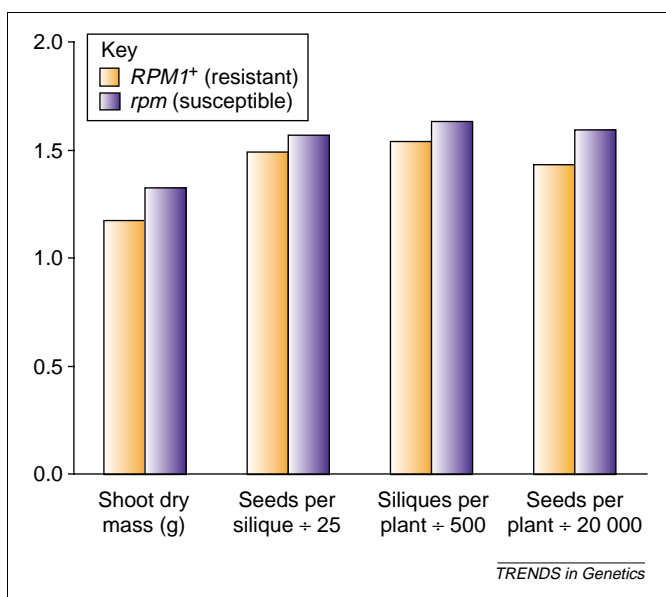


Figure 1. Costs of *RPM1* resistance in *Arabidopsis thaliana*, redrawn from Ref. [2]. In near-isogenic lines (NILs) produced by genetic engineering, those with the resistance gene (*RPM1*<sup>+</sup>; orange bars) had lower above-ground biomass, fewer seeds per silique, siliques per plant and total seeds per plant than susceptible plants (*rpm*; blue bars). These effects were consistent across four pairs of NILs [2]. Data shown are means across the four NILs.

<http://tigs.trends.com>

plant [8,9]. The question of whether or not gene-for-gene resistance is costly therefore does not necessarily have a bearing on the costs imposed by other types of resistance.

### More genes, higher cost?

Two different types of costs should be distinguished: adding a new *R*-gene and variation between the effective alleles of an existing *R*-gene that recognize the presence of different parasite *Avr* genes. The study by Tian *et al.* [2] is the first to investigate knowingly the cost of an additional *R*-gene. Previous studies that searched for costs of gene-for-gene resistance – all of which found none – probably investigated allelic variation in functional *R*-genes. The most comprehensive study used ten pairs of near-isogenic lines of the barley CULTIVAR Manchuria; in each pair one line had an *R*-gene that induced defence against powdery mildew whereas the other line did not. Each of these *R*-genes is effective against mildew fungi with different, matching *Avr* genes. These lines were developed using 12–15 backcrosses to Manchuria and so were very similar except for the sequence flanking the *R*-genes themselves. In no case did the *R*-gene significantly affect grain yield, grain size or a range of other traits [10]. Likewise, no costs were associated with *R*-genes that induced defence against mildew that segregated in populations of barley [11] or rye [12]. Three of the *R*-genes in Manchuria near-isogenic lines, *Mla1* [13], *Mla6* [14] and *Mla13* [6], have since been cloned. They have similar sequences and appear to be allelic at the *Mla* locus.

These data suggest that each new *R*-gene might be costly but allelic variation within an existing *R*-gene family might not. This could explain features of the organization of *R*-genes and alleles in the plant genome. On the one hand, *A. thaliana* has fewer than 200 NB-LRR (nucleotide-binding, leucine-rich-repeat) genes [5], the largest known class of *R*-gene. These account for only approximately one-thousandth of the genome, despite the vast range and diversity of potential parasites. If adding new *R*-genes were cost-free, one would expect massive proliferation of *R*-genes so that the plant could recognize as many of its enemies as possible. On the other hand, several *R*-genes have many alleles at the same locus, each of which corresponds to a different pathogen *Avr* gene. *Mla* in barley for resistance to powdery mildew is a well-known example of this [6,13,14]. More generally, many *R*-genes are clustered in certain regions of the genome and might have originated by gene duplication [15]. This suggests that generation of new alleles of existing *R*-genes is economical in evolutionary terms.

There could be a balance of the costs and benefits of adding a new *R*-gene compared with adding a new *R*-allele at an existing locus. In the first case, the total number of genotypes would be doubled but at the cost of reduced fitness. In the second case, there would be many fewer new genotypes, for example, 5% more if a new *R*-allele occurred at a locus that already had 20 alleles – but this increase would incur little or no cost to the plant.

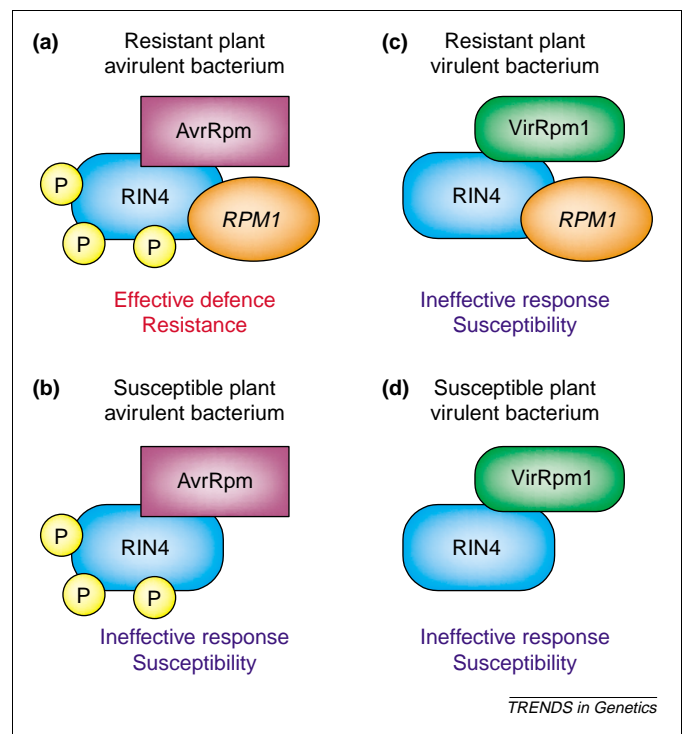
### Complex costs

There are several possible reasons why *RPM1* resistance is costly [2]. Overexpression of an *R*-gene under certain

environmental conditions might lead to inappropriate and damaging induction of plant defences [16]. Also, basal levels of expression of an *R*-gene in the absence of disease might induce plant defences to a certain level; although there is as yet no direct evidence for such induction, overexpression of some *R*-genes in the absence of the parasite causes activation of plant defences [9]. If these two effects are generic consequences of gene-for-gene resistance, however, it is doubtful that either would result in as large a cost as that found by Tian *et al.* [2], because of the physiological load of the combined effect of many *R*-genes.

Alternatively, as Tian *et al.* note [2], a cost could arise not from the presence of *RPM1* alone but from the interaction of the *RPM1* protein with other plant proteins. Evidence is emerging that *R*-proteins do not recognize Avr proteins directly, but through the interaction of Avr with other plant proteins, known as pathogenicity targets (PTs). Effectively, *R*-proteins guard PT proteins against damage by pathogens [17,18]. The only PT for which there is direct evidence for a tri-partite interaction (*R* + PT + Avr) is *RPM1* interacting protein (RIN4) [19–21], one of several *A. thaliana* proteins with which *RPM1* is associated (Figure 2). RIN4 is required for normal seedling growth, MERISTEM function and fertility [19]. The presence of the corresponding Avr-proteins, AvrRpm1 and AvrB, leads to phosphorylation of RIN4. This is recognized by *RPM1*, resulting in the activation of the host defences [19]. *RPM1* also interacts with another PT, AtTIP49a, as does recognition of *Peronospora parasitica* 5 (RPP5), the product of an *R*-gene against DOWNY MILDEW, caused by the OOMYCETE *Peronospora parasitica*. Like RIN4, AtTIP49a is required for normal development, being essential for meristem function, flower formation and seed set [22]. AtTIP49a also suppresses RPP5-mediated resistance to *P. parasitica*, such that plants in which AtTIP49a transcription is reduced are more resistant to the parasite. However, reduction of AtTIP49a levels did not noticeably enhance *RPM1*-mediated resistance to *P. syringae* pv *maculicola*, although it might have been difficult to detect any enhancement because this resistance is already strong in plants with wild-type AtTIP49a [22].

AtTIP49a interacts with the *R*-genes *RPM1* and *RPP5*, downregulates RPP5-dependent resistance and is essential for normal meristem development. *RPM1* also interacts with RIN4, which is also essential for plant growth and development, and the cost of *RPM1* was measured in terms of seed production. It is therefore reasonable to speculate that the cost of resistance identified by Tian *et al.* [2] might result from *RPM1* interfering with the activity of the developmentally important proteins AtTIP49a, RIN4 or both, so that less seed is produced than in *rpm1* plants. A test of this hypothesis would require comparative functional analysis of RIN4 in *RPM1*<sup>+</sup> and *rpm1* plants and has not yet been done. Whether or not this particular hypothesis turns out to be correct, it could well be that costs of resistance are specific to individual *R*-genes and to the proteins they guard. The nature and size of the cost, indeed the existence of any cost at all, might be specific to each *R*-gene. Indeed, there might be no general model for costs of gene-for-gene resistance.

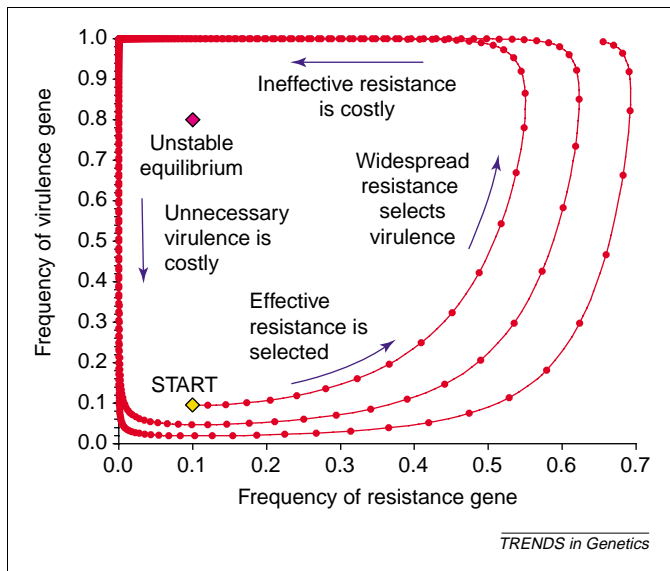


**Figure 2.** The guard hypothesis for gene-for-gene interactions [17], illustrated by the plant resistance (*R*) gene *RPM1* and the pathogen avirulence (*Avr*) gene *AvrRpm1*. *RPM1* interacting protein (RIN4) is required for normal development of *Arabidopsis thaliana* and is the pathogenicity target (PT) for the Avr protein AvrRpm1 of the bacterial pathogen *Pseudomonas syringae* pv *maculicola*. *RPM1* 'guards' RIN4 and is the *R*-protein in the gene-for-gene relationship, corresponding to AvrRpm1 [19]. Interaction with AvrRpm1 results in phosphorylation of RIN4. (a) In a resistant plant, RIN4 phosphorylation is recognized by *RPM1*, which induces effective defence against *P. syringae* pv *maculicola*, so the plant is resistant to disease (top left). (b) If *RPM1* is absent, RIN4 phosphorylation by interaction with AvrRpm1 is not recognized, so host defences are not induced and disease results (bottom left). (c),(d) If AvrRpm1 is modified or absent and RIN4 is therefore not phosphorylated, so that *P. syringae* pv *maculicola* is virulent, *P. syringae* pv *maculicola* is not recognized whether *RPM1* is (c) present or (d) not. *RPM1* also guards RIN4 against another *P. syringae* pv *maculicola* Avr protein, AvrB [19], and RIN4 is also the PT for yet another Avr protein, AvrRpt2. The *R*-protein that guards this protein is RPS2 [20,21].

### Are costs necessary?

Population geneticists have recently questioned whether there is any need for costs of resistance for polymorphism to be maintained in gene-for-gene systems. Figure 3 depicts a standard, simple model of cycles of *R*- and *V*-gene frequencies [23]. This and most other theories of gene-for-gene coevolution include a cost of host resistance because, without it, the *R*-allele would become fixed in the host population and the virulence allele of the *Avr* gene would be fixed in the parasite population. This model is unrealistic in several respects. First, the equilibrium point is unstable, so that trajectories of gene frequencies cycle outwards and polymorphism is not maintained indefinitely. Second, the system comes so close to the boundaries, especially at the top and left of the diagram, that it is likely that alleles are lost quickly from a population of finite size. Third, rapid loss of alleles can only be avoided if the costs of resistance and virulence are unrealistically high. These problems have inspired several developments of the model to explain what factors might contribute to maintaining polymorphism, notably sub-division of host populations [24].

In Figure 3, selection and near-elimination of susceptibility and avirulence alleles is rapid because dispersal is



**Figure 3.** Dynamics of gene frequencies in a simple model of gene-for-gene interactions between a plant and a parasite [23]. In each generation, each plant is attacked by one parasite and all parasites have an equal chance of attacking each plant. A resistance gene for which the matching parasite avirulence is common (i.e. the virulence allele is rare) has a selective advantage because it protects the host against the majority of parasites (lower part of the cycle). As it becomes common, it selects for virulence in the parasite (right) – or more precisely, against functional avirulence. A resistance gene that is no longer effective against the pathogen falls in frequency because it is costly (top of the cycle), then the cost of the virulence that is no longer required to overcome host resistance causes it to be removed from the parasite population (left of the cycle). Here, the cost to the plant of being diseased ( $s$ ) is 0.25 and the cost of having the resistance gene ( $t$ ) is 0.05. The cost to an avirulent pathogen of being unable to infect a resistant host ( $c$ ) is 0.9 and the cost of being virulent ( $d$ ) is 0.09. The trajectories of gene frequencies spiral around and away from an unstable equilibrium point where the frequency of resistance is  $d/c = 0.1$  and the frequency of virulence is  $(s - t)/s = 0.8$ . The model was run for 500 generations.

panmictic, such that the entire pathogen population is dispersed over the entire plant population in each generation. Avirulent pathogens are therefore removed because they rapidly encounter resistant plants, whereas susceptible plants have reduced reproduction because they are rapidly attacked by virulent pathogens. By contrast, Thrall and Burdon [25] investigated a METAPOPULATION model with restricted dispersal. Each sub-population comprises plants with diverse resistance genotypes, infected by pathogens with diverse virulences. The greater the severity of disease in a sub-population, the more plants are killed. Dispersal of seeds between sub-populations allows re-colonization of sites where plants have been exterminated by disease, whereas spore dispersal allows re-infection of recently re-colonized sub-populations. The importance of the spatial scale of dispersal for population dynamics of plants and pathogens that interact by gene-for-gene relationships has been demonstrated experimentally for flax rust [26].

In the metapopulation model [25], high polymorphism in both  $R$ - and  $A$ -genes was maintained even if resistance and virulence incurred no direct costs. Instead, a different kind of cost operates. Plant populations with low diversity in  $R$ -genotypes are vulnerable to attack by few pathogen genotypes, whereas extermination of plants where the population is uniform causes the death of the pathogens on those plants. The costs are therefore not of resistance or

virulence as such, but of uniformity of resistance and virulence. In an earlier metapopulation model, it was proposed that the cost of uniformity at the  $R$ -locus could arise from a lack of diversity for resistance to another parasite [27]. If this is true, it means that plant biologists must extend their horizons even beyond complex interactions within a single plant to complex interactions within entire ecosystems – a daunting prospect.

### Must resistance be costly?

Is the cost of  $RPM1$  a paradigm or a peculiarity? On the one hand, long-lived alleles have been detected in several  $R$ -genes in *A. thaliana* [28], and so mechanisms must be sought to account for durable polymorphism in host-parasite interactions. The data of Tian *et al.* [2] imply that a cost of resistance might be part of the story, helping to maintain polymorphism in the  $RPM1$  gene for millions of years. On the other hand, the cost of  $RPM1$  is so high that it is most unlikely that all the hundred or so  $R$ -genes in *A. thaliana* have similar costs. The complex interactions of  $RPM1$  with other proteins suggest that the costs of  $RPM1$  might not be typical of  $R$ -genes in general unless all  $R$ -proteins interact with many others. Moreover, costs of resistance might not actually be required to maintain polymorphism in realistic, spatially structured populations. A working hypothesis, therefore, is that the existence of a cost of resistance and the nature and size of that cost might be properties of each individual  $R$ -gene and that variation between effective  $R$ -alleles at the same locus is cost-free, whereas the cost of uniformity of resistance might be a general property of populations of all host organisms, animals and plants.

### References

- 1 Flor, H.H. (1971) The current status of the gene for gene concept. *Annu. Rev. Phytopathol.* 9, 275–296
- 2 Tian, D. *et al.* (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423, 74–77
- 3 Stahl, E.A. *et al.* (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400, 667–671
- 4 Grant, M.R. *et al.* (1995) Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* 269, 843–846
- 5 Palomino, M.M. *et al.* (2002) Patterns of positive selection in the complete *NBS-LRR* gene family of *Arabidopsis thaliana*. *Genome Res.* 12, 1305–1315
- 6 Halterman, D.A. *et al.* (2003) Powdery mildew-induced *Mla* mRNAs are alternatively spliced and contain multiple upstream open reading frames. *Plant Physiol.* 131, 558–567
- 7 Ayliffe, M.A. *et al.* (1999) Analysis of alternative transcripts of the flax *L6* rust resistance gene. *Plant J.* 17, 287–292
- 8 Bergelson, J. and Purrington, C.B. (1996) Surveying patterns in the cost of resistance in plants. *Am. Nat.* 148, 536–558
- 9 Brown, J.K.M. (2002) Yield penalties of disease resistance in crops. *Curr. Opin. Plant Biol.* 5, 339–344
- 10 Jørgensen, J.H. and Jensen, H.P. (1990) Effect of ‘unnecessary’ powdery mildew resistance genes on agronomic properties of spring barley. *Norsk Landbruksforskning* (Suppl. 9), 125–130
- 11 Kjær, B. *et al.* (1991) Associations between 23 quantitative traits and 10 genetic markers in a barley cross. *Plant Breed.* 106, 261–274
- 12 Welz, H.G. *et al.* (1995) Two unnecessary powdery mildew resistance genes in a synthetic rye population are neutral on fitness. *Euphytica* 81, 163–170
- 13 Zhou, F.S. *et al.* (2001) Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a *Rar1*-independent signaling pathway. *Plant Cell* 13, 337–350
- 14 Halterman, D. *et al.* (2001) The *MLA6* coiled-coil, *NBS-LRR* protein

- confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f.sp. hordei in barley and wheat. *Plant J.* 25, 335–348
- 15 Holub, E.B. (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* 2, 516–527
- 16 Wang, Z.-W. *et al.* (2001) Expression of the *Pib* rice-blast-resistance gene family is up-regulated by environmental conditions favouring infection and by chemical signals that trigger secondary plant defences. *Plant Mol. Biol.* 47, 653–661
- 17 van der Biezen, E.A. and Jones, J.D.G. (1998) Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem. Sci.* 23, 454–456
- 18 Dangl, J.L. and Jones, J.D.G. (2001) Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–832
- 19 Mackey, D. *et al.* (2002) RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* 108, 743–754
- 20 Axtell, M.J. and Staskawicz, B.J. (2003) Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* 112, 369–377
- 21 Mackey, D. *et al.* (2003) *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulated RPS2-mediated resistance. *Cell* 112, 379–389
- 22 Holt, B.F. *et al.* (2002) An evolutionarily conserved mediator of plant disease resistance gene function is required for normal *Arabidopsis* development. *Dev. Cell* 2, 807–817
- 23 Leonard, K.J. (1977) Selection pressures and plant pathogens. *Ann. New York Acad. Sci.* 287, 207–222
- 24 Leonard, K.J. (1977) Modelling gene frequency dynamics. In *The Gene-for-Gene Relationship in Plant-Parasite Interactions* (Crute, I.R. *et al.*, eds), pp. 211–230, CAB International
- 25 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, 169–194
- 26 Thrall, P.H. *et al.* (2003) Influence of spatial structure on pathogen colonization and extinction: a test using an experimental metapopulation. *Plant Pathol.* 52, 350–361
- 27 Damgaard, C. (1999) Coevolution of a plant host-pathogen gene-for-gene system in a metapopulation model without cost of resistance or cost of virulence. *J. Theor. Biol.* 201, 1–12
- 28 Bergelson, J. *et al.* (2001) Evolutionary dynamics of plant R-genes. *Science* 292, 2281–2285

0168-9525/\$ - see front matter © 2003 Elsevier Ltd. All rights reserved.  
doi:10.1016/j.tig.2003.10.008

## Tracing the dawn of *Plasmodium falciparum* with mitochondrial genome sequences

David J. Conway

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street London WC1E 7HT, UK

**Until recently, little light had been shed on the murky origins of human malaria. Did *Plasmodium falciparum*, the most virulent malaria parasite, emerge as a common pathogen only in the past few thousand years, as suggested by some analyses of its nucleotide sequence diversity? Or, was it an ancient scourge of early humans >100 000 years ago, as suggested by others? A recent study, using complete mitochondrial DNA sequence polymorphism data and new analytical methods, points to an intermediate date of origin and expansion out of Africa. Subsequent population growth in each continent is less well resolved.**

*Plasmodium falciparum*, which causes malaria, has probably resulted in more human mortality than any other infectious agent and has a continued impact today. Thus, it is important to understand the evolution of this pathogen. Analyses of nucleotide sequence diversity in *P. falciparum* have been used to support divergent views on its origin. Very low levels of polymorphism at synonymous nucleotide sites [1] and within introns [2] of some protein coding genes indicated that the global *P. falciparum* parasite population might have originated from a very small founding population only a few thousand years ago. By contrast, further analysis of public database sequences indicated that nucleotide polymorphism was more extensive, supporting a view that the parasite has had a large effective population size for hundreds of

thousands of years [3]. Analysis of complete sequences of the parasite mitochondrial genome (mtDNA) yielded an intermediate estimate of tens of thousands of years [4]. These contrasting data were discussed previously [5–7], although differing interpretations remained. More recently, a substantial dataset has been generated on most of the genes contained on chromosome 3 (one of the 14 nuclear chromosomes of the parasite), surveyed in five different *P. falciparum* isolates, and this indicates that some of the nucleotide diversity traces back to ~100 000 years ago [8]. However, it has also been shown that most of the polymorphism in two closely related paralogous genes on chromosome 11 has probably been generated by gene conversion between the paralogues. Furthermore, it has been speculated that such processes could be widespread in the genome and could inflate estimates of allelic divergence [9]. Most protein-coding genes in the *P. falciparum* genome are single loci (and do not have adjacent closely related paralogues) [10], as are their orthologues in the genomes of other *Plasmodium* species [11]. Thus, it is not clear whether gene conversion should be of widespread concern or if it greatly challenges the analysis of chromosome 3 diversity [8]. Given such uncertainty, there is potential merit in independent approaches to trace the evolutionary history of *P. falciparum*.

### Roots and shoots

Joy and colleagues [12] studied the six-kilobase mtDNA because it is uni-parentally inherited (only through the

Corresponding author: David J. Conway (david.conway@lshtm.ac.uk).