Genetic architecture and pleiotropy shape costs of Rps2-mediated resistance in Arabidopsis thaliana

Alice MacQueen¹, Xiaqin Sun¹ and Joy Bergelson*}

The mounting evidence that R genes incur large fitness costs raises a question: how can there be a 5–10% fitness reduction for all 149 R genes in the Arabidopsis thaliana genome? The R genes tested to date segregate for insertion–deletion (indel) polymorphisms where susceptible alleles are complete deletions. Since costs of resistance are measured as the differential fitness of isolines carrying resistant and susceptible alleles, indels reveal costs that may be masked when susceptible alleles are expressed. Rps2 segregates for two expressed clades of alleles, one resistant and one susceptible. Plants with resistant Rps2 are not less fit than those with a susceptible Rps2 allele in the absence of disease. Instead, all alleles provide a fitness benefit relative to an artificial deletion because of the role of RPS2 as a negative regulator of defence. Our results highlight the interplay between genomic architecture and the magnitude of costs of resistance.

U
nderstanding how plants maximize fitness in response to intermittent pathogen presence is of central importance in plant pathology. Natural plant pathosystems frequently involve long-maintained polymorphisms for host resistance1–3 and current theory on the maintenance of stable resistance polymorphisms requires costs of resistance and/or virulence acting in combination with frequency-dependent selection. In contrast, in agricultural contexts, resistance genes often have useful lifespans of only a few years and high costs of resistance are undesirable because of their negative effects on plant performance. Pathogen resistance may involve two distinct fitness costs. The first, which we term a cost of surveillance, accrues from harbouring R genes that allow a resistance response on attack, and the second, a cost of defence, accrues from activation of the resistance response during attack.

A cost of surveillance is measured in the absence of disease and has been determined for two R genes, Rps5 and Rpm1, in Arabidopsis thaliana. Rps5 exists in nature as a long-lived insertion–deletion polymorphism (indel) for resistance (R) and susceptibility. Rpm1 similarly exists as a long-lived indel polymorphism, although secondarily disrupted S alleles are also present within the resistance clade. In both cases, resistant isolines suffer a 5–10% fitness cost relative to null isolines. However, costs of this magnitude would correspond to an impossibly high genetic load if seen for all, or many, of the ~149 R genes in A. thaliana. We propose that indels are an architecture with unusually high costs of surveillance. Null alleles of indels cannot carry any burden of mis-expression or mis-activation that would reduce relative costs of carrying resistant versus susceptible alleles. Furthermore, null alleles do not have the potential to evolve pleiotropic or alternative functions, another means of ameliorating costs of surveillance. R genes are found with a great diversity of genetic architectures, including single loci with many, functional alleles and arrays of tandem duplicated R genes. Here, we explore the possibility that R gene genetic architectures with alternative functional alleles have substantially smaller costs than those of the indels Rpm1 and Rps5. We posit that the large costs associated with these two R genes are precisely the reason susceptible alleles are deleted.

Rps2 exists as an ancient balanced polymorphism with two long-lived clades of alleles, one resistant and one susceptible to Pseudomonas syringae pv. avrRpt2. Both clades are maintained at intermediate frequencies in local populations. Rps2 is also present in every accession sequenced to date. (Out of 300 accesses with sequencing data (see Supplemental methods), none has missing data or deletions called for Rps2.) To measure the surveillance cost associated with resistant alleles of Rps2, we assayed fitness in the absence of disease for a transgenically created allelic series of Rps2 that controls for the insertion site of the R genes. We extended the basic strategy of using a Cre-lox system to compare isolines with and without a resistant allele (as done for Rpm1, ref. 8 and Rps5, ref. 9) to precisely integrate five alleles of Rps2 into the same genomic location. These alleles were inserted into a Col-0 genetic background in which Rps2 had been knocked out. We additionally verified the robustness of our results by using three genomic locations for the insertion of Rps2. Our results reveal a bidirectional interplay between genomic architecture and fitness costs of resistance.

Results
No cost of surveillance for R alleles relative to S alleles of Rps2. To test for surveillance costs of Rps2 alleles, we measured the lifetime fitness of three resistant (Rps2R) and one partially resistant (Rps2S) isolines relative to one susceptible (Rps2S) allele at each of three genomic insertion sites in a field experiment performed in the absence of pathogens carrying avrRpt2 (Fig. 1, Supplementary Figs. 1 and Supplementary Table 1). After correction for multiple testing, there were no significant differences in the fitness proxies of isogenic plants carrying Rps2R and Rps2S alleles at the same insertion site (Supplementary Tables 2–4). Significant differences in correlations between fitness proxies might reveal differential resource allocation across resistance classes; however, such differences in fitness proxy correlations were not evident between Rps2R, Rps2S and Rps2R lines after a multiple testing correction (Fisher’s z-test, NS). Two fitness proxies supported higher, rather than lower, fitness for Rps2R alleles than Rps2S alleles.
Figure 1 | Natural variation in Rps2 captured by the transgenic allelic series. Red and black bars indicate the susceptible and resistant clades of Rps2 alleles, respectively. A, The two clades of Rps2 alleles were inferred from the coding sequence of Rps2 using maximum likelihood, and included 80 genomes from Cao et al. and Sanger sequencing of the five alleles used in this study. The percentage of trees in which the associated taxa clustered together out of 1,000 bootstrap replicates is shown next to the branches. B, Amino acid variation in the alleles used in this study. Cyan bar indicates the leucine-rich repeat region of Rps2; Rps2 and Rps2 alleles are resistant and partially resistant to Pseudomonas syringae pv. avirRpt2.
Expression levels can alter the penetrance of phenotypes and overexpression of Rps2 can lead to non-specific activation of the hypersensitive response (HR), or even lethality, if expression levels are too high\(^2,3\). The expression of each allele varied within and across insertion sites when transgenic plants were measured in growth chamber conditions (Supplementary Fig. 2), and in many cases Rps2 expression was two- to fourfold higher in the isolines than in the accessions from which the alleles derived (Supplementary Fig. 3). To investigate the relationship between fitness variation and Rps2 expression, we modelled lifetime seed production in the field as a linear function of average Rps2 expression at 23 days for each allele nested within allele class (Fig. 2f and Supplementary Table 5). We also included as a covariate the date at which each plant was collected from the field. For both the Rps2\(^R\) and Rps2\(^S\) classes, basal expression of Rps2 was negatively correlated with lifetime fitness (Fig. 2f; \(F = 9.666; d.f. = 8, 602; R^2 = 0.114, P = 0.0028, 0.024\)), although the negative correlation for the Rps2\(^R\) class was not significant (\(P = 0.73\)). These results indicate that Rps2 overexpression is costly in the absence of pathogens for both \(R\) and \(S\) clade alleles. Adding Rps2 expression into the fitness model did not reveal differences in performance that were masked by differences in expression between Rps2\(^R\), Rps2\(^S\) and Rps2\(^S\) lines (Supplementary Table 5).

An artificial Rps2 knockout is significantly less fit in the absence of pathogens. If resistant and susceptible alleles have equivalent fitness in the absence of pathogens, then the benefit of Rps2\(^R\) resistance should have driven the Rps2\(^R\) clade to fixation. However, both clades have been maintained for millions of years in \(A. thaliana\)^14. We hypothesized that Rps2\(^R\), and perhaps Rps2\(^S\), alleles must have another beneficial function to permit maintenance of the Rps2\(^S\) clade. To explore this possibility, we compared the fitness of all lines with an expressed allele of Rps2 (collectively, Rps2\(^S\)) to three artificial knockout isolines (Rps2\(^KO\)), using the same field experiment performed in the absence of pathogens carrying avrRpt2. In 19 out of 21 comparisons using seven fitness proxies for each of the three insertion sites, Rps2\(^KO\) isolines demonstrated higher performance than Rps2\(^KO\) isolines, although after correction for multiple testing, only five of these instances were significant (Supplementary Tables 6–8). In terms of lifetime seed set, Rps2\(^KO\) individuals suffered up to a 54% reduction relative to Rps2\(^S\) isolines (Fig. 2a–c); this fitness cost was significant for the five lines with the lowest Rps2 expression. This pattern follows from the negative correlation between basal Rps2 expression and lifetime fitness in both Rps2\(^R\) and Rps2\(^S\) allele classes (Fig. 2f), such that highly expressing \(A. thaliana\) lines suffered a fitness reduction much like that observed in the knockout lines. Interestingly, the comparisons revealing a significant cost of the Rps2 knockout include Rps2\(^S\) lines with expression levels most similar to native accessions (Supplementary Fig. 3 and Fig. 2f). Rps2\(^KO\) individuals also had a significantly weaker correlation between plant weight and total collected seed (Fisher’s \(z\) test \(P = 4.8 \times 10^{-9}\) ) than plants with any other allele of Rps2 (correlation coefficient of 0.85 compared to all \(>0.93\)); this pattern was driven by a number of Rps2\(^KO\) plants with much lower seed sets than expected for their weight. Taken together, these results suggest that low expression of any allele of Rps2 is beneficial in the absence of known Rps2-mediated pathogens carrying avrRpt2.

We considered two hypotheses to explain the observed benefit of all Rps2 alleles in the absence of \(P. syringae\) pv. \(avrRpt2\). First, the presence of a different, and undetected, pathogen recognized by all alleles of Rps2 in the field may have provided a benefit to isolines carrying alleles susceptible to \(avrRpt2\). Alternatively, a pleiotropic function for Rps2 in the absence of disease may have contributed to its benefit. To discriminate between these hypotheses, we first repeated our fitness experiment in a growth chamber that mimicked the stressful environmental conditions of our field environment but was known to be free of RPS2-recognized pathogens. Owing to size constraints of the growth chamber, we used isolines from only one insertion site in this experiment. As in the field experiment, after correction for multiple testing, there were no significant differences in fitness proxies between Rps2\(^R\) or Rps2\(^S\) lines relative to Rps2\(^S\) lines (Fig. 2d and Supplementary Table 9; \(F = 9.20, d.f. = 9, 1,009, P > 0.003\)). Again, Rps2\(^S\) lines set significantly more seed than Rps2\(^KO\) lines (Supplementary Table 10; \(F = 39.8, d.f. = 2, 875, P = 0.006\)). Thus, the growth chamber results recapitulated the results seen in the field.

As a final confirmation that the observed fitness difference was not due to an interaction with an unknown microbe, we grew our isolines from one insertion site in sterile conditions on agar. Again, Rps2\(^S\) plants had a higher weight than Rps2\(^KO\) plants at 21 days (Fig. 2e; \(F = 9.63, d.f. = 114, P = 0.0024\)). This result excluded the possibility that the presence of Rps2 carried a fitness benefit because of recognition of pathogens.

Our isolines were all created using a Cre-lox system in the rps2-101C mutant background. It should be noted that, although this background is often used as an Rps2 null\(^7,19–26\), it nonetheless produces a 235 amino acid, N-terminus fragment of an RPS2 protein which contains the entirety of the RPS2 coiled-coil domain. In addition, our constructs could have antisense transcription of RPS2, as is seen for the native copy of RPS2, or ectopic expression of the last three exons of \(At4g26100\). Although previous work has demonstrated that truncation mutants of RPS2 are not autoactive\(^20,27\) and do not interact with typical RPS2-interacting proteins\(^20\), we sought to test that this truncated RPS2-101C protein, antisense transcript and/or the presence of nptII and lox sites did not contribute to plant performance. To do this, we compared the fitness of a transgenic Rps2\(^KO\) line, created in the rps2-101C background and containing lox sites at insertion site 2, with an independently created Col-0 artificial microRNA (amiRNA) knockdown of Rps2 grown under sterile conditions (Supplementary Fig. 4). There were no significant differences in weight between these two susceptible lines (\(P = 0.65\)). We additionally failed to detect a difference in the weight of two resistant lines, Col-0 and a transgenic Rps2\(^KO\) line with the Col-0 allele at insertion site 2 (\(P > 0.45\)). Thus, neither the rps2-101C background and its associated truncated RPS2 protein nor the lox sites and nptII impact plant performance directly or through an interaction with Rps2. Furthermore, both Rps2\(^S\) plant genotypes had significantly higher weight than both genotypes without Rps2 (\(P < 0.0005\)) when grown under sterile conditions. A comparison of Rps2 isolines not created in the rps2-101C background confirmed the growth benefit of Rps2; the Col-0 amiRNA knockdown line was significantly less fit than Col-0 (\(P = 0.0026\)). Thus, removal of RPS2 protein is associated with reduced performance both in the rps2-101C and Col-0 backgrounds. In combination, these results demonstrate a beneficial pleiotropic function of Rps2 measurable in stressful abiotic environments in the absence of pathogens.
Figure 2 | Significant fitness variation among lines in the allelic series in the absence of pathogen. KO is the rps2-101C mutant with empty lox sites at the insertion site for that line, referred to in the text as the Rps2KO line; 101C is the Rps2 null mutant without an inserted lox site. Lines with Rps2 inserted at three genomic locations, or insertion sites, were tested. Black bar is under Rps2R lines, grey bar is under Rps2pR lines, red bar is under Rps2S lines and white bar is under Rps2KO lines. Letters above bars within each insertion site indicate grouping using Tukey’s post hoc test. Lines with ‘NT’ were not grown in this experiment. 

a, Field fitness results. a, Field fitness for insertion site 1. The isoline with the Col-0 allele at insertion site 1 did not germinate. 

b, Field fitness for insertion site 2. 

c, Field fitness for insertion site 3. 

d, Growth chamber fitness for insertion site 2. 

e, Sterile condition fitness for insertion site 2. The isoline with the Ler-0 allele did not germinate in this experiment. 

f, Average Rps2 expression at three weeks is negatively correlated with fitness in the field. The x-axis shows unitless relative expression of the isogenic lines (points) and the native accessions (vertical dotted line). Black and red dashed lines are the regression lines for the R and S clades, respectively, for the relationship between fitness and expression nested in allelic class. Black points are Rps2R lines from the resistant clade of Rps2, and red points are Rps2S lines from the susceptible clade of Rps2. The average fitness of the Rps2 knockouts is plotted at the horizontal dotted line.
was minimal (Supplementary Fig. 5). The \textit{Rps2} line had 14 genes that were upregulated and two genes that were downregulated relative to the \textit{Rps2S} line (Supplementary Fig. 5). These genes were enriched for gene ontology (GO) annotations of response to stress, particularly for response to water stress (Supplementary Table 11; \( P = 1.24 \times 10^{-5} \)).

The field fitness data displayed two patterns that we further explored with transcriptome data. First, we observed that \textit{Rps2} knockout lines had lower lifetime fitness than \textit{Rps2+} lines, essentially irrespective of \textit{Rps2} expression level (Fig. 2f). Second, given \textit{Rps2} presence, there was an inverse relationship between \textit{Rps2} expression level and fitness (Fig. 2f).

We explored the first of these two patterns, of \textit{Rps2} presence or absence, by determining the expression profiles of three \textit{Rps2} lines and contrasting them with the \textit{Rps2KO} line. \textit{Rps2+} lines upregulated 538 genes relative to \textit{Rps2KO} lines and downregulated 312 genes. Genes upregulated in \textit{Rps2+} plants were enriched for GO annotations involving photosynthesis and light response (Fig. 3a and Supplementary Table 12; \( P = 2.47 \times 10^{-3}, 3.91 \times 10^{-5} \)), whereas genes downregulated in \textit{Rps2+} plants were enriched for stimulus response, stress response, biotic stimulus response and defence response annotations (Fig. 3b,c and Supplementary Table 13; \( P = 1.18 \times 10^{-16}, 3.57 \times 10^{-16}, 7.41 \times 10^{-11}, 9.01 \times 10^{-11} \)).

**Figure 3** | \textit{Rps2} knockout lines differentially express stress response, defence response and growth-related genes relative to all lines with an allele of \textit{Rps2}. \textit{R} clade and \textit{S} clade are resistant and susceptible \textit{Rps2} lines, Col-0 and Wu-0, with similar levels of expression from insertion site 2, and high \textit{R} is the Col-0 allele of \textit{Rps2} from insertion site 3, which has a higher level of \textit{Rps2} expression. \textbf{a-}c. Heatmaps and dendrograms of gene sets as described below. Genes are in rows, and biological replicates are in columns, with both dendrograms grouped by similarity of expression in the gene set displayed. KO is the \textit{rps2}-101C mutant with empty lox sites at the insertion site. 

\textbf{a.} Differentially expressed genes with GO annotations related to photosynthesis or response to light stimulus. 

\textbf{b.} Differentially expressed genes with GO annotations of response to stimulus or response to stress. 

\textbf{c.} Differentially expressed genes with GO annotations of defence response or response to biotic stimulus. 

\textbf{d.} The overlap of differentially expressed genes for three contrasts of lines with an \textit{Rps2} allele and the knockout. Orange values are upregulated and blue are downregulated relative to the knockout.
We found substantial overlap in genes differentially expressed in independent contrasts of single Rps2\(^{-}\) lines and the Rps2\(^{KO}\) line (Fig. 3d); genes identified in each single-line comparison were consistently enriched for the same GO annotations (Fig. 3d and Supplementary Tables 14–19). Thus, removing both R and S alleles of Rps2 from A. thaliana predominantly increased expression of genes that are induced in response to stress and pathogens.

We explored the second of these two patterns, Rps2 expression level, by comparing the expression profiles of two Col-0 Rps2\(^{+}\) lines with high and low Rps2 expression under sterile conditions. Expression of 36 genes was upregulated in the high Rps2 expression line relative to the low Rps2 expression line, and expression of 189 genes was downregulated (Fig. 4a). Upregulated genes in the high Rps2 expression line were enriched for genes involved in response to stimulus, particularly for thalianol metabolic processes (Fig. 4a and Supplementary Table 20; \(P = 1.05 \times 10^{-5}\)). Downregulated genes in the high Rps2 expression line were enriched for response to stimuli and stress, particularly for response to biotic stimulus and for the defence response (Fig. 4a,b and Supplementary Table 21; \(P = 3.38 \times 10^{-10}, 1.75 \times 10^{-3}, 1.67 \times 10^{-3}, 4.03 \times 10^{-4}\)). In contrast to the previous comparison between Rps2\(^{+}\) and Rps2\(^{KO}\), defence-related genes were enriched in the downregulated gene set in the line with lower field fitness, or the line with higher Rps2 expression level, rather than upregulated as in the Rps2 knockout. The set of downregulated genes contained a subset that was differentially expressed in only the high Rps2 expression line; this gene subset was also enriched for GO categories of response to stress and the defence response (Fig. 4c and Supplementary Table 22; \(P = 1.04 \times 10^{-2}, 1.28 \times 10^{-2}\)). Rps2 overexpression therefore downregulated an additional, unique set of defence response genes than those induced in Rps2 null mutants (Fig. 4b,c). Thus, we observe enrichment for defence related genes that distinguish the highly fecund, low Rps2 expression line from both the Rps2 knockout and high Rps2 expression lines, although the genes that contribute to these enrichments are different.

**Discussion**

Costs of resistance contribute to the long-term maintenance of polymorphism at defence genes because they help explain the persistence of susceptible alleles. However, it is hard to understand why the production of minute quantities of a recognition protein, such as those produced by R gene loci, would entail a large physiological cost. It was therefore surprising that our emerging picture of the costs associated with R gene loci is that they are large, on the order of 10%\(^8,9\). Here, we report on the absence of a cost associated with Rps2, an R gene that segregates for the maintenance of alternative alleles\(^16\). This result makes sense in that both R and S alleles are expressed in our Rps2 isolines; thus, the difference between R and S genotypes is small relative to the difference in isolines segregating for indel polymorphisms, as in the previous R genes for which costs have been measured\(^8,9\). Since a large fraction of R genes in the genome harbour multiple alleles\(^28,29\), our results help explain how host genomes can tolerate the genetic load associated with R gene resistance. We suggest that whereas stable indel polymorphisms may be maintained by large costs of resistance, stable non-indel R gene polymorphisms are more likely to be maintained by a variety of ecological and physiological mechanisms, as elaborated below. Thus, this work reveals a fundamental effect of genetic architecture on the manifestation of costs of resistance.

Our creation of an artificial indel polymorphism for Rps2 revealed a fitness benefit of up to 40% associated with the presence of Rps2, albeit an equivalent benefit for R and S alleles (Fig. 2). The benefit of carrying an allele of Rps2 appears to result from its function as a negative regulator of the defence response, as loss of Rps2 causes the upregulation of a number of genes involved in induced responses to stress (Fig. 3 and Supplementary Table 13). The critical function of Rps2 provides a clear explanation for why S alleles are not deleted, as they typically are for Rpm1 and Rps5; however, it does not explain the maintenance of both R and S alleles at Rps2. There are several possible explanations for the long-term maintenance of these clades. First, the S alleles may encode the ability to recognize effectors or pathogens that have yet to be identified,
leading to selection for the retention of the functional ‘S’ allele. Alternatively, the previously measured cost of attack\(^2\), in which the cost of response by R alleles of Rps2 is a larger physiological burden than infection, may favour S alleles in certain environments. Such a benefit of susceptibility, even though environmentally restricted, could promote stable Rps2 polymorphism\(^3\). Spatial or temporal variation in other costs of resistance, namely in costs of surveillance or defence, could also promote stable Rps2 polymorphism\(^4,5\). Finally, spatial or temporal variation in costs of virulence in pathogens carrying avrRps2 could similarly promote stable polymorphism at Rps2\(^4,5\).

More generally, functional R gene alleles, when expressed, have the potential to carry a physiological cost because of expression or mis-expression\(^6,7\). For Rps2, we observed a cost associated with increasing levels of expression (Fig. 2f). Selection should act to minimize the costs associated with R loci, especially costs of S alleles that have no pleiotropic, beneficial function. We suggest that a natural consequence of this selective process should be the deletion of S alleles, because deletions can carry no costs associated with their (mis)expression. Indeed, only when S alleles harbour beneficial effects should they be retained by selection, as is the case for Rps2. Our results thus demonstrate that defence loci segregating for functional alternatives, rather than for indel polymorphisms, limit the manifestation of costs of surveillance. Furthermore, our results suggest that genetic architecture both impacts, and is impacted by, physiological costs associated with segregating R gene alleles. Given the substantial variation in R gene evolutionary histories\(^2\) and genetic architectures\(^1,2,3,5\), it will be fascinating to further disentangle the complex interplay of genetic, physiological and ecological factors in the generation of diversity.

**Methods**

**Cre-lox insertion of RPS2.** We introduced five intact alleles of Rps2 into the same genomic location using a Cre-lox system in an rps2-101C mutant of Col-0\(^6\), a plant line with a stop codon in RPS2 at amino acid 235 that is a presumed null mutation\(^6\) (Supplementary Fig. 1). We also introduced an empty integration vector, without Rps2, to obtain empty vector insertions in the isogenic RPS2 null background (hereafter Rps2\(^{-}\)). We repeated this process for each of three genomic locations, creating 18 isolines in all (Supplementary Table 1). Three alleles from the resistant clade, Col-0, Cto-0 and Ler-0 (R clade alleles or Rps2\(^{-}\) lines) were characterized as resistant in their native genetic background\(^6,10\). One allele from the R clade, Wa-0 (Rps2\(^{-}\)), was characterized as partially resistant in its native genetic background\(^6\). One allele from the susceptible clade, Wu-0 (S clade allele or Rps2\(^{+}\) lines) was characterized as susceptible in its native genetic background\(^6\). The Rps2\(^{+}\) and Rps2\(^{-}\) lines exhibited elevated HR and resistance compared with the Rps2\(^{-}\) and Rps2\(^{+}\) lines when infected with *P. syringae* pv. *avrRpt2*\(^2\) (Supplementary Figs 6–7). Further details are included in the Supplementary Methods.

**Field fitness experiment.** Seedlings of each of 17 RPS2 lines were germinated in 98-cell trays containing 50:50 Metromix 200:Farfod C2 in the University of Chicago greenhouses. Plants with the Col-0 allele at insertion site 1 did not germinate. Seedlings were thinned on day 7 of growth and flat locations were randomly cycled daily in the greenhouse to standardize growth conditions. On day 14, 100 seedlings per line were transplanted to a rolled field site in Downers Grove, Illinois, in a randomized block design in which each block contained a plant from each Rps2 line. Plants were set out in 15 rows of nine blocks, spaced by 0.25 m within rows and by 1 m between rows. Plants were irrigated for 1 week to reduce transplantation shock, and then sustained only by natural rainfall. The field was hand weeded once and plants received no other protection from competition or pests. Fifty-five F\(^{0}\) and Rps2\(^{-}\) lines exhibited elevated HR and resistance compared with the Rps2\(^{-}\) and Rps2\(^{+}\) lines when infected with *P. syringae* pv. *avrRpt2*\(^2\) (Supplementary Figs 6–7). Further details are included in the Supplementary Methods.

**Growth chamber fitness experiment.** A total of 1,400 seedlings of seven lines with Rps2 at insertion site 2 were germinated in 36-cell trays containing 25:25:50 Metromix 200:Farfod C2:Turface in the University of Chicago growth chambers. Owing to growth chamber size constraints, only lines from insertion site 2 were included. Seedlings were thinned and accessions were randomized within flats on day 7 of growth. After day 14, plants were watered every other day to mimic stressful growth conditions in the field. After 6 weeks of growth, we stopped watering and allowed the plants to dry for 2 weeks before processing. Two fitness proxies were measured: dry weight and unabsceded seed set.

**Sterile plant fitness experiments.** Two sets of lines were grown in sterile conditions. First, isogenic lines with Rps2 at insertion site 2 were grown to measure fitness of plants with alleles of Rps2 relative to the Rps2 knockout. Second, four lines were measured to determine whether the truncated RPS2 protein present in the rps2-101C impacted the survival of *P. syringae* pv. *avrRpt2* resistance: Col-0, an amiRNA knockdown of RPS2 in a Col-0 background, and two lines on an isogenic rps2-101C Rps2 null background at insertion site 2, one containing an insert of the Col-0 allele and one containing only the empty box cassette. Further information on sterile conditions is included in the Supplementary Methods.

**Fitness analysis.** R was used to specify nested linear models using the lm function from the stats package. Two sets of linear models were used to generate confidence intervals for each of seven fitness proxies at each insertion site for the field data, and for two fitness proxies for the growth chamber data. The field linear models included an effect of the date the plant was collected from the field. The growth chamber linear models included an effect of the date the plant was processed. The first set of models nested allele into either one of three (S, R, pR) or one of two (KO, Rps2\(^{-}\)) allelic classes and considered each genomic insertion site independently. The second set of models combined data for all genomic insertion sites to nest Rps2 expression level into one of three allele classes (S, R, pR).

**Quantitative real-time PCR.** Expression of all Rps2 isogenic lines was measured with qPCR using primers for Rps2 and normalizing between samples using three reference genes: *PP2A*, *Helicase* and *BHLH* (ref. 36). A subset of isogenic strains was used to compare Rps2 expression in natural accessions to expression in the allelic series. Details of qPCR are described in the Supplementary Methods.

**Whole transcriptome profiling.** Plants with the Col-0 allele at insertion site 2 (R), the Col-0 allele at insertion site 3 (High R), the Wu-0 allele at insertion site 2 (S) or an empty vector at insertion site 2 (KO) were grown in sterile growth media as in the sterile plant fitness experiments. The Col-0 allele was chosen as the representative R allele for two reasons: (1) non-coding divergence was smallest between Col-0 and Wu-0 (Supplementary Fig. 1b) and (2) Rps2 gene expression variation existed to allow observation of expression level effects. Specific contrasts included (1) R vs. S; (2) R, (3) High R, (4) S vs. KO; and (5) High R vs. R* RNA-seq and analysis followed standard protocols described in the Supplementary Methods.

Received 17 February 2016; accepted 20 June 2016; published 18 July 2016

**References**

1. Tian, D., Araki, H., Stahl, E., Bergelson, J. & Kreitman, M. Signature of balancing selection in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 11525–11530 (2002).
2. Bakker, E., Toomajian, C., Kreitman, M. & Bergelson, J. A genome-wide survey of R gene polymorphisms in *Arabidopsis*. *Plant Cell Online* **18**, 1803–1818 (2006).
3. Thrall, P. et al. Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecol. Lett.* **15**, 425–435 (2012).
4. Brown, J. & Tellier, A. Plant-parasite coevolution: Bridging the gap between genetics and ecology. *Annu. Rev. Phytopathol.* **49**, 345–367 (2011).
5. Brown, J. K. Durable resistance of crops to disease: a Darwinian perspective. *Annu. Rev. Phytopathol.* **53**, 513–539 (2015).
6. Moreno-Gámez, S., Stephan, W. & Tellier, A. Effect of disease prevalence and spatial heterogeneity on polymorphism maintenance in host–parasite interactions. *Plant Pathol.* **62**, 133–141 (2013).
7. Korves, T. & Bergelson, J. A novel cost of R gene resistance in the presence of disease. *Am. Nat.* **163**, 489–504 (2004).
8. Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M. & Bergelson, J. Fitness costs of *R*-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74–77 (2003).
9. Karasov, T. et al. The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* **512**, 436–440 (2014).
10. Grant, M. et al. Structure of the *Arabidopsis* RPM1 gene enabling dual specificity disease resistance. *Science* **269**, 843–846 (1995).
11. Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. & Bergelson, J. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**, 667–671 (1999).
12. Rose, L., Atwell, S., Grant, M. & Holub, E. Parallel loss-of-function at the RPM1 bacterial resistance locus in *Arabidopsis thaliana*. *Front. Plant Sci.* **3**, 287 (2012).
13. Meyers, B. C., Kozik, A., Griego, A., Kuang, H. & Michelmore, R. W. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**, 809–834 (2003).
14. Kunkel, B., Bent, A., Dahlbeck, D., Innes, R. & Staskawicz, B. RPS2, an Arabidopsis disease resistance locus specifying recognition of Pseudomonas syringae strains expressing the avirulence gene avrRpt2. Plant Cell Online 5, 865–875 (1993).

15. Yu, G. L., Katagiri, F. & Ausubel, F. Arabidopsis mutations at the RPS2 locus result in loss of resistance to Pseudomonas syringae strains expressing the avirulence gene avrRpt2. Mol. Plant Microbe Interact. 6, 434–443 (1993).

16. Mauricio, R. et al. Natural selection for polymorphism in the disease resistance gene Rps2 of Arabidopsis thaliana. Genetics 163, 735–746 (2003).

17. Bergelson, J. & Purrington, C. B. Surveying patterns in the cost of resistance in plants. Am. Nat. 148, 536–558 (1996).

18. Raj, A. & Rifkin, S. Variability in gene expression underlies incomplete penetrance. Nature 463 913–918 (2010).

19. Mindrinos, M., Katagiri, F., Yu, G. L. & Ausubel, F. M. The A. thaliana disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78, 1089–1099 (1994).

20. Tao, Y., Yuan, F., Leister, R. T., Ausubel, F. M. & Katagiri, F. Mutational analysis of the Arabidopsis nucleotide binding site-leucine-rich repeat resistance gene RPS2. Plant Cell 12, 2541–2554 (2000).

21. McNellis, T. W., Mudgett, M. B., Li, K. & Aoyama, T. Glucocorticoid-inducible expression of a bacterial avirulence gene in transgenic Arabidopsis induces hypersensitive cell death. Plant J. 14, 247–257 (1998).

22. Leister, R. T. & Katagiri, F. A resistance gene product of the nucleotide binding site – leucine rich repeats class can form a complex with bacterial avirulence proteins in vivo. Plant J. 22, 345–354 (2000).

23. Banerjee, D., Zhang, X. & Bent, A. The leucine-rich repeat domain can determine effective interaction between RPS2 and other host factors in Arabidopsis RPS2-mediated disease resistance. Genetics 158, 439–450 (2001).

24. Axtell, M. J., McNellis, T. W., Mudgett, M. B., Hsu, C. S. & Staskawicz, B. J. Mutational analysis of the Arabidopsis RPS2 disease resistance gene and the corresponding pseudomonas syringae avrRpt2 avirulence gene. Mol. Plant Microbe Interact. 14, 181–188 (2001).

25. Day, B. et al. Molecular basis for the RIN4 negative regulation of RPS2 disease resistance. Plant Cell Online 17, 1292–1305 (2005).

26. Day, B., Dahlbeck, D. & Staskawicz, B. NDR1 interaction with RIN4 mediates the differential activation of multiple disease resistance pathways in Arabidopsis. Plant Cell Online 18, 2782–2791 (2006).

27. Qi, D., DeYoung, B. & Innes, R. Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein. Plant Physiol. 158, 1819–1832 (2012).

28. Bergelson, J., Kreitman, M., Stahl, E. A. & Tian, D. Evolutionary dynamics of plant R-genes. Science 292, 2281–2285 (2001).

29. Gan, X. et al. Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature 477, 419–423 (2011).

30. Laine, A. L. & Tellier, A. Heterogeneous selection promotes maintenance of polymorphism in host–parasite interactions. Oikos 117, 1281–1288 (2008).

31. Guo, Y. L. L. et al. Genome-wide comparison of nucleotide-binding-site-leucine-rich repeat-encoding genes in Arabidopsis. Plant Physiol. 157, 757–769 (2011).

32. Vergunst, A. C., Jansen, L. E. & Hooykaas, P. J. Site-specific integration of Agrobacterium T-DNA in Arabidopsis thaliana mediated by Cre recombinase. Nucleic Acids Res. 26, 2729–2734 (1998).

33. Caicedo, A. L., Schaal, B. A. & Kunkel, B. N. Diversity and molecular evolution of the RPS2 resistance gene in Arabidopsis thaliana. Proc. Natl Acad. Sci. USA 96, 302–306 (1999).

34. Gutman, D. S. & Greenberg, J. T. Functional analysis of the type III effectors AvrRpt2 and AvrRpm1 of Pseudomonas syringae with the use of a single-copy genomic integration system. Mol. Plant Microbe Interact. 14, 145–155 (2001).

35. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W. R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol. 139, 5–17 (2005).

36. Cao, J. et al. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nature Genet. 43, 956–963 (2011).

Acknowledgements
We thank H. King, L. Merwin, C. Meyer, W. Muliyati, A. Olesen, N. Shahoor and T. Stewart for their assistance in the field; J. Greenberg for donation of strains for infection; P. Hooykaas for pSDM3110, M. Vetter for her high-throughput infection protocol; and B. Brachi, T. Karasov, M. Kreitman and L. Merwin for helpful discussions. This research was supported by NSF and NIH grants to J.B. and a grant from the National Natural Science Foundation of China to X.Q.S. (Grant no. 31470448).

Author contributions
X.Q.S. created the allelic series; A.M., X.Q.S. and J.B. designed the experiments; A.M., X.Q.S. and J.B. wrote the paper.

Additional information
Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.B.

Competing interests
The authors declare no competing financial interests.