

# Genetic architecture of *Arabidopsis thaliana* response to infection by *Pseudomonas syringae*

PX Kover<sup>1</sup>, JB Wolf<sup>1</sup>, BN Kunkel<sup>2</sup> and JM Cheverud<sup>3</sup>

<sup>1</sup>School of Biological Sciences, University of Manchester, Stopford Building 3.614, Oxford Road, Manchester M13 9PT, UK; <sup>2</sup>Department of Biology, Washington University, Campus Box 1137, St Louis, MO 63130-4899, USA; <sup>3</sup>Department of Anatomy & Neurobiology, Washington University School of Medicine, 660 S Euclid Ave., St Louis, MO 63110, USA

Plant pathogens can severely reduce host yield and fitness. Thus, investigating the genetic basis of plant response to pathogens is important to further understand plant–pathogen coevolution and to improve crop production. The interaction between *Arabidopsis thaliana* and *Pseudomonas syringae* is an important model for studying the genetic basis of plant–pathogen interactions. Studies in this model have led to the discovery of many genes that differentiate a resistant from a susceptible plant. However, little is known about the genetic basis of quantitative variation in response to *P. syringae*. In this study, we investigate the genetic basis of three aspects of *A. thaliana*'s response to *P. syringae*: symptom severity, bacterial population size and fruit production using a

quantitative trait loci (QTL) analysis. We found two QTL for symptom severity and two for fruit production (possible candidate genes for observed QTL are discussed). We also found significant two-locus epistatic effect on symptom severity and fruit production. Although bacterial population size and symptom severity were strongly phenotypically correlated, we did not detect any QTL for bacterial population size. Despite the detected genetic variation observed for susceptibility, we found only a weak overall relationship between susceptibility traits and fitness, suggesting that these traits may not respond to selection.

*Heredity* (2005) **94**, 507–517. doi:10.1038/sj.hdy.6800651  
Published online 16 March 2005

**Keywords:** plant; pathogen; symptom severity; bacteria; fitness; QTL

## Introduction

Pathogens are ubiquitous and often severely reduce host fitness. As a result, they can have important effects on plant ecology, such as altering community level diversity, or affecting the evolution of host morphology and life-history traits (Clay and Van der Putten, 1999). By affecting fitness-related traits, pathogens also often have important economical effects on crops by reducing growth and yield (Wolfe, 2000). The coevolutionary dynamics between host and pathogen depends on the genetic architecture of resistance and virulence, as well as the selective value of different resistance-related traits (Kover and Caicedo 2001). Thus, understanding the underlying genetic architecture of plant–pathogen interactions can provide important insights into ecological, evolutionary and agricultural systems.

Much progress in understanding the genetic basis of disease resistance in plants comes from studies on the model system *Arabidopsis thaliana* (Kunkel *et al.*, 1993; Glazebrook, 1999). Studies of the genetic basis of simple forms of disease resistance in *A. thaliana* have identified single Mendelian genes of large effect, known as resistance (*R*) genes (Glazebrook, 1999; Dangl and Jones, 2001). These genes govern a plant's ability to recognize pathogens expressing specific elicitors (the product of *Avr*

genes), resulting in the activation of plant defense responses that serve to limit pathogen growth and disease development (Whalen *et al.*, 1991; Glazebrook *et al.*, 1997).

Disease resistance mediated by *R* genes (also called 'gene-for-gene resistance') behaves as a qualitative trait, where the presence or absence of an *R* gene differentiates between a resistant and a susceptible plant. Thus, disease resistance in *A. thaliana* is usually considered a qualitative trait. However, closer examination of *A. thaliana*'s response to virulent strains shows a continuous variation in disease susceptibility (Whalen *et al.*, 1991; Kover and Schaal, 2002). For example, the accession Nossen (No-0) shows fewer symptoms and less pathogen growth than the accession Columbia (Col-0) when challenged with a virulent strain of the bacterial pathogen *Pseudomonas syringae* (pv. *tomato* strain PstDC3000) for which no *R* gene has been identified (Chen *et al.*, 2000). Although the genetic basis of *R* gene-mediated resistance has been extensively studied at the genetic and molecular level, the genetic basis of the quantitative variation in disease susceptibility is only beginning to be investigated (Glazebrook *et al.*, 1996; Kloek *et al.*, 2001). In addition to furthering our understanding of the defense mechanisms of *A. thaliana*, studying the genetic basis of quantitative variation in susceptibility to pathogens can provide important insights into the evolutionary interaction between plants and pathogens in natural populations, where variation in susceptibility is often quantitative (Lande and Arnold, 1983; Alexander, 1992).

The genetic basis of quantitative variation is typically polygenic. Thus, analysis of the genetic basis for variation in the response of plants to pathogens can be

Correspondence: PX Kover, School of Biological Sciences, University of Manchester, Stopford Building 3.614, Oxford Road, Manchester M13 9PT, UK. E-mail: kover@man.ac.uk

Received 3 September 2004; accepted 12 January 2005; published online 16 March 2005

addressed using quantitative trait loci (QTL) analysis, which allows for the simultaneous identification of multiple genomic regions responsible for quantitative variation in trait expression (Lander and Botstein, 1989). The interaction between *A. thaliana* and the bacterial pathogen *P. syringae* is an ideal system to investigate the genetic basis of quantitative variation in susceptibility using QTL analysis. Natural populations of *A. thaliana* in the US have been reported to be infected with *P. syringae* and *P. viridiflava* (Jakob *et al*, 2002), and quantitative variation in disease susceptibility has been previously observed in this system (Kover and Schaal, 2002). In addition, the genome of *A. thaliana* has been completely sequenced, making a large number of molecular markers available (a continuously updated list is available online at The Arabidopsis Information Resource (TAIR), [www.arabidopsis.org](http://www.arabidopsis.org)). Furthermore, this system also presents the opportunity to measure different components of plant response to pathogens, allowing the simultaneous investigation of different components of their interaction for a more complete analysis of susceptibility.

Here, we use QTL analysis to investigate the genetic basis of quantitative variation in disease susceptibility in *A. thaliana*. The experimental population chosen for the analysis is the F<sub>2</sub> generation from the intercross between No-0 and Col-0 lines. These parental lines show a significant difference in susceptibility to *P. syringae* (Chen *et al*, 2000; Kover and Schaal, 2002). We searched for QTL affecting three different susceptibility-related traits: (1) the extent of chlorosis and necrotic lesions and patches in infected leaves, which allows for the identification of genetic factors that reduce the severity of disease symptoms; (2) the number of bacterial cells per cm<sup>2</sup> of host leaf tissue, which allows for the identification of genetic factors that may reduce pathogen entrance and/or bacterial growth and (3) the number of fruits produced by each plant, which measures variation in plant fitness (yield) under infection. Since infection by *P. syringae* reduces *A. thaliana* fitness (Kover and Schaal, 2002), it is expected that susceptibility would be negatively associated with fitness, and therefore, be under selection. However, the costs of being resistant (Bergelson and Purrington, 1996; Tian *et al*, 2003), as well as those of tolerance traits (Stowe *et al*, 2000; Kover and Schaal, 2002) can breakdown the negative relationship between susceptibility and fitness. Thus, we also investigated the relationship between different components of susceptibility and fitness using a selection analysis (Lande and Arnold 1983). Specifically, we address the following questions: (1) can we identify genetic factors (QTL) that underlie the differences in susceptibility between the two *A. thaliana* lines; (2) what is the genetic architecture of these factors? Does dominance and epistasis affect the expression of susceptibility?; (3) are there likely candidate genes for the identified QTL and (4) what is the relationship between susceptibility traits and fitness?

## Material and methods

The F<sub>2</sub> progeny used in this study were produced by initially crossing emasculated Col-0 *rps2-201C* flowers with pollen from No-0 *rps2-201C* flowers. F<sub>1</sub> plants from this cross, confirmed to be heterozygous using genetic

markers, were self-pollinated to produce F<sub>2</sub> seeds. All F<sub>2</sub> seeds used in the QTL mapping were collected from a single F<sub>1</sub> individual. We chose to use a F<sub>2</sub> design (instead of the commonly used sets of recombinant inbred lines) because it allows for the investigation of the role of dominance. Both parental lines used in the analysis carry mutations at the *RPS2* disease resistance locus for the purpose of a different experiment. The isolation of the Col-0 *rps2-201C* mutant is described in Kunkel *et al* (1993), and the generation of the No-0 *rps2-201C* line is described in Chen *et al* (2000). *RPS2* is an *R* gene capable of recognizing pathogens carrying the avirulence gene *AvrRPS2*. Plants in our experiment were challenged with the bacteria *P. syringae* pv. *tomato* strain DC3000 (*Pst*DC3000), which does not carry *AvrRPS2*. Thus, it is unlikely that the presence of the *rps2* mutation in these lines is important in the context of this study. For simplicity, the parental lines will be from now on referred only as 'Col-0' and 'No-0'.

Seeds from the F<sub>2</sub> sibship were planted in six trays containing 50 individual cells, for a total of 276 plants. For comparison, we also planted the original parental stocks (No-0 *rps2-201C* and Col-0 *rps2-201C*) in two cells of each tray. Seeds were cold-treated for 3 days at -4°C, and later germinated and grown under constant environment in a growth chamber kept at 25°C and 12 h of light/day. Tray position in the growth chamber was randomly assigned and rotated every week to minimize the effect of microenvironmental variation.

## Pathogen inoculation and disease assays

Plants, 4-week old, were inoculated by inverting the trays and submerging entire plant rosettes in a bacterial solution containing 10 mM MgCl<sub>2</sub>, 0.02% L-77 Silwet and 10<sup>8</sup> cells/ml of *Pst*DC3000 as described in Whalen *et al* (1991). Trays were covered with a plastic dome for the first 24 h to maintain high humidity. At 3 days after inoculation, four disks of leaf tissue were collected from each plant with a cork borer (0.25 cm<sup>2</sup> area) to determine the number of bacterial cells present per cm<sup>2</sup> of leaf tissue. Disks were collected randomly from different leaves before any symptoms were visible. The disks were ground in 10 mM MgCl<sub>2</sub> solution and plated after appropriate dilutions were made on NYG agar plates (containing 1 mg/ml of Rifampicin). The number of colony-forming units (CFUs) per plate was counted 24 h later. To determine the severity of disease symptoms, each plant was visually inspected for disease symptoms 5 days after inoculation. Symptoms were scored on a predefined scale that takes into consideration the presence and extent of chlorosis and necrotic disease lesions. The scale ranges from 1 (no signs of disease symptoms) to 5 (extensive chlorosis and patches of necrotic lesions). To insure scoring accuracy, two different people independently scored each plant. Since the two scores were tightly correlated ( $r^2=0.87$ ), the mean between the two scores were used in the analysis.

Since *A. thaliana* is an annual plant, the effect of pathogen infection on lifetime fitness can be estimated by total seed production upon senescence. Thus, after plants were scored for disease symptoms, day length was gradually increased to 16 h to induce flowering. Plants were kept in the growth chamber until senescence, when total fitness was estimated by counting total fruit

production. The number of fruits produced by a plant has been previously shown to correlate tightly with total seed production (Mauricio and Rausher, 1997). To determine the role of life-history traits on fitness, we measured rosette size one day before inoculation, and recorded flowering time. Plants were checked every day for bolting. Flowering time corresponds to the first day a flowering bud was observed.

### Phenotypic analysis

For each of the five traits measured, we calculated the average and standard deviation within the parental lines and F<sub>2</sub> progeny. To test whether parental lines significantly differ for these traits, we used univariate *t*-tests. Transgressive segregation was determined by evaluating whether any F<sub>2</sub> progeny had trait values that were more extreme than the most extreme parental phenotype plus (or minus) 2 standard deviations. To test for correlation between traits in the F<sub>2</sub> progeny, we calculated pairwise Pearson's correlations between all traits measured.

To investigate the potential selective value of reduced susceptibility, we performed a selection analysis following the regression method proposed by Lande and Arnold (1983). Since this experiment was performed in a growth chamber, the observed pattern of selection will not necessarily reflect patterns of natural selection. However, it can still provide valuable information about the relationship between susceptibility and fitness. To evaluate the total effect of each trait on fitness, we calculated the univariate regression of the standardized symptom scores, size of bacterial population, rosette size and flowering time on relative fitness. This is equivalent to calculating the standardized selection differentials (ie, the total covariance of a trait with fitness). To obtain the relative fitness of each individual, we divided the number of fruits produced by each F<sub>2</sub> individual by the average number of fruits produced in the F<sub>2</sub> progeny. To evaluate the direct effect of the same traits, independent of correlations between traits, we calculated the partial

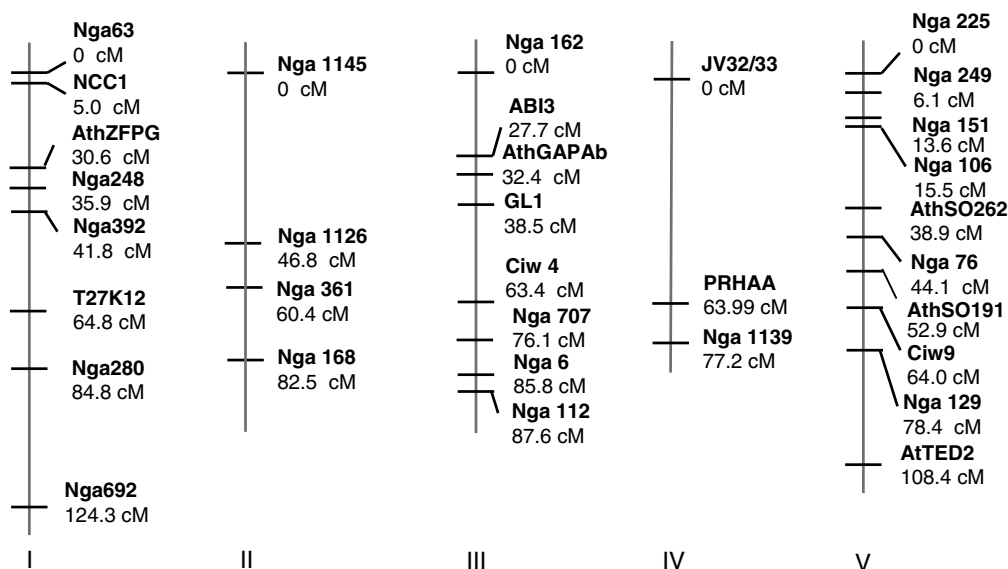
regression coefficients of standardized traits on relative fitness using a multiple regression (the standardized selection gradients).

### QTL analysis

One large leaf from each F<sub>2</sub> plant was collected shortly after the onset of flowering, and DNA was extracted using the Dneasy™ kit from Quiagen. DNA from the two parental lines was screened with a 100 different markers to identify markers that were polymorphic between parents. The DNA from each F<sub>2</sub> individual was genotyped with 28 microsatellite and five CAPS (cleaved amplified polymorphic sequences) markers (see Figure 1). All primers used to genotype these markers are described at TAIR ([www.arabidopsis.org](http://www.arabidopsis.org)). Conditions for PCR amplifications generally followed the Bell and Ecker (1994) protocol, but annealing temperature was individually optimized for each primer pair (specific annealing temperatures used are available upon request).

The genetic map for the F<sub>2</sub> cross was generated using MAPMAKER 3.0b (Lander *et al*, 1987). Since the *A. thaliana* genome has been completely sequenced, the order of markers and their chromosome assignment (according to the information present in [www.arabidopsis.org](http://www.arabidopsis.org)) was imposed in the analysis. Recombination distances between markers were estimated using the Kosambi function.

QTL analysis was performed using QTL Cartographer software version 2.0 (Basten *et al*, 1994). We initially performed an interval mapping with tests being performed every 2 cM. Statistical significance for the LOD scores obtained in the QTL analysis was determined individually for each trait by performing 10 000 permutations using the method developed by Churchill and Doerge (1994). Composite interval mapping (CIM) was used to further analyze traits for which a putative QTL was identified using interval mapping. CIM can improve QTL results by partitioning out the contribution of QTL of large effect and of linkage disequilibrium. However,



**Figure 1** Recombination map for the cross between No-0 and Col-0 lines showing markers used in QTL analysis. Recombination distances were determined using Mapmaker.

the results from composite interval mapping have to be interpreted carefully since the outcome of the analysis depends on the covariate markers chosen to do the analysis (Broman, 2001). To insure that our results were not dependent on one particular model, we ran CIM models with different number of covariate markers (1–5) and window size (5–10 cM) for each trait. Potential marker covariates were identified using a stepwise forward and backward regression, which ranks the five markers that explain the highest proportion of the variation. Confidence intervals for QTL position were established as the region around the peak location where there is one order of magnitude (1 LOD) change in the probability of a false-positive (Lander and Botstein, 1989).

We searched for pairwise epistatic interactions between all genotyped markers using canonical correlations analysis in SAS as described in Cheverud and Routman (1995) and Routman and Cheverud (1997). While the QTL cartographer software does allow for tests of epistatic effects, it only performs tests on locations previously determined to have a main effect. Since genetic factors without a main effect can have significant epistatic effects, we chose to perform the complete analysis for all pairs of scored markers located on different chromosomes using SAS. Pairs of loci that showed significant interaction ( $P < 0.05$ ) were further analyzed with a full regression model where phenotypic values were regressed on additive, dominant, additive by additive ( $A \times A$ ), additive by dominant ( $A \times D$ ), dominant by additive ( $D \times A$ ) and dominant by dominant ( $D \times D$ ) genotypic effects. Estimates of all effects and their significance were calculated using the method described in Cheverud and Routman (1995) and Cheverud *et al* (1997). The significance threshold used to evaluate the significance of epistatic effects was determined by performing a Bonferroni correction for the number of independent comparisons given the covariation between markers (as described in Vaughn *et al*, 1999 and Cheverud, 2001).

Total phenotypic variance explained by QTL and epistatic interaction detected in our study was determined by fitting the markers closest to each significant QTL in a regression model. Calculations were performed using PROC REG in SAS.

Two approaches were used to screen candidate genes for QTL identified in this analysis. We scanned the annotated sequence of *A. thaliana* (available at TAIR, Garcia-Hernandez *et al*, 2002) within the confidence interval of each QTL for genes with known or predicted roles in interactions with pathogens. The equivalence between the available physical map for *Arabidopsis* and

our recombination map was made using the genetic markers with known positions in the physical map. We also compared the position of genes known to be important in modulating pathogen growth or symptom development with the positions of observed QTL. Genes considered in the latter approach are ones known to participate in the salicylic acid (SA), ethylene, jasmonic acid synthesis and signalling pathways (Kunkel and Brooks, 2002). We also considered genes involved in cell death and overall disease susceptibility (Greenberg and Ausubel, 1993; Glazebrook *et al*, 2003).

## Results

### Trait distribution and correlation

All five traits measured differed significantly between the parental types (Table 1). Col-0 has a significantly smaller rosette, develops more severe disease symptoms when infected with *Pst*DC3000, harbors a larger bacterial population in its leaves, flowers later and produces fewer fruits than the No-0 parent (Table 1). All five traits vary continuously among the  $F_2$  progeny (Figure 2). Transgressive segregation (ie progeny with phenotypes more extreme than either parent) was observed for rosette size, flowering time and fruit production (Figure 2). Transgressive segregation can be due to parental lines having a mixture of alleles that increase and decrease the traits, or the presence of epistasis. It may also be caused by genotype by environment interaction variation. However, this is less likely to explain our results considering that the transgressive  $F_2$  phenotypes are more extreme than the most extreme parental phenotype plus (or minus) 2 standard deviations.

There were significant correlations among some of the traits in the  $F_2$  (see Table 2). Disease symptom severity and bacterial growth were positively correlated with each other ( $P = 0.0001$ ), but neither was significantly correlated with fruit production. Among life-history traits, we observed a significant negative correlation between rosette size and number of days to flower ( $P = 0.00001$ ). Both of these traits were correlated with fruit production. However, rosette size was positively correlated ( $P = 0.0001$ ), while flowering time was negatively correlated ( $P = 0.0001$ ) with fruit production.

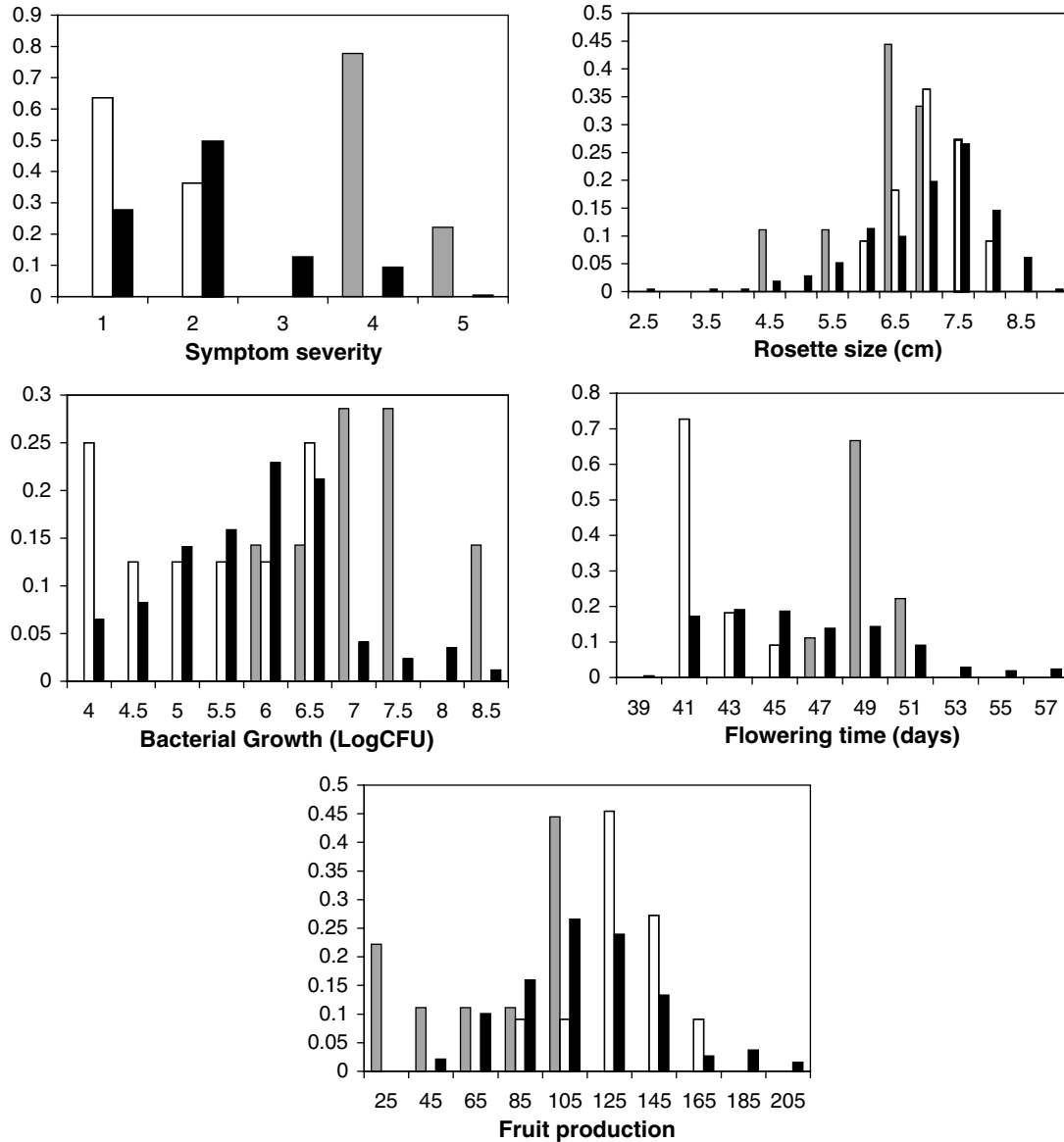
### QTL associated with symptom severity

The recombination map obtained for this cross is shown in Figure 1 ( $F_2$  genotypes are available upon request). This map spans 480 cM and encompasses 97 of the 125 MB genome of *A. thaliana* (by comparison to the physical map available at TAIR). The recombination

**Table 1** Mean and standard deviation of measured traits in the two parental lines, Col-0 and No-0, and the significance of a *t*-test comparison of parental means

Trait	$F_2$	Col-0	No-0	P
Rosette size (cm)	6.8 ± 1.02 (212)	6.09 ± 0.84 (9)	6.81 ± 0.54 (11)	0.046
Symptom score	2.1 ± 0.87 (170)	4.18 ± 0.36 (9)	1.53 ± 0.38 (11)	$2 \times 10^{-12}$
Bacterial pop. size (log (CFU/cm <sup>2</sup> ))	5.7 ± 0.95 (212)	6.91 ± 0.72 (7)	5.09 ± 0.98 (8)	0.002
Days to flowering	45.61 ± 3.93 (209)	48.89 ± 1.17 (9)	41.55 ± 1.04 (11)	$4 \times 10^{-11}$
#Fruits	103.12 ± 32.11 (188)	70.00 ± 32.57 (9)	119.36 ± 21.86 (11)	0.002

Numbers in parenthesis indicate the number of individuals measured.



**Figure 2** Frequency distribution of traits measured (for details, see text). White bars indicate distribution for No-0 parents; hatched pattern indicates distribution for Col-0 parents and black bars indicate the distribution on F2 progeny.

**Table 2** Pearson's correlation between all traits measured on the F2 progeny

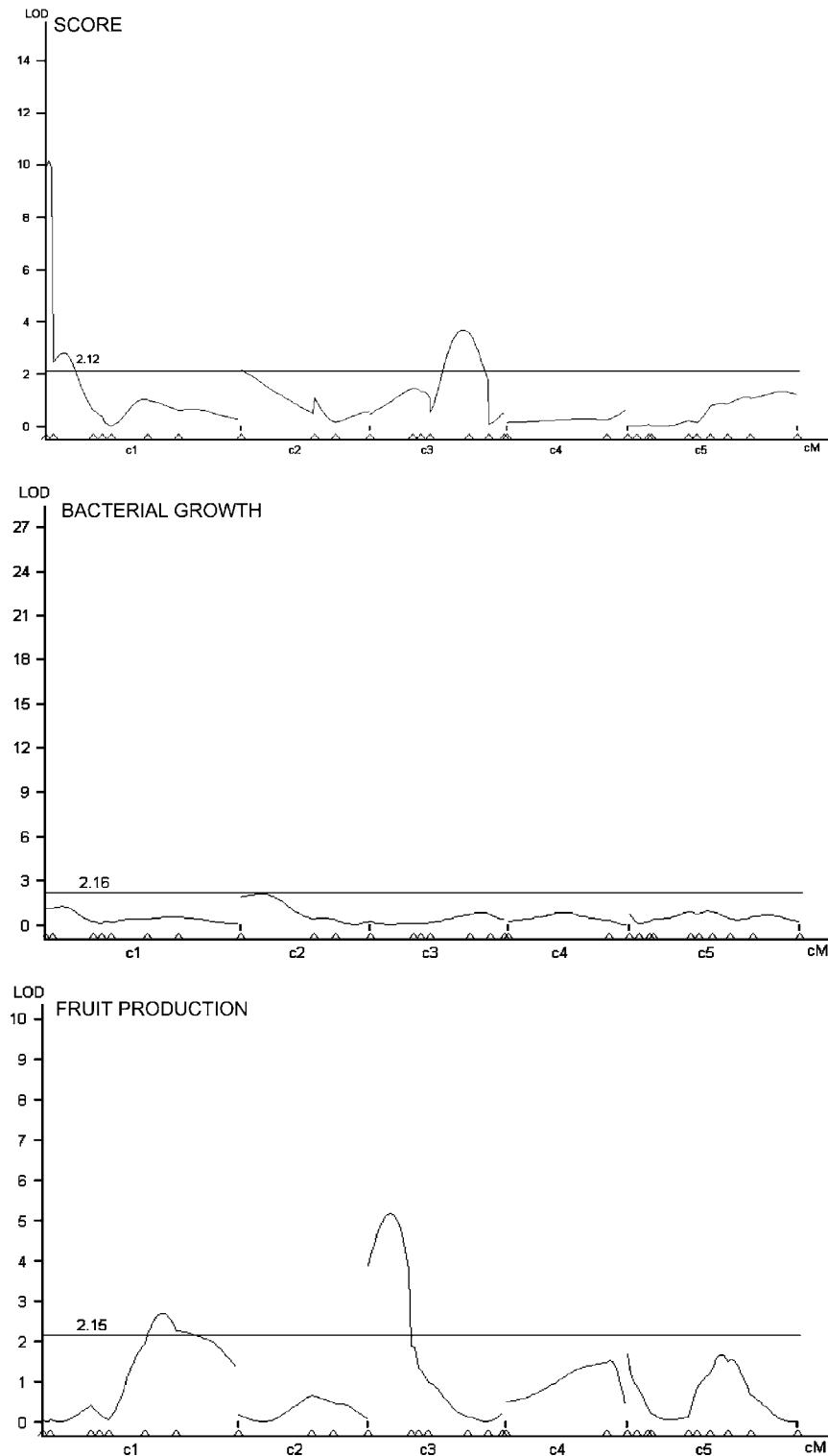
Trait	Rosette size	Symptom score	LogCFU	Days to flowering
Symptom score	-0.114			
Bacterial pop. size	0.019	<b>0.411</b>		
Days to flowering	<b>-0.501</b>	-0.056	-0.044	
Fruit production	<b>0.350</b>	0.030	-0.115	<b>-0.357</b>

Bold values indicate correlations that were significant ( $P < 0.05$ ) after a bonferroni correction for multiple comparisons.

distances observed between markers are comparable to the distances reported between the same markers in other mapping studies (eg Lister and Dean, 1993).

To determine the number and position of QTL that affect symptom severity, we ran three different CIM models. The same three locations were significant at the empirically determined threshold value corresponding

to  $P = 0.05$  ( $LOD = 2.12$ ) whether we used one, two or five markers as covariates. The main difference was that the window size used in the model had a slight affect on the position, and significance of the two peaks identified at the top of chromosome 1. In Figure 3, we show the LOD distribution for the CIM model with five covariate factors and a window size of 4.9 cm. We present the



**Figure 3** LOD distribution for the four traits studied. The five chromosomes are presented in order (c1–c5). Marker positions in each chromosome are indicated by arrowheads in the X-axis. Horizontal line indicates the significance LOD threshold value calculated using permutations (see text).

model with a small window because it separates the effects of the markers Nga63 and NCC1, which are 5 cM apart. Although two independent peaks can still be identified with a window of 4.9 cM, the multiple regression analysis suggests that most of the variation

associated with NCC1 is due to a QTL nearer to Nga63. Thus, we will consider that there is a single QTL in the top of chromosome I, located at 2 cM.

Table 3 shows the position, effect and percentage of phenotypic variation ( $R^2$ ) explained by each of the two

**Table 3** QTL positions identified by composite interval mapping (for more details, see text)

Trait	Chrom.	Position (cM)	1 LOD confidence interval (cM)	LOD score	Additive value (a)	Dominance value (d)	R <sup>2</sup>	Closest marker
Symptom severity	I	2.0	0–4	10.18	–0.458	–0.372	0.174	Nga63
	III	58.5	50–69.4	3.68	–0.396	–0.098 (NS)	0.045	Ciw4
Fruit production	I	74.8	64.8–98.8	3.53	–11.508	15.691 (NS)	0.039	Nga280
	III	16.0	4.0–26.0	5.08	12.706	–18.884	0.097	ABI3

identified QTL. Negative additive effects indicate that the allele from the No-0 parent reduces the trait value. Thus, in both QTL for symptom severity, the allele from the less susceptible parent (No-0) reduces the symptom score. The negative dominance value of the putative QTL close to Nga63 indicates that the No-0 allele is partially dominant over the Col-0 allele.

Of the 418 possible pairwise interactions tested, 13 were significant at  $P < 0.05$ . Only one was significant at  $P < 0.0002$  (the appropriate threshold with the Bonferroni correction for multiple epistasis tests): the interaction between markers Nga392 located on chromosome I and Nga76 located on chromosome V ( $P = 0.00002$ ). This interaction is mostly due to  $D \times D$  effects ( $P = 0.00002$ ). A full model with the two QTL with a main effect and the significant epistatic interaction explains 23.2% of the phenotypic variance in score symptoms.

No genes with known or predicted roles in pathogen response occur within the confidence interval for the QTL identified on chromosome I. However, this interval contains a putative protein with the typical characteristics of a disease-resistance protein of the CC-NBS-LRR class (AT1G10920). Other proteins of this class include disease-resistance genes such as *RPS2*, *RPS5* and *RPM1*, which are involved in the recognition of pathogens (Dangl and Jones (2001), but see Meyers *et al*, 2003). The QTL on chromosome III coincides with the position of two genes of known function in plant/pathogen interactions: *ACD1* (accelerated cell death 1) and *EDS1* (enhanced disease susceptibility 1). Within the confidence interval for the QTL on chromosome III, there are also three other putative proteins with characteristics of disease-resistance proteins (AT350950, AT351560 and AT351570).

#### QTL associated with bacterial population size

No significant association (LOD threshold = 2.16) between bacterial population size and marker genotypes was observed with single marker regression or IM (Figure 3). However, at the top of the chromosome II (21.6 cM), there is a peak that is nearly significant (LOD = 2.12). This possible location for a QTL affecting bacterial growth does not match any of the putative locations for loci affecting disease symptom severity.

Although 35 pairwise interactions were significant at  $P < 0.05$ , none of them were significant at  $P < 0.00021$ . By chance alone, we would expect to find 21 significant interactions among the 418 tested. Thus, it is possible that some of the 35 epistatic interactions are really affecting bacterial growth, but we cannot determine which of these are true interactions and which are false positives simply due to the large number of tests performed.

#### QTL associated with fruit production (fitness)

The results for fruit production are qualitatively the same whether we use IM or CIM. Figure 3 shows the LOD distribution for the CIM model with five covariates and 10 cM window size, which gives a slightly sharper peak for the QTL on chromosome 3. The two QTL significant at the empirically determined threshold value (LOD = 2.15) are presented in Table 3. Positive additive effects indicate that the allele from No-0 increases fruit production. The opposite signs for the additive effects at the two putative QTL indicates that No-0 alleles increase fruit production in one QTL, but reduces in the other. This mixture of allelic effects within parental line may explain the observed transgressive segregation for fruit production. The negative value of the dominance effect indicates that the Col-0 allele is dominant over the No-0 allele in the QTL closest to ABI3. The location of these markers does not coincide with any of the locations identified for symptoms or bacterial growth.

Of the 418 pairwise interactions tested, 13 were significant at  $P < 0.05$  and two were significant at  $P < 0.0002$ . Both, AthGAPAb and ABI3 on chromosome III had significant interactions with AthSO262 on chromosome V, with the interaction being mostly due to  $D \times D$  effects. Since AthGAPAb and ABI3 are 4.7 cM apart, it is likely that both interactions are due to a common QTL located between the two markers. The interacting QTL is likely to be closer to AthGAPAb because the interaction between AthGAPAb has a higher  $F$  ( $F = 6.42$ ,  $P = 0.00008$ ). A full regression model with the interaction between AthGAPAb and AthSO262 plus the two QTL with main effects explains 17.2% of the total phenotypic variance in fruit production.

#### Selection analysis

The multiple regression analysis indicates that all four measured traits: rosette size, symptom severity, bacterial growth and flowering time have significant effects (as measured by partial regression coefficients) on relative fitness (Table 4). This means that each individual trait has a direct effect on relative fitness when all other traits are kept constant. This analysis shows that, under our experimental conditions, infected plants are under selection for larger rosette size, earlier flowering time, smaller bacterial population sizes and higher symptom scores. The  $R^2$  for this model was 0.238, indicating that there is still a large portion of the variance in fitness unexplained. However, this result is not unexpected since so many aspects of plant growth and development can affect lifetime fruit production.

The univariate regression coefficients indicate that only rosette size and flowering time has an overall correlation with relative fitness (Table 4). This means that

**Table 4** Selection analysis for susceptibility and life-history traits

Trait	Univariate regression coefficient	P	Partial regression coefficient	SE	P
Rosette size	<b>0.118</b>	0.001	<b>0.109</b>	0.025	0.001
Days to flowering	<b>-0.107</b>	0.001	<b>-0.072</b>	0.024	0.003
Symptom score	0.016	0.530	<b>0.070</b>	0.026	0.007
LogCFU	-0.036	0.158	<b>-0.060</b>	0.025	0.018

Univariate regression coefficients are equivalent to selection differentials and indicate the net selection on a trait, while the partial regression coefficients are the selection gradients indicating the direct effect of selection on each trait, when all other phenotypes held constant. Data used in this analysis include 153 F<sub>2</sub> individuals for which we had a complete data set. Bold values indicate significant regression coefficients.

only these two traits would be expected to evolve in this population, since the other traits show no net correlation with fitness. The discrepancy between the multivariate and univariate analyses results from the fact that bacterial population size and symptom score have partial selection gradients of similar magnitude and opposite signs while being positively correlated at the phenotypic level. Thus, the independent effect of selection on these two traits cancels out when considering the net force of selection acting on either trait in the univariate model. This may imply some sort of fitness trade-off between these traits.

## Discussion

*A. thaliana* has been an important model for the identification of the genetic basis of disease resistance in plants. As a consequence, many genes that affect plant response to pathogens have been identified in *A. thaliana* (Glazebrook *et al*, 1996; Glazebrook, 1999). However, most of these studies have concentrated on genes and pathways that affect the qualitative response of *A. thaliana* to pathogens; that is, genes that determines whether a plant is resistant or susceptible. In contrast, we know very little about genes that affect quantitative variation in disease susceptibility. Quantitative variation in disease susceptibility is a common form of variation in response to pathogens, where most plants will be susceptible to infection but to different degrees (Geiger and Heun, 1989; Wyss and Muller-Scharer, 1999; Kover and Caicedo, 2001). Although QTL studies are now commonly used in *A. thaliana* to investigate the genetic basis of the natural variation of quantitative traits such as bolting time, tolerance to salinity and floral morphology (Alonso-Blanco *et al*, 1998; Ungerer *et al*, 2002; Quesada *et al*, 2002), this is the first QTL study on *A. thaliana* responses to infection by *P. syringae*. Only two studies have investigated the genetic architecture of disease resistance in *A. thaliana* (Schiff *et al*, 2001; Wilson *et al*, 2001), both of which investigated QTL for resistance to powdery mildew.

Our study revealed two QTL for symptom severity and two QTL for fruit production segregating between accessions No-0 and Col-0 under conditions of *P. syringae* infection. We also found that epistatic interactions significantly affect the expression of these traits. This result is similar to the results of other QTL studies in *A. thaliana* where a moderate number of QTL per trait are seen and epistasis is common (Alonso-Blanco *et al*, 1998; Mitchell-Olds and Pedersen, 1998; Juenger *et al*, 2000). Despite the fact that bacterial growth is highly variable among ecotypes (Kover and Schaal, 2002) and is

significantly different among the parental lines used in this cross (which show about a 2 standard deviation difference in mean bacterial population size), we did not find any significant QTL affecting bacterial growth. This could be due to the fact that we did not have polymorphic markers in regions where QTL for bacterial growth are located (eg middle of chromosome IV), or because there are no QTL of large enough effect to be detected. It is also possible that there is a large number of QTL of small effect, which would require a denser map and a larger sample size to detect. It is important to remember that the statistical power to detect QTL is a function of sample size, number of markers and the magnitude of the QTL effects (Beavis, 1994). Thus, considering the size of our F<sub>2</sub> progeny and the limited number of markers on chromosomes II and IV, the observed number of QTL in this study is probably an underestimate of the total number of QTL that affect susceptibility in this cross.

Several genes that affect plant response to pathogens, such as *R* genes that mediate specific recognition of *P. syringae* strains (Glazebrook *et al*, 1997), and genes in the SA, ethylene and jasmonic acid pathways (Glazebrook, 1999; Glazebrook *et al*, 2003) were considered as candidate genes for the QTL observed in our study. Some of the QTL for symptom severity map close to candidate genes in these categories. *EDS1*, for example, is a particularly interesting candidate gene for the QTL for symptom severity identified on chromosome III because plants carrying a mutant *EDS1* gene develop more severe symptoms under infection by *P. syringae* than wild-type plants (Glazebrook *et al*, 1996). However, any positional candidate gene at this point should be regarded only as hypothetical. The average 1 LOD confidence interval in this study was 19.84 cM, which is roughly equivalent to 4600 kb (assuming 233 kb per cM) and can encompass hundreds of genes (approximately 1050 genes, considering 4.4 kb/gene). Thus, the possible role of the identified candidate genes in quantitative disease resistance should be further tested through association studies (Long *et al*, 1998), complementation approaches (Mackay and Fry, 1996) or using near isogenic lines (Doebley *et al*, 1995).

Although genes that specifically determine fruit production have not been identified, some genes that affect root architecture (*AXR4*), circadian rhythm (*CCA1*) and photo perception (*PHYB*) have been shown to also affect plant fitness (Ballare and Scopel, 1997; Gilliland *et al*, 2002; Green *et al*, 2002) and can be considered potential candidate genes. In reality though, there are too many genes and pathways that can conceivably affect

traits as complex as fruit production. Thus we did not identify *a priori* candidate genes specific for fruit production, and did not attempt to perform an exhaustive search for candidate genes for the fruit production QTL. Since infection with *Pst*DC3000 affects fruit production (Kover and Schaal, 2002), it is possible that QTL identified in this study for fruit production are quite different from QTL that determine total number of fruits produced by an uninfected plant. It is interesting though that the QTL on chromosome V for total number of fruits is in a similar location as a QTL found in a set of uninfected recombinant inbred lines between accessions Cvi and Ler (Ungerer *et al*, 2002). None of the QTL locations for fruit production identified in this study were close to a gene of known function in disease resistance.

Disease susceptibility in plants is empirically evaluated in many different ways. In natural populations susceptibility is often estimated by symptom severity, while the susceptibility of crop varieties is usually evaluated by their relative yield. In many experimental systems, where it is possible to measure pathogen titer, susceptibility is measured by pathogen growth. These different estimates of plant susceptibility are equivalent if disease symptoms are a direct consequence of pathogen growth, and if fitness loss to pathogens is directly correlated with disease symptoms. While such relationship is commonly assumed in evolutionary models (eg in models of the evolution of virulence in pathogens), we still have very little information about the relationship between these different components of host response to infection at the genetic and molecular level. Previous studies have used simultaneous mapping to test for pleiotropic effects on correlated traits and possible selection trade-offs (eg Ungerer *et al*, 2002). Although testing for pleiotropy is notoriously difficult (Cheverud *et al*, 1997; Knott and Haley, 2000), all QTL identified in this study were in such distinct regions of the genome that we can rule out pleiotropy. This result suggests that symptom severity, bacterial growth and fruit production in infected plants are under independent genetic control. However, given the power limitations of QTL analysis, there are alternative interpretations for this result. For example, it is possible that two traits share many QTL in common, but that all QTL that they share have small enough effects that they are not detectable. Still, the results of the regression analysis further support the idea that the susceptibility traits studied here are not simply different measures of a single trait, since symptom severity and bacterial growth had partial regression coefficients on relative fitness with opposite signs. Thus, although our study cannot provide a final answer to whether these traits are genetically independent, it does suggest that future studies should pursue the nature of the relationship between these traits.

The consequences of pathogen infection for trait evolution in plants depend on the genetic architecture of disease-related traits and the relationship between those traits and fitness. Here, we investigated the relationship of symptom expression and bacterial growth with fitness using a regression analysis as suggested by Lande and Arnold (1983). This analysis estimates both the independent (ie, partial) effect of traits on relative fitness and their overall association with relative fitness.

In the analysis of the independent effects of traits on fitness, partial regression coefficients indicate the change in relative fitness associated with a standard deviation change in the value of a trait, holding the values of all other traits in the model constant. The partial regression analysis showed that all four traits had a significant relationship with fitness. The signs of the regression coefficients indicate that, given our experimental conditions, infected plants are under selection to have larger rosettes, flower earlier, reduce bacterial growth and increase symptom severity. The direction and magnitude of selection on rosette size and flowering time is similar to what has been reported in other studies (reviewed in Pigliucci, 2003). Selection for less bacterial growth is not surprising considering that infection by *P. syringae* has been shown to reduce fitness in *A. thaliana* in other studies (Kover and Schaal, 2002; Korves and Bergelson, 2003). However, the positive relationship between symptom severity and fitness is very surprising. It indicates that all else being equal, plants with more severe symptoms produced more seeds. While such a positive association between symptom severity and fitness is nonintuitive, it could result from a real biological processes, such as 'overcompensation'. Overcompensation is a phenomenon originally described for interactions between plant and herbivores, where plants that are damaged by herbivores show higher seed production than nondamaged plants (reviewed in Stowe *et al*, 2000). While the mechanism behind overcompensation is not understood, it is quite possible that it also occurs in plant-pathogen interactions. We also cannot rule out the possibility that the observed positive association between symptom and fitness is caused by selection on a trait not included in our analysis, which is correlated with both fitness and symptom severity. In any case, the relationship between disease symptoms and fitness clearly needs further investigation. Should this pattern prove to be consistent, it may provide valuable insights into the evolution of the phenotypic manifestation of disease.

Univariate regression coefficients were used to estimate overall relationship between traits and relative fitness. These overall associations result from both the direct effects that traits have on fitness and the indirect effects caused by selection acting on correlated traits. The separate analyses of direct effects and overall associations of traits with fitness is important because some traits can have a direct effect on fitness, but no overall association with relative fitness due to correlations with other traits. We find such a case in this analysis, where bacterial growth and symptom severity show a positive phenotypic correlation and are under selection of similar magnitude but in opposite directions. The net result of this antagonistic relationship is that, although we see significant partial regression coefficients for these traits, we see no overall association between either of these susceptibility traits and fitness. This result suggests that symptom severity and bacterial growth would show no evolutionary change, despite both having independent effects on fitness, because their antagonistic relationship results in no net selection on either trait. As with any other selection study, the observed patterns of selection are dependent on the environmental conditions experienced by the plants in the growth chamber and only reflect plant response to infection at a specific developmental

stage. Although no phenotypic correlation was observed between either rosette size or flowering time and the two susceptibility-related traits in our study, other studies have shown that infected plants tend to flower earlier than uninfected plants (Peters, 1999; Korves and Bergelson, 2003). Since we found that selection favors earlier flowering in infected plants, it is possible that selection on tolerance-related traits (ie traits that alter the effect of infection on fitness) are more important than resistance-related traits (ie traits that affect the degree to which plants become infected).

## Acknowledgements

We thank M Dyer and the greenhouse staff of Washington University for maintaining our plant during the experiment, R Collier for technical assistance, A Kloek for providing us with seeds and B Schaal for her interest and support. We also thank Kent Holsinger and two anonymous reviewers for invaluable comments and suggestions that much improved this manuscript. PXX acknowledges support from an NSF minority post-doctoral grant, and JBW was supported by NSF DEB-0236956.

## References

- Alexander HM (1992). Evolution of disease resistance in natural plant populations. In: Fritz RS, Simms E (eds) *Plant Resistance to Herbivores and Pathogens*. The University of Chicago press: Chicago, pp 326–344.
- Alonso-Blanco C, El-Assal SE-D, Coupland G, Koornneef M (1998). Analysis of natural allelic variation at flowering time loci in the Landsberg *erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. *Genetics* **149**: 749–764.
- Ballare C, Scopel A (1997). Phytochrome signalling in plant canopies: testing its population-level implications with photoreceptor mutants of *Arabidopsis*. *Funct Ecol* **11**: 441–450.
- Basten C, Weir BS, Zeng ZB (1994). Zmap-a QTL cartographer. In: Smith C, Gavora JS, Benkel B, Chesnais J, Fairfull W *et al* (eds) *The Organizing Committee, 5th World Congress on Genetics Applied to Animal Production*. Guelph, Ontario, Canada. Vol 22, pp 65–66.
- Beavis W (1994). The power and deceit of QTL experiments: lessons from comparative QTL studies. *Proceedings of the 49th Annual Corn and Sorghum Research Conference*, pp 250–266.
- Bell C, Ecker J (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**: 137–144.
- Bergelson J, Purrington CB (1996). Surveying patterns in the cost of resistance in plants. *Am. Nat.* **148**: 536–558.
- Broman KW (2001). Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal* **30**: 44–52.
- Chen Z, Kloek A, Boch J, Katagiri F, Kunkel B (2000). The *Pseudomonas syringae* avrRpt2 gene product promotes pathogen virulence from inside plant cells. *Mol Plant-Microbe Interact* **13**: 1312–1321.
- Cheverud J (2001). A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* **87**: 52–58.
- Cheverud J, Routman E (1995). Epistasis and its contribution to genetic variance. *Genetics* **139**: 1455–1461.
- Cheverud JM, Routman EJ, Irschick DJ (1997). Pleiotropic effects of individual gene loci on mandibular morphology. *Evolution* **51**: 2006–2016.
- Churchill G, Doerge R (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.
- Clay K, Van der Putten WH (1999). Pathogens and plant life histories. In: Vuorisalo T, Mutikainen P (eds) *Life History Evolution in Plants*. Kluwer: Dordrecht, pp 275–301.
- Dangl J, Jones JD (2001). Plant pathogens and integrated defence responses to infection. *Nature* **411**: 826–830.
- Doebley J, Stec A, Gustus G (1995). *Teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**: 333–346.
- Garcia-Hernandez M, Berardini TZ, Chen G, Crist D, Doyle A, Huala E *et al* (2002). TAIR: a resource for integrated *Arabidopsis* data. *Funct Integr Genom* **2**: 239.
- Geiger H, Heun M (1989). Genetics of quantitative resistance to fungal diseases. *Annu Rev Phytopathol* **27**: 317–341.
- Gilliland L, Kandasamy M, Pawloski L, Meagher R (2002). Both vegetative and reproductive actin isovariants complement the stunted root hair phenotype of the *Arabidopsis act-2* mutation. *Plant Physiol* **130**: 2199–2209.
- Glazebrook J (1999). Genes controlling expression of defense responses in *Arabidopsis*. *Curr Opin Plant Biol* **2**: 280–286.
- Glazebrook J, Chen W, Estes B, Chang H, Nawrath C, Metraux J *et al* (2003). Topology of the network integrating salicylate and jasmonate signal transduction derived from global expression phenotyping. *Plant J* **34**: 217–228.
- Glazebrook J, Rogers EE, Ausubel FM (1996). Isolation of *Arabidopsis* mutants with enhanced disease susceptibility by direct screening. *Genetics* **143**: 973–982.
- Glazebrook J, Rogers EE, Ausubel FM (1997). Use of *Arabidopsis* for genetic dissection of plant defense responses. *Annu Rev Genet* **31**: 547–569.
- Green R, Tingay S, Wang Z, Tobin E (2002). Circadian rhythms confer a higher level of fitness to *Arabidopsis* plants. *Plant Physiol* **129**: 576–584.
- Greenberg JT, Ausubel FM (1993). *Arabidopsis* mutants compromised for the control of cellular damage during pathogenesis and aging. *Plant J* **4**: 327–341.
- Jakob K, Goss E, Araki H, Van T, Kreitman M, Bergelson J (2002). *Pseudomonas viridiflava* and *P. syringae*: natural pathogens of *Arabidopsis thaliana*. *Mol. Plant Microbe Interact* **15**: 1195–1203.
- Juenger T, Purugganan M, Mackay TFC (2000). Quantitative trait loci for floral morphology in *Arabidopsis thaliana*. *Genetics* **156**: 1379–1392.
- Kloek A, Verbsky M, Sharma S, JE JS, Vogel J, Klessig D *et al* (2001). Resistance to *Pseudomonas syringae* conferred by an *Arabidopsis thaliana* coronatine-insensitive (*coi1*) mutation occurs through two distinct mechanisms. *Plant J* **26**: 509–522.
- Knott S, Haley C (2000). Multitrait least squares for quantitative trait loci detection. *Genetics* **156**: 899–911.
- Korves TM, Bergelson J (2003). A developmental response to pathogen infection in *Arabidopsis*. *Plant Physiol* **133**: 339–347.
- Kover PX, Caicedo AL (2001). The genetic architecture of resistance and the role of parasites in maintaining sexual recombination. *Mol Ecol* **10**: 1–17.
- Kover PX, Schaal BA (2002). Genetic variation for disease resistance and tolerance among *A. thaliana* accessions. *Proc Natl Acad USA* **99**: 11270–11274.
- Kunkel B, Brooks D (2002). Cross talk between signalling pathways in pathogen defense. *Curr Opin Plant Biol* **5**: 325–331.
- Kunkel BN, Bent AF, Dahlbeck D, Innes RW, Staskawicz BJ (1993). RPS2, an *Arabidopsis* disease resistance locus specifying recognition of *Pseudomonas syringae* strains expressing the avirulence gene *avrRpt2*. *The Plant Cell* **5**: 865–875.
- Lande R, Arnold S (1983). The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.
- Lander ES, Botstein D (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- Lander ES, Green S, Abrahamson P, Barlow A, Daley M *et al* (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.

- Lister C, Dean C (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J* **4**: 745–750.
- Long AD, Lyman RF, Langley CH, Mackay TFC (1998). Two sites in the delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149**: 999–1017.
- Mackay T, Fry J (1996). Polygenic mutation in *Drosophila melanogaster*: genetic interactions between selection lines and candidate quantitative trait loci. *Genetics* **144**: 671–688.
- Mauricio R, Rausher M (1997). Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense. *Evolution* **51**: 1435–1444.
- Meyers B, Kozik A, Griego A, Kuang H, Michelmore R (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**: 809–834.
- Mitchell-Olds T, Pedersen D (1998). The molecular basis of quantitative genetic variation in central and secondary metabolism in *Arabidopsis*. *Genetics* **149**: 739–747.
- Peters AD (1999). The effects of pathogen infection and mutation on life history. *J Evolution Biol* **12**: 460–470.
- Pigliucci M (2003). Selection in a model system: ecological genetics of flowering time in *Arabidopsis thaliana*. *Ecology* **84**: 1700–1712.
- Quesada V, Garcia-Martinez S, Piqueras P, Ponce MR, Micol JL (2002). Genetic architecture of NaCl tolerance in *Arabidopsis*. *Plant Physiol* **130**: 951–963.
- Routman EJ, Cheverud JM (1997). Gene effects on a quantitative trait: two locus epistatic effects measured at microsatellite markers and at estimated QTL. *Evolution* **51**: 1654–1662.
- Schiff CL, Wilson IW, Sommerville SC (2001). Quantitative trait loci analysis of powdery mildew disease resistance in *Arabidopsis thaliana*: 2. Columbia-0 and Warschau-1 accessions. *Plant Pathol* **50**: 690–701.
- Stowe KA, Marquis RJ, Hochwender CG, Simms EL (2000). The evolutionary ecology of tolerance to consumer damage. *Ann Rev Ecol Syst* **31**: 565–595.
- Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003). Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**: 74–77.
- Ungerer MC, Halldorsdottir SS, Modliszewski JL, Mackay TFC, Purugganan M (2002). Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. *Genetics* **160**: 1133–1151.
- Vaughn T, Pletscher L, Peripato A, King-Ellison K, Adams E, Erikson C et al (1999). Mapping quantitative trait loci for murine growth: a closer look at genetic architecture. *Genet Res* **74**: 313–322.
- Whalen M, Innes R, Bent A, Staskawicz B (1991). Identification of *Pseudomonas syringae* pathogens of *Arabidopsis* and a bacterial locus determining avirulence on both *Arabidopsis* and soybean. *Plant Cell* **3**: 49–59.
- Wilson IW, Schiff CL, Hughes DE, Somerville SC (2001). Quantitative trait loci analysis of powdery mildew disease resistance in the *Arabidopsis thaliana* accession Kashmir-1. *Genetics* **158**: 1301–1309.
- Wolfe M (2000). Crop strength through diversity. *Nature* **406**: 681–682.
- Wyss G, Muller-Scharer H (1999). Infection process and resistance in the weed pathosystem *Senecio vulgaris* – *Puccinia lagenophorae* and implications for biological control. *Can J Bot* **77**: 361–369.