Genetic basis of *Arabidopsis thaliana* responses to infection with naïve and adapted isolates of turnip mosaic virus

Anamarija Butković[1,†,§], Thomas James Ellis[2,†], Rubén González[1,†,#], Benjamin Jaegle[2], Magnus Nordborg[2,*], Santiago F. Elena[1,3,*]

[1] Instituto de Biología Integrativa de Sistemas (I[2]SysBio), CSIC-Universitat de València, Parc Científic UV, Catedrático Agustín Escardino 9, Paterna, 46980 València, Spain,

[2] Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna BioCenter, Doktor-Bohr-Gasse 3, 1030 Vienna, Austria,

[3] The Santa Fe Institute, Santa Fe, 1399 Hyde Park Road, NM 87501, USA.

[†]These authors contributed equally

[§]Current address: Archaeal Virology Unit, Institut Pasteur, 75015 Paris, France.

[#]Current address: Institut de Biologie de l'Ecole Normale Supérieure, CNRS, INSERM, 75005 Paris, France.

[*]**Corresponding authors' e-mail:** magnus.nordborg@gmi.oeaw.ac.at and santiago.elena@csic.es

ORCIDs:

A.B.: 0000-0002-1435-0912

T.J.E.: 0000-0002-8511-0254

R.G.: 0000-0001-9588-1127

B.J.: 0000-0002-2679-1328

M.N.: 0000-0001-7178-9748

S.F.E.: 0000-0001-8249-5593

Summary

Plant viruses account for enormous agricultural losses worldwide, and the most effective way to combat them is to identify genetic material conferring plant resistance to these pathogens. Aiming to identify genetic associations with responses to infection, we screened a large panel of *Arabidopsis thaliana* natural inbred lines for four disease-related traits caused by infection by *A. thaliana*-naïve and -adapted isolates of the natural pathogen turnip mosaic virus (TuMV). We detected a strong, replicable association in a 1.5 Mb region on chromosome 2 with a 10-fold increase in relative risk of systemic necrosis. The region contains several plausible causal genes as well as abundant structural variation, including an insertion of a *Copia* transposon into an TIR-NBS-LRR coding gene, that could be either a driver or a consequence of the disease-resistance locus. When inoculated with TuMV, loss-of-function mutant plants of this gene exhibited different symptoms than wild-type plants, being the magnitude and sign of the difference dependent on the degree of adaptation of the viral isolate to *A. thaliana*. This increase in symptoms severity was specific for infections with the naïve isolate. Necrosis-associated alleles are found worldwide, and their distribution is consistent with a trade-off between resistance during viral outbreaks and a cost of resistance otherwise, leading to negative frequency-dependent selection.

Keywords: emerging viruses; experimental evolution; frequency-dependent selection; GWAS; *Potyvirus*; resistance genes; virus evolution

Introduction

Plant viruses represent an enormous threat to crop yields and food security (Tomlinson, 1987; Oerke, 2006; Jones, 2021). Infected plants are difficult to treat, so it is of vital importance to identify genetic material which is resistant to infection (Monnot et al., 2021). In spite of this, the genetic architecture of plant responses to viral infections has received much less attention than the response to bacterial and fungal pathogens (Bartoli & Roux, 2017; Monnot et al., 2021). In agricultural settings, plants predominantly form monocultures, thus allowing for more virulent and specialized viruses that cause more detrimental effects on the host (McDonald & Stukenbrock, 2016; González et al., 2019). Through their specialization in one host species, viruses also evolve better counter defenses than their naive counterparts (Brosseau et al., 2020).

Among plant viruses, members of *Potyviridae* family are of special interest. Potyviruses affect a wide variety of crops, especially from the families *Brassicaceae* and *Solanaceae*, and are among the most widespread crop pathogens worldwide (Revers & García, 2015). Turnip mosaic virus (TuMV species *Turnip mosaic virus*, genus *Potyvirus*, family *Potyviridae*) is among the most damaging of the potyviruses (Tomlinson, 1987), and also has a high incidence in wild populations of *Arabidopsis thaliana* (L.) HEYNH (Pagán et al., 2010). *A. thaliana* is a perfect model organism for investigating the mechanisms behind viral resistance in plants because of the genetic resources available. Also, because of its near genetic resemblance to the agriculturally significant brassicas, this system permits the study of a variety of crop viruses, thereby broadening the scope of research (Pagán et al., 2010; Ouibrahim & Caranta, 2013). Viral outbreaks are frequent in natural populations of *A. thaliana*, and there is substantial genetic variation in resistance, indicating that viral coevolution represents a meaningful selection pressure in this species (Pagán et al., 2010). Several previous genome-wide association (GWA) studies of the response of *A. thaliana* to viral infection have been carried out. Rubio et al. (2019) used TuMV and 317 lines grown under field conditions, while Butković et al. (2021) also used TuMV but 450 lines kept in laboratory conditions. Montes et al. (2021) and by Liu et al. (2022) used 156 and 496 inbred lines respectively to measure the response of *A. thaliana* to infection by cucumber mosaic virus under controlled conditions. These studies have demonstrated that genetic variation for virus response exists, and that individual loci with large effects on virus response segregate in *A. thaliana*.

84 Here, we report the results of GWA studies using two isolates of TuMV and one of
85 the largest (1,050) set of *A. thaliana* inbred lines studied so far. We compare an
86 "ancestral" isolate of TuMV that was obtained from a calla lily (*Zantedeschia* sp.) and
87 was naïve to *A. thaliana* (Chen et al., 2003), to its "evolved" descendant that had been
88 submitted to12 serial passages of experimental evolution on *A. thaliana* Col-0 line
89 (González et al., 2019). The two virus isolates differ in their adaptation to Col-0 and their
90 infection phenotypes, where the evolved virus induces stronger symptoms and faster
91 disease progression. We aim to identify plant genes that play a role in TuMV infection
92 and whether some of these genes may respond differentially to each viral isolate.
93
94

95 # Materials and methods

96 **Viruses and inoculation procedure**

97 We obtained the ancestral virus from infected *Nicotiana benthamiana* DOMIN plants
98 inoculated with a transcript product from a p35STunos infectious plasmid that contains
99 TuMV genome cDNA (GenBank line AF530055.2), corresponding to YC5 isolate from
100 calla lily (Chen et al., 2003). This cDNA was under the control of the cauliflower mosaic
101 virus 35S promoter and a NOS terminator. We obtained the evolved virus via 12 passages
102 of experimental evolution on *A. thaliana* line Col-0; see González et al. (2019) for full
103 details. The evolved virus has two mutations (CI/T1293I and VPg/N2039H) not present
104 in its antecessor. In both cases, symptomatic tissues were pooled, frozen with liquid N_2
105 and homogenized into a fine powder using a Mixer Mill MM400 (Retsch GmbH, Haan,
106 Germany). For inoculations, we mixed 100 mg of powdered liquid-nitrogen-frozen
107 infected tissue in 1 mL of phosphate buffer (50 mM phosphate pH 7.0, 3% PEG6000,
108 10% Carborundum) and rubbed 5 µL of this mixture into three central rosette leaves per
109 plant. To minimize the noise that significant variations in vegetative development could
110 produce, we infected all the lines at roughly the same growth stage; when they reached
111 growth stage 3.2 - 3.5 in the Boyes et al. (2001) scale. In our growing conditions the lines
112 used in this study reached that stage approximately 21 days after germination.
113

114 **Plant growth and phenotyping**

115 We screened a total of 1,050 lines from the Arabidopsis 1001 Genomes Project (Alonso-
116 Blanco et al., 2016) for disease phenotypes (Appendix S1). We grew selected lines in

117 climatic chambers under a long-day regime (16 hours light/8 hours darkness) with 24 ºC
118 day and 20 ºC night, 45% relative humidity and light intensity of 125 μmol m^{-2} s^{-1} using
119 a 1:3 mixture of 450 nm blue and 670 nm purple LEDs. Due to limited growth-chamber
120 space and labor force, we inoculated and phenotyped lines in four independent cohorts of
121 4,800 plants each. Five lines in the fourth block that did not reach the proper size on the
122 day of the inoculation were inoculated four days after the other lines. We inoculated eight
123 plants per line with the ancestral virus and another eight with the evolved virus, as well
124 as two mock-inoculated plants that served as negative controls. We placed four lines
125 inoculated with each viral isolate and the corresponding mock-inoculated plants in the
126 same tray and randomized tray positions every day.

127 We measured four disease-related traits daily for 21 days post inoculation (dpi), at
128 which point infection phenotypes reached a steady plateau. We measured infectivity,
129 AUDPS (an analysis of the disease progression using the number of infected plants during
130 the 21 dpi (Simko & Piepho, 2012), severity of symptoms (based on a scale showed in
131 Fig. 1A), and necrosis (systemic necrosis or the most severe stage of disease). Necrosis
132 was considered present in a line if at least one plant showed full-leaf systemic necrosis,
133 as shown with number 5 in Fig. 1A.

134 To test for statistically significant differences in disease phenotypes in response to
135 the two viruses we constructed linear models using the phenotype as the response
136 variable, and viral isolate and block as explanatory variables. For infectivity and necrosis,
137 we used a Binomial generalized linear model (GLM) with a logit link function. For
138 AUDPS, we log transformed the data to improve normality, and fitted a GLM with a
139 Gaussian link function. For severity of symptoms, we applied a cumulative link model
140 with a logit link function using the R package "ordinal"
141 (https://github.com/runehaubo/ordinal). Because *P*-value estimation is sensitive to
142 departures from normality on the scale of the link function, we calculated a null
143 distribution for the difference due to viral isolate by permutation. For each phenotype we
144 randomly shuffled the response variable 1000 times and refitted the linear model. The
145 (one-sided) *P*-value is then the proportion of permuted samples with a greater difference
146 in phenotype between the two viral isolates than was found for the observed data. Since
147 it was previously found that the evolved virus induced a stronger response in the reference
148 line Col-0 (Butković et al., 2020), we report one-sided *P*-values.

149 To validate the association between necrosis and symptom severity with the
150 candidate region on chromosome 2, we tested an additional cohort including all 51 lines
151 that had previously shown necrosis in response to one or both viruses together with 67
152 randomly chosen non-necrotic lines from the previous screen. This confirmation cohort
153 was inoculated and grown in the same way as the four previous cohorts.
154 In addition, we selected two candidate genes within this region, *AT2G14080*
155 (encoding for a nucleotide-binding-site leucine-rich repeat disease-resistance protein with
156 an N-terminal Toll/interleukin receptor (TIR-NBS-LRR)) and *AT2G14120* (encoding for
157 *DYNAMIN RELATED PROTEIN 3B* (*DRP3B*)) and ordered T-DNA knock-out (KO)
158 mutants from the NASC stock center (http://arabidopsis.info/BrowsePage). Mutants
159 were chosen on the following criteria: (*i*) have to be in the Col-0 background, (*ii*) must
160 be T-DNA inserts that cause gene knock-outs, (*iii*) must be homozygous, and (*iv*) last but
161 not least, they must have already been related to virus infection. We used the SALK
162 homozygous line SALK_105230C (NASC ID: N656836) to study *AT2G14080* and
163 SALK_112233C (N667049) for *DRP3B*. Visual inspection of inoculated plants was done
164 daily for 21 dpi, annotating the number of infected plants and the severity of their
165 symptoms. Symptoms severity data were fitted to a generalized linear mixed model
166 (GLMM) with a Poisson probability distribution and a logarithmic link function. Plant
167 genotype and virus isolate were treated as orthogonal inter-individual factors and dpi as
168 intra-individual factor (repeated measures). The main effects as well as their interactions
169 were evaluated using type III sum of squares. These analyses were done with SPSS
170 version 28.0.1.0 (IBM, Armonk NY, USA).
171
172 **Genetic associations**
173 We used the Python package LIMIX 3.0.3 (Lippert et al., 2014) in Python 3.7.3 to
174 perform multi-trait genome-wide association analyses of disease-related traits based on
175 the statistical framework developed in (Korte et al., 2012; Segura et al., 2012). We looked
176 for associations with the 10,709,949 SNPs published by Alonso-Blanco et al. (2016). We
177 used a liberal minor-allele-frequency cut-off of 3% to allow us to detect rare variants
178 associated with necrosis, since only 3.7% and 4.9% of lines showed necrosis in response
179 to the ancestral and evolved viruses respectively, leaving 2,257,599 SNPs. We accounted
180 for cohort effects by including these as fixed cofactors, and for population structure by
181 modeling the covariance between lines due to relatedness as a random effect (Bergelson

182 et al., 2001; Kang et al., 2008). In a second analysis, we repeated this GWA for necrosis
183 including the genotype at chromosome 2 position 5,927,469 as a covariate to account for
184 the confounding effects of this locus (Korte et al., 2012). For all GWA, we used a
185 significance threshold of $-\log P = 7.65$, corresponding to a Bonferroni-corrected P-value
186 of 0.05.

187 We estimated narrow-sense SNP heritability (h^2_{SNP}) using the R package "Sommer"
188 4.1.4 (Covarrubias-Pazaran, 2016) by regressing phenotypes onto the matrix of pairwise
189 relatedness between individuals (Yang et al., 2010). To estimate variance explained by
190 the major association for necrosis, and to assess the sensitivity of GWA and heritability
191 to the presence of this locus we also repeated both GWA and heritability estimates
192 including genotypes at the SNP most strongly associated with necrosis as a covariate.
193

194 **Structural variation around locus *AT2G14080***

195 **Confirming the insertion using long-read genome assemblies**

196 We examined haplotype structure in the 225 Kb surrounding the region associated with
197 necrosis using data from 53 full genomes sequenced using PacBio long reads
198 (Arabidopsis 1001 Genomes Consortium, pers. comm.), and assembled using the
199 methods previously described for the 1001 Genomes Plus Project (Koren et al., 2017).
200 We extracted sequences around the peak of association close to *AT2G14080*. We
201 identified the precise location of that gene, as well as eight annotated genes and
202 transposable elements on each side from the Araport 11 database
203 (https://www.araport.org/), using BLAST (Camacho et al., 2009) if they showed 70%
204 percent identity and were at least 70% of the length of the gene or transposable element
205 in the reference genome. When comparing whole genomes, it is not meaningful to plot
206 SNP positions with the coordinate system of a reference genome, so to visualize the
207 results we plotted haplotypes on a relative scale centered between *AT2G14050* and
208 *AT2G14120* (*DRP3B*), since these genes were largely syntenic between lines and flanked
209 most of the region associated with necrosis. Finally, we sorted lines based on the distance
210 between those two genes (Fig. 4).

211 To confirm the insertions observed in the necrotic/susceptible lines 351, 870, 7346
212 and 9,102, we mapped the raw PacBio reads to the new assembly using Minimap2 (Li,
213 2021) with the option "-ax map-pb" and checked if any breaks coincided with the
214 insertion position that would suggest an artifact of the assembly. This was not the case.

215

216 **Genotyping the insertion in the 1001 Genomes**

217 We attempted to confirm the presence of the insertion in the full sample using existing
218 short-read sequencing data (Alonso-Blanco et al., 2016). We used multiple approaches,
219 all based on the flanking sequence of the insertion. First, using two different mappers,
220 bwa-mem (http://bio-bwa.sourceforge.net) as well as bowtie (Langmead et al., 2009), we
221 mapped all the short-read data to the insertion sequence plus 5 Kb of flanking sequence
222 (on either side) from line 870, filtering the bam files using samtools "rmdup" (Bonfield
223 et al., 2021) to remove duplicates. For each line, the number of paired reads mapping to
224 each side of the insertion borders were extracted and counted. We assumed that the
225 presence of such reads would confirm the presence of the insertion, whereas lines without
226 the insertion would only have paired reads flanking the whole insertion. However, the
227 mapping quality at the flanking regions was too low for this approach to work. Many
228 flanking regions had no coverage from short-read sequencing (even when duplicated
229 reads were kept). The reasons for this are not clear, but presumably reflect high levels of
230 repetitiveness and high levels of polymorphism in the region. In a second approach, we
231 use bbmap (https://sourceforge.net/projects/bbmap) to look for the exact sequence at the
232 insertion site (±50 bp) within the raw fastq file. However, when we used the four lines
233 with a confirmed insertion from *de novo* long-read assemblies as a control, this approach
234 only worked in two cases. Again, the reasons are not clear, but the approach is clearly
235 too unreliable for genotyping.

236

237 **Geographic distribution of major and minor necrosis alleles**

238 We plotted the worldwide distribution of necrosis phenotypes and genotypes at the most
239 strongly-associated SNP using R packages "maps" and "mapdata" (Becker et al., 2018,
240 2021). We defined "resistant" and "susceptible" genotypes at each associated SNP based
241 on the allele most strongly associated with systemic necrosis. To test whether the lines
242 harboring susceptible alleles are more or less geographically dispersed than would be
243 expected by chance, we first calculated the mean distances between all pairs of lines with
244 the susceptible alleles, and compared this to mean distances between lines with the minor
245 allele at 10,000 randomly chosen lines with minor allele counts between 10 and 200
246 (minor allele counts at significant ranged from 33 to 178).

247

248

Results

Increased disease phenotypes in response to the evolved virus

We inoculated 1,050 lines of *A. thaliana* (Alonso-Blanco et al., 2016) with the ancestral (Chen et al., 2003) or evolved (González et al., 2019) isolates of TuMV and characterized the response to each virus based on four phenotypes: (1) infectivity (proportion of plants showing symptoms), (2) AUDPS (area under the disease progress stairs, a measure of disease progress (Simko & Piepho, 2012)), (3) severity of symptoms on a semi-quantitative scale from 0 - 5 (Fig. 1A) (Butković et al., 2020), and (4) necrosis (a binary trait reflecting presence/absence of systemic necrotic symptoms). Note that along the manuscript we refer to necrosis as the most severe stage of infection: systemic necrosis that results in the death of the plant (Fig. 1A). We do not refer to the hypersensitive response that induces local necrotic spots and limits the spread of a pathogen.

Lines infected with the evolved virus showed more severe disease symptoms than those infected with the ancestral virus. On average, we found a 27% increase in infectivity (Fig. 1B; $P < 0.001$), 37% increase in AUDPS (Fig. 1C; $P < 0.001$), and 23% increase in severity of symptoms (Fig. 1D; $P < 0.001$). We also found a 31% increase in necrosis (Fig. 1E), although this is not statistically significant when cohort effects are taken into account ($P = 0.08$). None of the 2,100 mock-inoculated plants showed any symptoms.

Although the severity of disease phenotypes increased in response to the evolved virus, the nature of that increase varied between phenotypes (Fig. 1B - 1E). Off-diagonal elements of the correlation matrix between phenotypes were largely symmetric, with strong correlations between AUDPS and infectivity, and weak correlations between necrosis and other traits, indicating that the overall pattern of correlations between phenotypes in response to a single virus were very similar (Fig. 1F). However, although necrosis was highly repeatable between viruses, we found only moderate correlations for AUDPS, infectivity and severity of symptoms in response to the two viruses (Fig. 1F, diagonal elements). This suggests that disease phenotypes in response to each virus varied between *A. thaliana* lines. AUDPS and infectivity showed low to moderate h^2_{SNP} for both isolates, with substantially higher heritability for necrosis (> 50%) and severity of symptoms (> 40%) in response to the evolved virus (Fig. 1G).

280

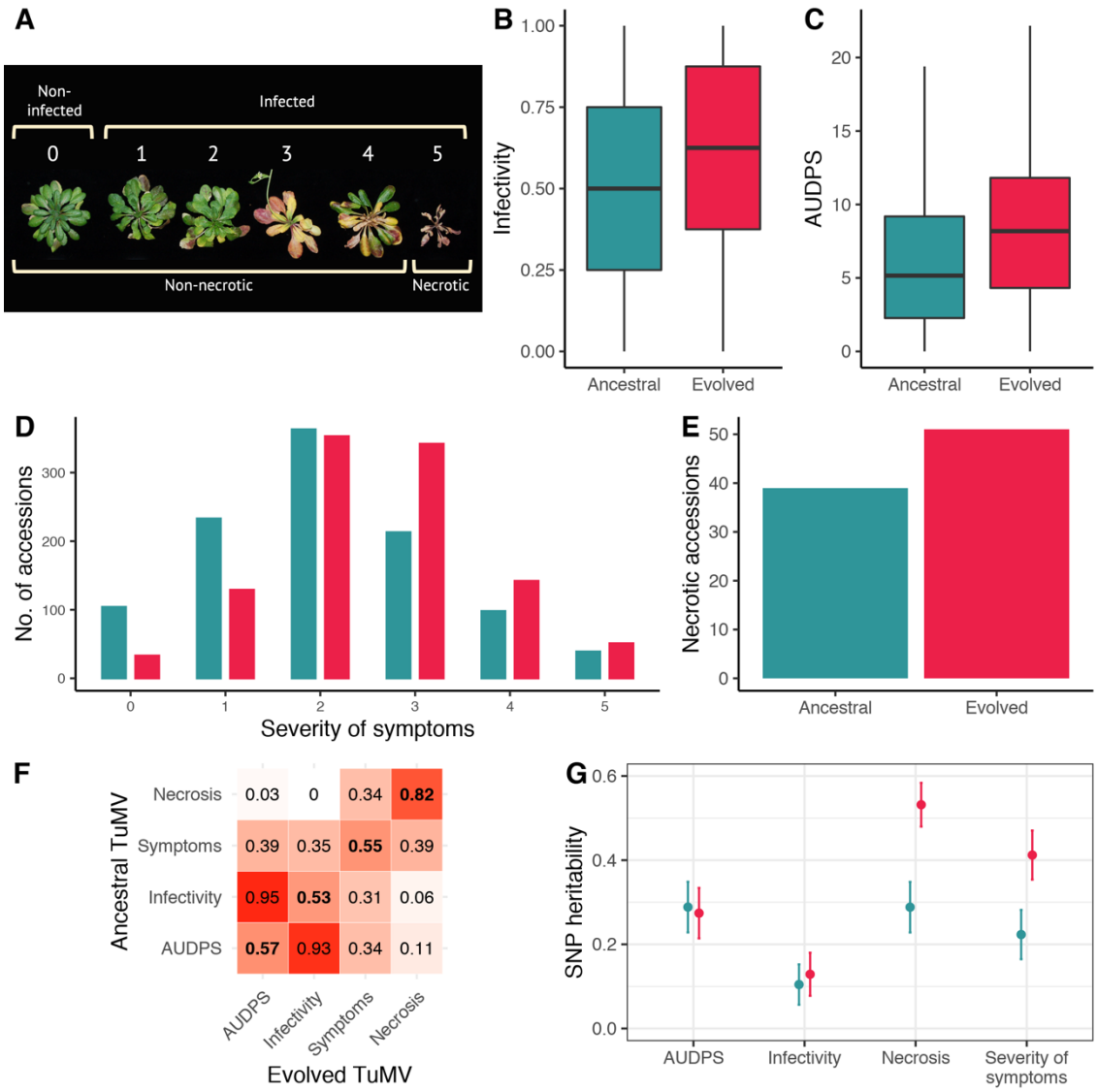


Figure 1. Disease phenotypes. (A) Illustration of the scale used to evaluate the severity of symptoms: (0) no symptoms or healthy plant (1) mild symptoms without chlorosis (2) visible chlorosis, (3) advanced chlorosis, (4) strong chlorosis and incipient necrosis, (5) necrosis and death of the plant. (B - E) Infection phenotypes across lines in response to the ancestral (blue) and evolved (red) TuMV isolates: infectivity (B), AUDPS (C), intensity of symptoms (D), and necrosis (E). (F) Correlation matrix between disease phenotypes in response to the ancestral (upper left) and evolved (lower right) isolates. The diagonal shows correlations between the same phenotype in response to each viral isolate. (G) SNP heritability (h^2_{SNP}) for each disease-related trait in response to each viral isolate.

When considering the way each line responds to infection with the ancestral or evolved TuMV isolate (Fig. 2A), we observed that the evolved isolate was significantly more infective (Fig. 2B; Friedman test: $\chi^2 = 528.122$, 2 d.f., $P < 0.001$), which means that adaptation to the Col-0 line leads to a virus isolate often more infective in other lines.

287 Indeed, the number of lines that did not show any symptoms after inoculation with each
288 viral isolate are largely different: 91 lines in case of the ancestral and only 11 for the
289 evolved. Twenty-two lines show no symptoms of infection regardless the viral isolate.
290

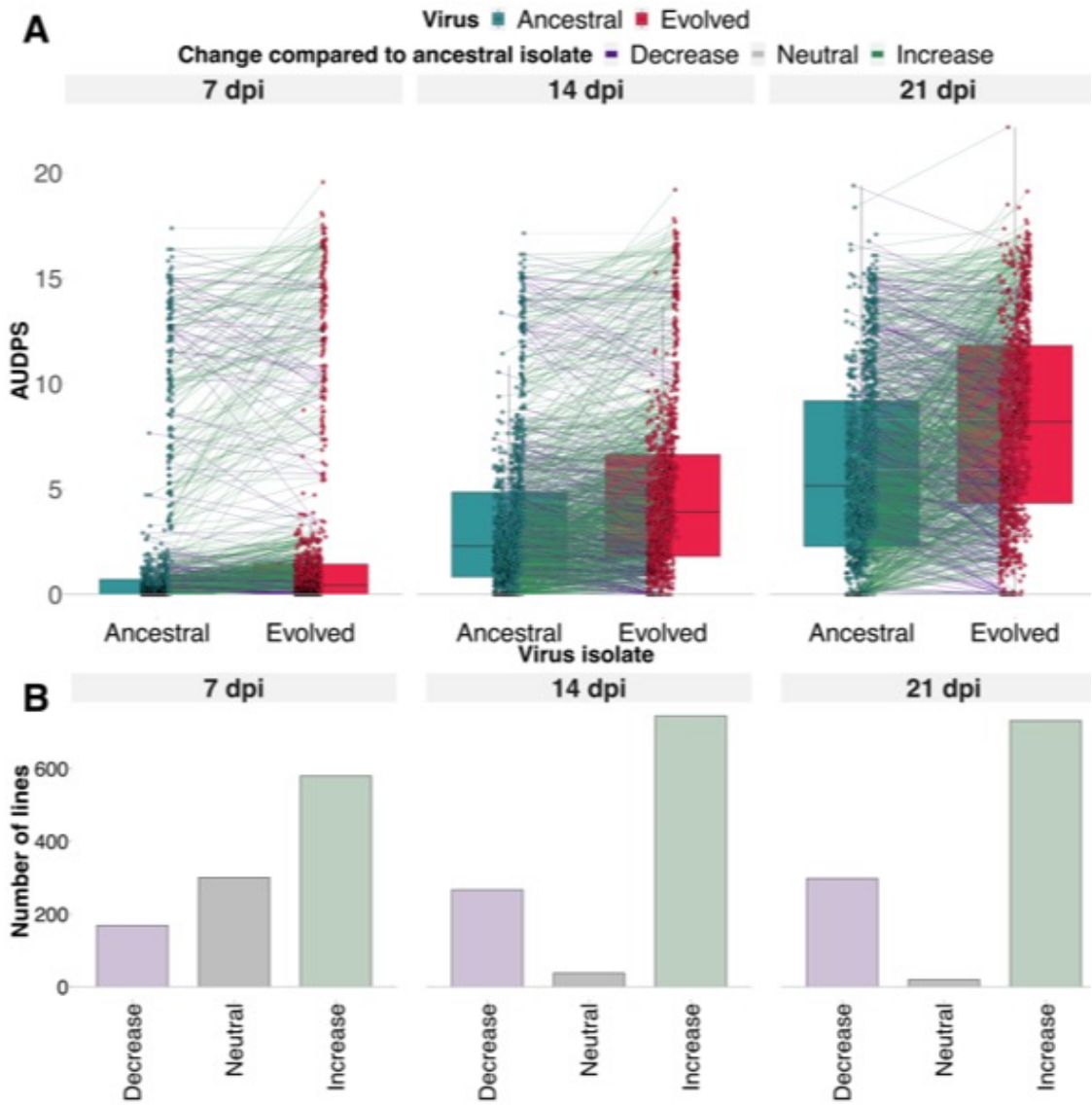


Figure 2. (A) The progression of disease (measured as AUDPS) was compared at different time points between the ancestral and evolved viral isolates in multiple *A. thaliana* lines. Each data point represents a different line, and the segments connecting the same line are color-coded green if the AUDPS has increased for the evolved virus, or purple if it has decreased. (B) Number of *A. thaliana* lines exhibiting increased, decreased, or nonsignificant changes in susceptibility when inoculated with the evolved or ancestral isolates.

291

292 **A major locus associated with necrosis and severity of symptoms**

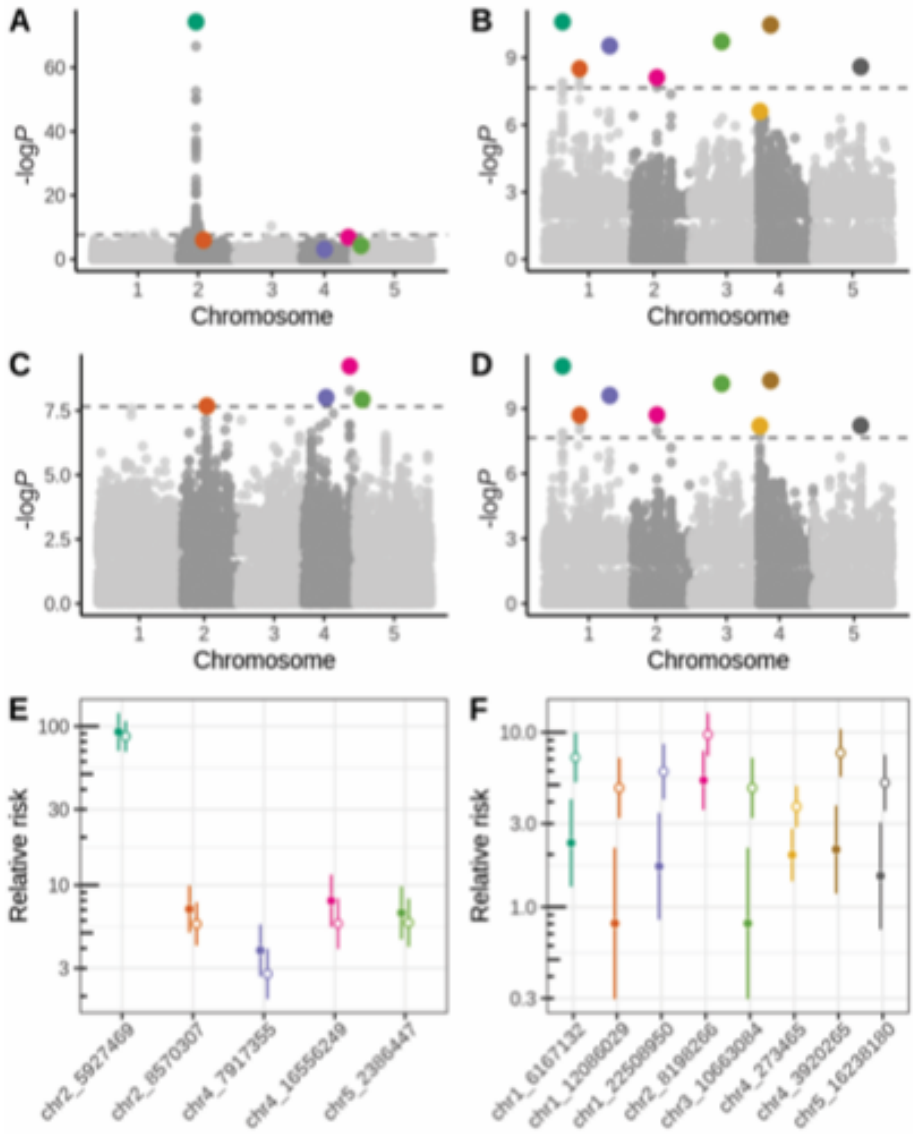
293 We used the multi-trait GWA analysis implemented in the software package LIMIX to
294 identify individual genetic loci that correlate with the response to each virus (Lippert et
295 al., 2014). LIMIX assesses the response to each virus jointly, and identifies loci
296 associated with (1) a shared response to both viruses, and (2) specific responses to
297 individual viral isolates.

298 We found a single strong peak of association for necrosis in response to both viruses
299 on chromosome 2 (Figs. 3A - 3D). The region of association encompasses five annotated
300 genes, and is within 300 bp of two more, as well as at least six annotated transposable
301 elements (Fig. 3, Table 1). 56% and 49% of the lines showing necrosis in response to the
302 ancestral and evolved viruses respectively harbor the susceptible allele at the most
303 strongly associated SNP at position 5,927,469. To validate this association, we repeated
304 the experimental procedure using 51 lines that showed systemic necrosis in the initial
305 experiment and 65 that did not, and recovered the association in the same region on
306 chromosome 2, with a peak at position 5,927,469 (Fig. S1). These results demonstrate
307 that a major association on chromosome 2 explains approximately half the cases of
308 necrosis.

309 The peak of association with necrosis overlaps with a peak found for severity of
310 symptoms (Fig. S2). Since necrosis is the final stage of disease progress in the semi-
311 quantitative scale used to score symptoms, this overlap in genetic architecture likely
312 reflects the same biological signal. For this reason, and because ordinal data are hard to
313 handle statistically, we focus on necrosis for the rest of this paper. We did not identify
314 any significant associations with AUDPS or infectivity (Figs. S2 and S3), consistent with
315 the low heritability for these traits (< 30%; Fig. 1B).

316 The presence of a strong association can interfere with the detection of other
317 associations (Lippert et al., 2014). To identify additional loci associated with systemic
318 necrosis we therefore repeated the GWA analysis including the genotype at the most
319 strongly associated SNP as a covariate. This revealed four additional loci showing a
320 significant association with the shared response to both TuMV isolates (Figs. 3A and 3C),
321 and eight loci with an isolate-specific response (Figs. 3B and 3D). Linkage
322 disequilibrium between these loci is weak (Fig. S4). Necrotic alleles at these loci were
323 rare (3.1 - 17.0% allele frequency; median = 3.6%) and showed up to 9.7-fold increase in
324 the relative risk of developing systemic necrosis compared to the non-necrotic alleles
325 (Table S1, Figs. 3E and 3F). Where loci displayed a virus-specific effect, this was

326 associated with increased relative risk of systemic necrosis in response to the evolved
327 viral isolate compared to that of the ancestral virus one (Fig. 3F). Overall, all but two
328 necrosis-developing lines harbored the necrotic allele in at least one of the loci identified.
329



330

Figure 3. Genetic associations with systemic necrosis. (A), (C) and (E) show SNPs associated with a common response to both viruses. (B), (D) and (F) show SNPs associated with a differential response to each viral isolate. (A) and (B) show associations with raw necrosis data. (C) and (D) show associations in a model that includes the genotype at the most strongly associated SNP in (A) as a cofactor. (E) and (F) show relative risk of necrosis in response to the ancestral (closed symbols) and evolved (open symbols) TuMV isolates for the susceptible alleles at (E) the five loci associated with the response to both isolates and (F) the eight loci associated with an isolate-specific response. Note the log scale of the ordinate axis. Colors in (E) and (F) correspond to those in the Manhattan plots above.

331

332 **Loss-of-function mutants of *AT2G14080* and *DRP3B* show TuMV isolate-dependent**
333 **differences in the severity of symptoms**
334 We assessed the impact of associated loci on response to virus infection by analyzing the
335 severity of symptoms in two KO mutants mapped within 1.5 Mb of the candidate region
336 of chromosome 2: the TIR-NBS-LRR protein encoded by *AT2G14080* and *DRP3B*. We
337 fitted a GLMM to assess the effect of infection with each viral isolate in the severity of
338 symptoms shown by the wild-type and both mutant plants. Focusing in the intra-
339 individual effects, the duration of infections shows a highly significant effect ($F_{2,162}$ =
340 40.481, $P < 0.001$), increasing a 328.6% ±82.2 (±1 SEM) in severity between 7 and 21
341 dpi (Fig. 4). No other intra-individual effect was found significant (*i.e.*, interaction of the
342 factors with dpi; in all cases $P \geq 0.430$). Regarding the inter-individual effects, all three
343 were significant. First, differences between the three plant genotypes exist in the average
344 severity of symptoms ($F_{2,162}$ = 5.379, P = 0.006), with the plants *drp3b* showing
345 symptoms 30.0% ±5.4 milder than wild-type plants but *at2g14080* showing 16.6% ±13.3
346 stronger symptoms than wild-type plants (Fig. 4). The average stronger symptoms in
347 *at2g14080* plants are consistent with the antiviral role of the encoded TIR-NBS-LRR
348 protein, while the average weaker symptoms in *drp3b* is consistent with a proviral
349 activity. Second, the two viral isolates also induce symptoms that significantly differ in
350 severity ($F_{1,162}$ = 9.001, P = 0.003), with the evolved isolate inducing 45.6% ±15.9 more
351 severe symptoms, on average, than the ancestral one (Fig. 4). Finally, and more
352 interestingly, a significant interaction between plant genotype and viral isolate exists
353 ($F_{2,162}$ = 3.907, P = 0.022). This effect means that the magnitude of the difference in
354 symptoms induced by the evolved and ancestral TuMV isolates depends on the host
355 genotype (Fig. 4). Whereas in wild-type plants, the evolved isolate induces symptoms
356 that are 124.9% ±47.3 stronger than those induced by the ancestral, in *at2g14080* plants
357 the difference is largely reduced to a non-significant 1.53% ±13.8. By contrast, the
358 symptoms induced by the evolved isolate in *drp3b* plants are 35.3% ±27.7 stronger than
359 what the ancestral isolate does.
360

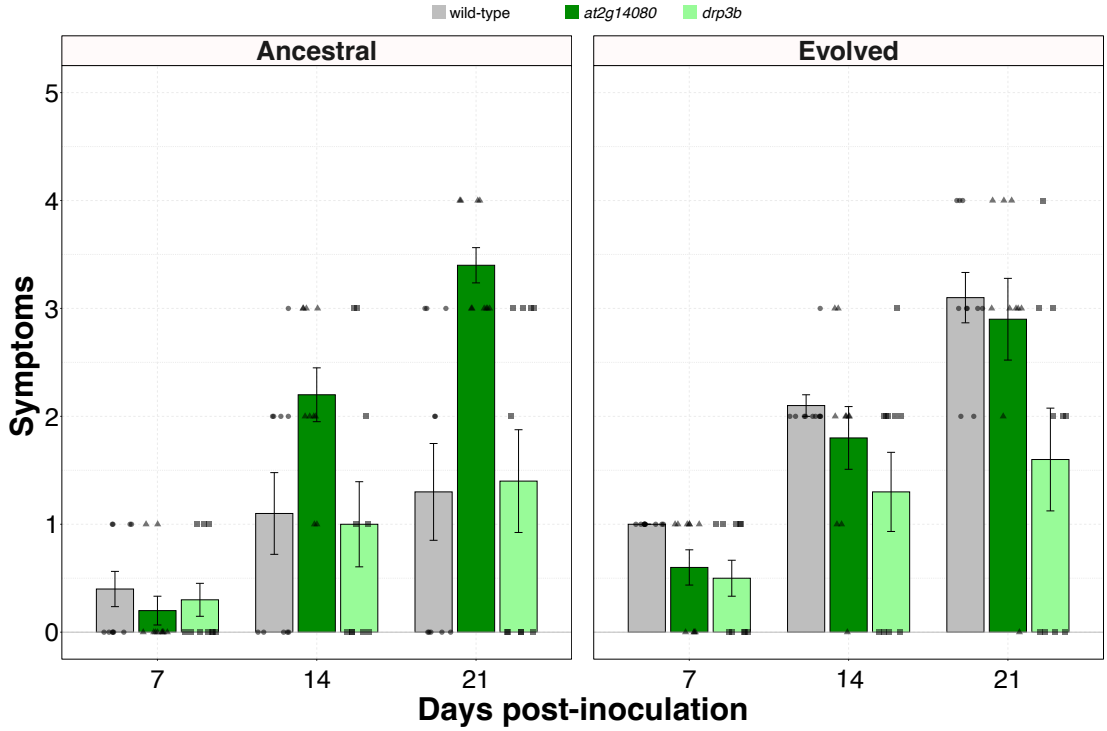


Figure 4. Symptoms severity at different stages of infection for the ancestral and the evolved TuMV isolates inoculated in wild-type, *at2g14080* and *drp3b* plants. Error bars represent ±1 SEM.

361

362 **A *Copia* element insertion in the primary candidate gene**

363 We explored haplotypes in the region associated with necrosis in 53 genomes assembled
364 *de novo* from long-read PacBio sequencing data as part of a different study
365 (https://1001genomes.org/). We aligned the sequences in the region around the strong
366 association for necrosis and severity of symptoms. To assess patterns of synteny, we
367 looked for homology with the two most strongly associated genes (*AT2G14080* and
368 *DRP3B*) and the eight annotated genes or transposable elements from the TAIR10
369 genome annotation on either side of these genes using BLAST. This analysis revealed
370 many structural polymorphisms (Fig. 5), including abundant intergenic indel variation,
371 duplications and a large presence/absence polymorphism downstream of *AT2G14080*
372 containing *AT2G14100* and a block of transposable elements. At least one line (9470)
373 shows a large-scale inversion for the entire region. The region around the strong
374 association for necrosis and severity of symptoms appears to be a hotspot for structural
375 variation.

376

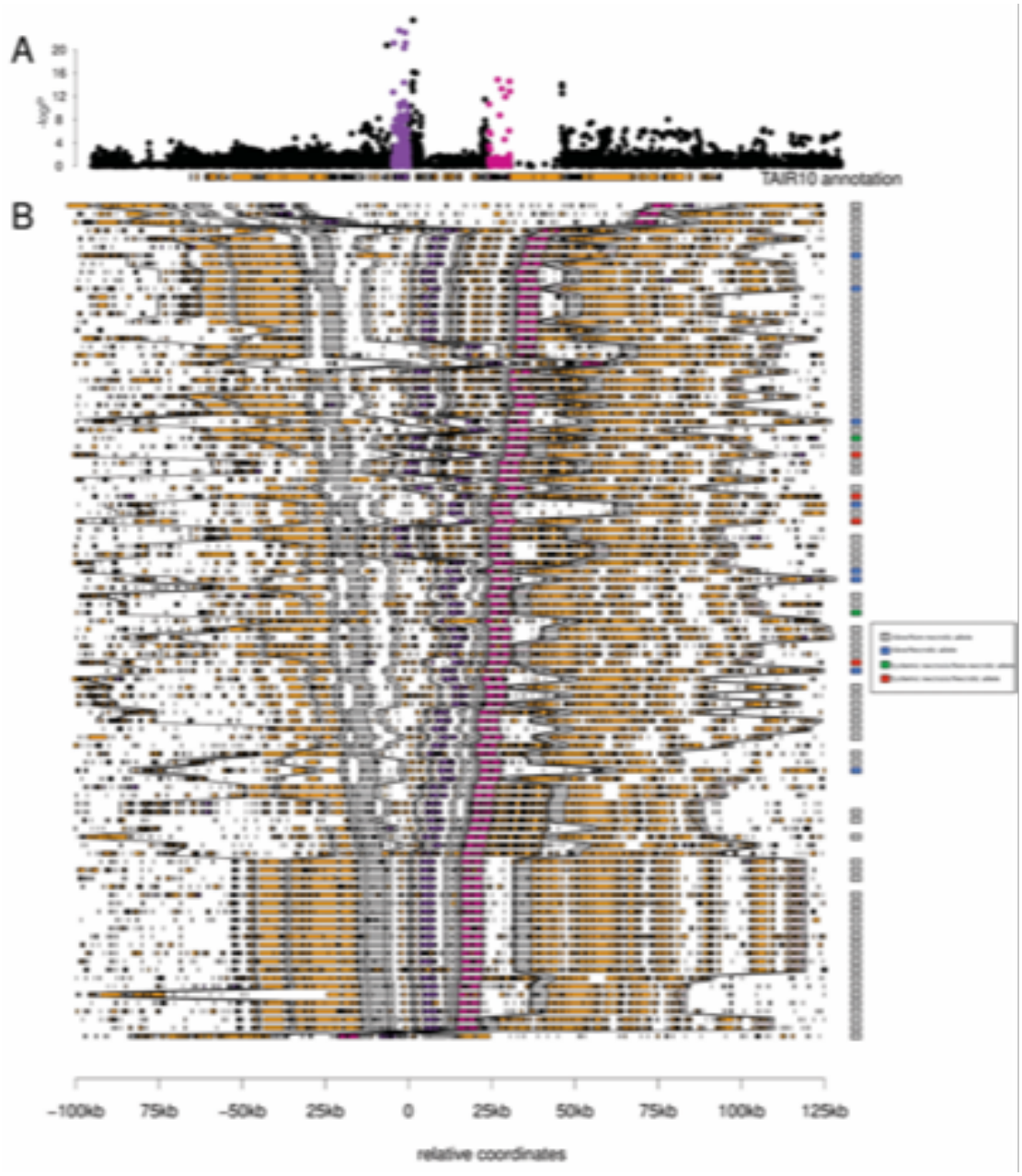


Figure 5. Multiple sequence alignment around the major association with necrosis on chromosome 2. (A) Enlarged view of the peak in Fig. 3A, with the TAIR10 annotation below. (B) A summary of structural variation in 53 assembled genomes. Transposable elements are shown in orange and coding genes in gray, with candidate genes *AT2G14080* and *DRP3B* highlighted in purple and pink respectively. Black lines connect homologous features. Each genome is labeled (by squares on the right) according to its necrosis phenotype and genotype at the most associated SNP. The lines without boxes have not been phenotyped.

We next looked for structural polymorphism that might be causal for the disease phenotypes. We categorized lines as "alive" *vs.* "systemic necrosis" based on whether they showed systemic necrosis in response to either virus, and as having the necrotic *vs.*

381 non-necrotic alleles at the most strongly associated SNP (Fig. 3). Unfortunately, only
382 four lines of the 53 sequenced both showed systemic necrosis and had the necrotic allele
383 at the major association, but all of them harbored a *Copia*-element inside the first intron
384 of *AT2G14080* that was present in only one other line. We tried to genotype the
385 presence/absence of this element in the full sample using short-read sequencing data, but
386 this turned out to be unachievable due to the repetitiveness of the sequence (see Materials
387 and methods). Thus, this element remains a promising candidate for a causal
388 polymorphism.
389
390 **The major susceptibility allele is spatially overdispersed**
391 The susceptible allele at the major association with necrosis on chromosome 2 is spread
392 throughout the global range of the GWA panel (Fig. 6A). To investigate whether this
393 broad spatial distribution could be due to chance, we compared the mean geographic
394 distance between lines with the susceptible alleles at each of the top SNPs identified by
395 GWA (Fig. 3) to the mean distances between lines with the minor allele at 10,000
396 randomly chosen loci with similar minor allele counts. For most associated loci, distances
397 between lines with susceptible alleles are well within the distribution of randomly chosen
398 loci, consistent with genetic drift allowing minor alleles to reach modest frequency on a
399 local spatial scale. In contrast, lines harboring the susceptible alleles at the major
400 association on chromosome 2 are further apart on average than 98.5% of randomly chosen
401 alleles at similar frequencies (Fig. 5C). This allele is found at a similar frequency in all
402 but three admixture groups identified by reference (Alonso-Blanco et al., 2016), so this
403 overdispersion cannot be accounted for by an association with a particular ancestry with
404 a high dispersal rate (Fig. 6B). We found similar spatial overdispersion for the associated
405 loci on chromosome 4 at positions 273,465 and 7,917,355. Alleles associated with
406 systemic necrosis at the major association, thus seem to be spatially overdispersed, in a
407 way which is independent of genetic background, and which is unlikely to be due to
408 genetic drift alone.
409

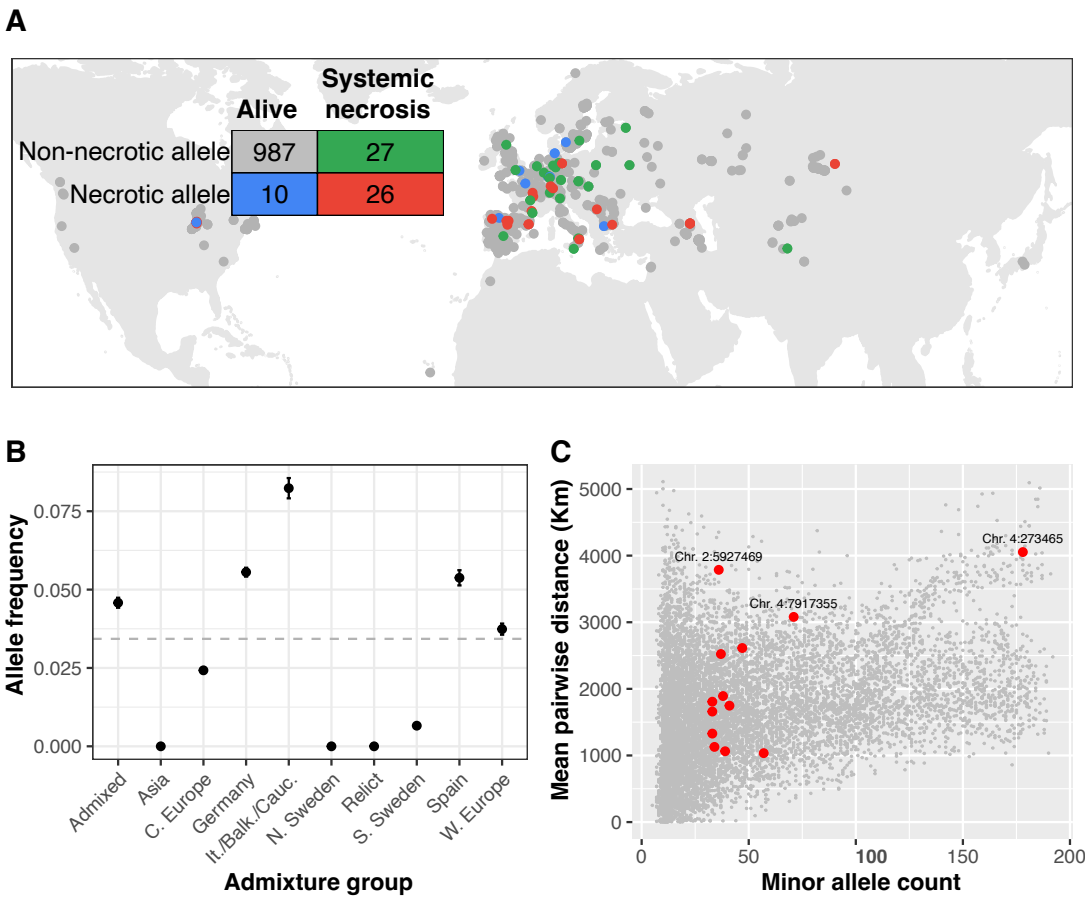


Figure 6. Global distribution of necrotic alleles. (A) Global distribution of genotypes (non-necrotic/necrotic alleles at the SNP most strongly associated with systemic necrosis), and phenotypes (whether lines showed systemic necrosis or not). The inset table shows the numbers of each genotype/phenotype; colors correspond to those on the map. (B) Frequencies of necrotic alleles at chromosome 2 position 5,927,469 within each admixture group. The horizontal line indicates global frequency. (C) Mean distances between pairs of loci harboring the necrotic allele at the 13 associated loci (red) and between minor alleles at 10,000 randomly chosen loci.

Discussion

We report results from the largest GWA study done so far aiming to identify *A. thaliana* genes involved in the infection response to one of its natural pathogens, TuMV, using an ancestral isolate naïve to *A. thaliana*, and a second isolate that had been experimentally evolved on *A. thaliana* line Col-0. The comparison of the results for the two isolates has pinpointed candidate host targets for virus adaptation.

Increased virulence of the evolved virus

420 Lines infected with the evolved virus showed increased disease symptoms compared to
421 plants infected with the ancestral virus (Fig. 1D - 1G). Systemic necrosis was strongly
422 correlated between virus treatments (Fig. 1C), and there was no evidence for a virus-
423 specific effect of alleles at the major association locus on chromosome 2 (Fig. 3E). The
424 increase in necrosis can thus probably be attributed to increased virulence of the evolved
425 virus. In contrast, the increase in AUDPS, infectivity and severity of symptoms was
426 associated with changes in the rank order of lines, reflected in only moderate correlations
427 between phenotypes in response to the two viral isolates (Fig. 1C). This suggests that the
428 change in these phenotypes reflects a more complex line-specific response to naïve and
429 adapted TuMV isolates.

430 The increase in virulence of the evolved isolate used in this study is associated with
431 two non-synonymous mutations acquired during adaptation to *A. thaliana* (Navarro et al.,
432 2022). First, mutation T1293I was found in the cylindrical inclusion protein (CI) that is
433 involved in viral genome replication and cell-to-cell movement (Deng et al., 2015), and
434 has been shown to interact with the photosystem I PSI-K protein in *A. thaliana* (Jiménez
435 et al., 2006). Second, mutation N2039H was found in the viral genome-linked protein
436 (VPg) that is involved in replication and intracellular movement (Wu et al., 2018), and is
437 a hub for interactions with other viral proteins (Bosque et al., 2014). Mutations in VPg
438 have been pervasively observed in evolution experiments of potyviruses in *A. thaliana*,
439 in all cases resulting in increased infectivity, virulence and viral load (Agudelo-Romero
440 et al., 2018; Hillung et al., 2014; González et al., 2021; Navarro et al, 2022; Melero et al.,
441 2023). Together, these observations indicate that genome replication and viral movement
442 underlie the increased virulence of the evolved virus.

443 Our study confirms that the adaptation to one susceptible *A. thaliana* line, in this
444 case Col-0, enables potyviruses to perform better in otherwise poorly susceptible lines.
445 This observation was first made by Lalić et al. (2010) after evolving a tobacco etch virus
446 isolate from tobacco in the susceptible line L*er*-0. Upon inoculation of 18 other lines,
447 they observed that the evolved isolate exhibited greater infectivity compared to the
448 ancestral one.

449

450 **The genetic architecture of TuMV resistance**

451 We identified a 1.5 Mb region on chromosome 2 that was strongly associated with the
452 severity of symptoms, and accounted for approximately half the cases of systemic

453 necrosis. This region encompasses several genes (Table 1), two of which are especially
454 strong candidates for causing the phenotype. First, *AT2G14080* encodes a known TIR-
455 NBS-LRR family protein. These proteins recognize and block translation of viral
456 transcripts in the cytoplasm, induce hypersensitive response leading to cell death, act as
457 a second line of defense if the initial antiviral response fails to restrict virus accumulation,
458 and are involved in signaling pathways (Li et al., 2001; Meyers et al., 2003; Bhattacharjee
459 et al., 2009; Kobayashi et al., 2010; Marone et al., 2013; van de Weyer et al., 2019). TIR-
460 NBS-LRR genes are the most numerous disease-resistance genes in the *A. thaliana*
461 genome and are under diversifying selection to respond to as broad a spectrum of
462 pathogens as possible (Ellis et al., 2003). This gene appears to have a significant role in
463 development of systemic necrosis during infection with both TuMV strains. Second,
464 *DRP3B* encodes a self-assembling GTPase involved in fission and fusion of membranes
465 in mitochondria and peroxisomes (Fujimoto et al., 2009; Wu et al., 2018). Its analogue,
466 *DRP2B,* has been shown to be co-opted by TuMV to enhance virulence, and treatment
467 with dynamin-specific inhibitors suppress growth by interfering with VPg (Wu et al.,
468 2018). As previously noted, one of the two amino-acid differences between the ancestral
469 and evolved TuMV isolates is in VPg (Navarro et al., 2022). These observations make
470 *AT2G14080* and *DRP3B* strong candidate genes for a role in viral resistance, either
471 independently or in tandem. To test this possibility, the ancestral and evolved isolates
472 were inoculated into KO mutants *at2g14080* and *drp3b*. We observed that, compared to
473 wild-type plants, *at2g14080* plants showed different responses for each viral isolate:
474 while the ancestral isolate induced more severe symptoms, those induced by the evolved
475 virus were similar or weaker in the wild-type plants (Fig. 4). These findings demonstrate
476 a difference in the host genes involved in the response to the two viral isolates and suggest
477 that *AT2G14080* is efficient for the defense against the ancestral but not for the evolved
478 isolate. The evolved viral isolate may be able to overcome the response imposed by
479 *AT2G14080* due to the mutations acquired during their adaptation to Col-0, most probably
480 because of the changes in the VPg viral protein that allow it to evade detection by the
481 TIR-NBS-LRR protein encoded by *AT2G14080*. In the case of the ancestral isolate, we
482 found no meaningful differences in symptomatology between wild-type and *drp3b* plants.
483 In sharp contrast, *drp3b* plants show milder symptoms than wild-type plants, confirming
484 a possible proviral effect of DRP3B that has been optimized during the previous
485 evolutionary history in Col-0. The possibility of epistatic effects between these two

486 candidate genes is highly likely considering the complex interplay between genetic
487 factors and environmental cues in determining the phenotypic outcomes of viral
488 infections.

489 Twelve additional loci (besides the major association on chromosome 2) were
490 significantly associated with the response to one (5) or both (8) viruses (Fig. 3). Lines
491 with the susceptible alleles at these loci were associated with up to 9.7-fold increased risk
492 of systemic necrosis compared to lines with the resistant allele (Figs. 2E - 2F), and the
493 low linkage disequilibrium between these loci indicates that they segregate independently
494 of the major association on chromosome 2, and of each other (Fig. S4). However, these
495 alleles are also at very low global frequencies (Table S1). On the one hand, this is
496 consistent with the expectation that selection against such deleterious alleles keeps
497 susceptible alleles rare. On the other hand, there is a substantial risk that alleles at such
498 low frequency should be associated with a binary phenotype at similar frequency just by
499 chance. As such, caution is warranted in interpreting a causative role for these
500 associations without further evidence.

501 We found very little overlap with associations found in two previous GWA studies
502 of TuMV resistance in *A. thaliana*. First, Butković et al. (2021) identified genome-wide
503 associations for AUDPS, severity of symptoms and infectivity under very similar
504 experimental conditions to those used here, but using two viral isolates that differed in
505 their host range (one evolved as a specialist in more permissive *A. thaliana* lines and the
506 other evolved as a generalist able to successfully infect more resistant lines) and fewer
507 host lines. That analysis recovered the association with *AT2G14080* for both isolates,
508 and identified eleven additional loci associated with the response to one or both viruses.
509 These eleven loci identified were not recovered in our larger dataset. This lack of
510 consistency can either indicate that these associations were spurious artifacts of a limited
511 sample size or, alternatively, a true biological effect due to the different TuMV isolates
512 used by Butković et al. (2021) and in this study. Second, we did not detect any of the
513 genes associated with viral load or infectivity in 317 lines grown under field conditions
514 identified by Rubio et al. (2019), and nor did that study detect the association on
515 chromosome 2 reported here. In this case, the lack of overlap may be explained not only
516 by the differences in sample sizes, but also by the differences in traits measured, viral
517 isolate and environmental conditions.
518

519 **The *Copia*-element insertion is a candidate causative variant**

520 We identified a *Copia*-element polymorphism in the first intron of *AT2G14080,* which is
521 found in all the four lines with long-read genome assemblies that show systemic necrosis
522 in response to either virus (and only one other line). Given this strong association, and
523 the fact that first introns often harbor regulatory elements, this insertion is a strong
524 candidate for a causative mutation. Whether it turns out to be casual or not, the abundance
525 of structural variation in this region highlights the need for restraint when trying to
526 identify any causal polymorphisms from SNP data alone, and more generally the
527 limitation of reference-biased polymorphism data.

528 It is also interesting to consider whether structural variation is driven by selection
529 for diversification of resistance genes, or whether particular structural variants themselves
530 cause variation in disease resistance. It has been previously observed that plant disease-
531 resistance genes often form clusters of duplicated copies within the genome, which would
532 contribute to structural variation between genomes (Leister, 2004). In particular, NBS-
533 LRR genes, of which *AT2G14080* is an example, are known to have undergone expansion
534 by numerous large- and small-scale duplication and translocation events (Leister, 2004).
535 However, we did not observe widespread duplications of *AT2G14080* or its domains, nor
536 evidence of synteny with other regions of the genome that would indicate translocations,
537 so in this case a simple transposon insertion remains the most likely explanation.

538

539 **The distribution of susceptible alleles is consistent with frequency-dependent**
540 **selection**

541 Given that potyvirus outbreaks are common in nature (Pagán et al., 2010) and
542 susceptibility is unquestionably deleterious, it is curious that susceptible alleles should be
543 at sufficiently high frequency to be detectable. There are three explanations for this.
544 First, selection may be sufficiently weak or the virus geographically restricted, such that
545 alleles can arise in one location and increase in frequency by genetic drift. If this were
546 true, we would expect to see a clustered geographic distribution of susceptible alleles,
547 reflected in under dispersion of distances between pairs of susceptible lines. While this
548 is true for most of the loci with small effects on necrosis, susceptible alleles at the major
549 association on chromosome 2 are spatially over dispersed worldwide, and found
550 throughout unrelated lines (Fig. 6). These patterns are difficult to explain via genetic
551 drift. Second, it may be that the genomic instability in the region surrounding this

552 association leads to a high turnover of new structural mutations that decrease viral
553 resistance. This is implausible because this would not generate linkage with individual
554 SNPs that could be detected by GWA. Moreover, we found a striking concordance
555 between a single *Copia*-element insertion into *AT2G14080* and necrosis, suggesting that
556 only a single variant is responsible for increased necrosis in this region. Neither genetic
557 drift nor mutation-selection balance can thus explain the persistence of susceptible alleles
558 at the major association for necrosis.
559 An alternative explanation is that susceptible alleles are beneficial in the absence
560 of the virus. This would cause a fitness trade-off between alleles in this region, as has
561 been previously demonstrated for many other pathogen-resistance systems in plants
562 (Bergelson et al., 2001; Todesco et al., 2010). Potyvirus outbreaks are common in natural
563 *A. thaliana* populations, but ephemeral (Pagán et al., 2010), causing susceptible alleles to
564 decrease in frequency during outbreaks, but to increase at other times. This would lead
565 to negative frequency-dependent selection maintaining the susceptible allele at a
566 frequency in proportion to the frequency of TuMV outbreaks. Despite being at low
567 frequency worldwide, the susceptible allele is found across the geographic and
568 genealogical distribution of *A. thaliana*, indicating that the polymorphism must either be
569 very old and has been maintained by balancing selection over long periods, or that alleles
570 have spread by adaptive introgression during viral outbreaks. These data do not allow us
571 to distinguish these hypotheses, nor are they mutually exclusive, but both are consistent
572 with a fitness trade-off at this locus. We conclude that the worldwide geographic
573 distribution of susceptible alleles is therefore consistent with negative frequency-
574 dependent selection maintaining the susceptible allele at low, but non-zero, frequency.
575
576

577 # Data availability

578 Phenotype data and code to reproduce the analyses presented here are given at GitHub
579 (https://github.com/ellisztamas/tumv_ms) and Zenodo (doi: 10.5281/zenodo.5789488).
580 Sequence data are available from the 1001 Genomes Project website
581 (https://1001genomes.org/) (Alonso-Blanco et al., 2016). Unless otherwise stated
582 analyses and plotting were done in R 4.0.3 (R Core Team, 2021) under RStudio Server
583 Pro 1.3.1093-1 (RStudio Team, 2021).
584

585

586 # Supporting information

587 Table S1. Necrosis-associated loci, with indication on whether the association was
588 TuMV isolate-specific or not
589 Fig. S1. Manhattan plots for the association between SNPs and necrosis in the replicate
590 dataset of 116 lines.
591 Fig. S2. Manhattan plots for associations with common disease responses to both TuMV
592 isolates.
593 Fig. S3. Manhattan plots for associations with disease responses specific to individual
594 TuMV isolates.
595 Appendix S1. Full list of *A. thaliana* lines for the GWAS analysis. (EXCEL).
596
597

598 # Acknowledgements

599 We thank Paula Agudo, Francisca de la Iglesia and Joanna Gunis for excellent technical
600 support. Work was funded by Spain's Agencia Estatal de Investigación-FEDER grant
601 PID2019-103998GB-I00 and Generalitat Valenciana grants GRISOLIAP/2018/005 and
602 PROMETEO/2019/012 to SFE. RG were supported by Agencia Estatal de Investigación-
603 FEDER contract BES-2016-077078. MN was funded by ERC AdG 789037 -
604 EPICLINES, ERA-CAPS (FWF I 3684-B25) and by the GMI.
605
606

607 # Authors contributions

608 MN and SFE planned and designed the research. AB and RG performed experiments.
609 AB, TJE, RG, SFE and BJ analyzed data. AB, TJE, RG, MN, and SFE wrote the
610 manuscript. AB, TJE and RG contributed equally.
611
612

613 # References

614 Agudelo-Romero P, Carbonell P, Pérez-Amador MA, Elena SF. Virus adaptation by
615 manipulation of host's gene expression. PLoS ONE. 2008;3: e2397.

616 Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J. 1,135 Genomes reveal the
617 global pattern of polymorphism in *Arabidopsis thaliana*. Cell. 2016;166: 481–491.
618 Arabidopsis 1001 Genomes Consortium. In preparation.
619 Bartoli C, Roux F. Genome-wide association studies in plant pathosystems: toward an
620 ecological genomics approach. Front Plant Sci. 2017;8: 763.
621 Becker RA, Wilks AR, Brownrigg R, Minka T, Deckmyn A. maps: draw geographical
622 maps. 2021. Available: https://CRAN.R-project.org/package=maps.
623 Becker RA, Wilks AR, Brownrigg R. mapdata: Extra map databases. 2018. Available:
624 https://CRAN.R-project.org/package=maps
625 Bergelson J, Kreitman M, Stahl EA, Tian D. Evolutionary dynamics of plant *R* genes.
626 Science. 2001;292: 2281–2285.
627 Bhattacharjee S, Zamora A, Azhar MT, Sacco MA, Lambert LH, Moffett P. Virus
628 resistance induced by NB-LRR proteins involves *Argonaute4*-dependent translational
629 control. Plant J. 2009;58: 940–951.
630 Bonfield JK, Marshall J, Danecek P, Li H, Ohan V, Whitwham A, *et al*. HTSlib: C library
631 for reading/writing high-throughput sequencing data. GigaScience. 2021;10: giab007.
632 Bosque G, Folch-Fortuny A, Picó J, Ferrer A, Elena SF. Topology analysis and
633 visualization of *Potyvirus* protein-protein interaction network. BMC Syst Biol.
634 2014;8: 129.
635 Boyes DC, Zayed AM, McCaskill AJ, Hoffman NE, Davis KR. Growth stage-based
636 phenotypic analysis of Arabidopsis: a model for high throughput functional genomics
637 in plants. Plant Cell. 2001;13: 1499–1510.
638 Brosseau C, Bolaji A, Roussin-Léveillée C, Zhao Z, Biga S, Moffett P. Natural variation
639 in the Arabidopsis *AGO2* gene is associated with susceptibility to potato virus X. New
640 Phytol. 2020;226: 866–878.
641 Butković A, González R, Cobo I, Elena SF. Adaptation of turnip mosaic potyvirus to a
642 specific niche reduces its genetic and environmental robustness. Virus Evol. 2020;6:
643 veaa041.
644 Butković A, González R, Rivarez MPS, Elena SF. A genome-wide association study
645 identifies *Arabidopsis thaliana* genes that contribute to differences in the outcome of
646 infection with two turnip mosaic potyvirus strains that differ in their evolutionary
647 history and degree of host specialization. Virus Evol. 2021;7: veab063.

648 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J. BLAST+: architecture and
649 applications. BMC Bioinformatics. 2009;10: 421.
650 Caparrós-Martín JA, McCarthy-Suárez I, Culiáñez-Macià FA. HAD hydrolase function
651 unveiled by substrate screening: enzymatic characterization of Arabidopsis thaliana
652 subclass I phosphosugar phosphatase AtSgpp. Planta. 2013;237: 943–954.
653 Chen CC, Chao CH, Chen CC, Yeh SD, Tsai HT. Identification of turnip mosaic virus
654 isolates causing yellow stripe and spot on calla lily. Plant Dis. 2003;87: 901–905.
655 Covarrubias-Pazaran G. Genome-assisted prediction of quantitative traits using the R
656 package sommer. PLoS ONE. 2016;11: e0156744.
657 Deng P, Wu Z, Wang A. The multifunctional protein CI of potyviruses plays interlinked
658 and distinct roles in viral genome replication and intercellular movement. Virol J.
659 2015;12.
660 Ellis J, Dodds P, Pryor T. Structure, function and evolution of plant disease resistance
661 genes. Curr Opin Plant Biol. 2000;3: 278–284.
662 Fujimoto M, Arimura S, Mano S, Kondo M, Saito C. Arabidopsis dynamin-related
663 proteins DRP3A and DRP3B are functionally redundant in mitochondrial fission, but
664 have distinct roles in peroxisomal fission. Plant J. 2009;58: 388–400.
665 Garcia-Moreno M, Järvelin AI, Castello A. Unconventional RNA-binding proteins step
666 into the virus-host battlefront. WIREs RNA. 2018;9: e1498.
667 González R, Butković A, Elena SF. Role of host genetic diversity for susceptibility-to-
668 infection in the evolution of virulence of a plant virus. Virus Evol. 2019;5: vez024.
669 González R, Butković A, Escaray FJ, Martínez-Latorre J, Melero I, Pérez-Parets E, *et al*.
670 Plant virus evolution under strong drought conditions results in a transition from
671 parasitism to mutulism. Proc Natl Acad Sci USA. 2021;118: e2020990118.
672 Hillung J, Cuevas JM, Valverde S, Elena SF. Experimental evolution of an emerging
673 plant virus in host genotypes that differ in their susceptibility to infection. Evolution.
674 2014;68: 2467–2680.
675 Jiménez I, López L, Alamillo JM, Valli A, García JA. Identification of a plum pox virus
676 CI-interacting protein from chloroplast that has a negative effect in virus infection.
677 Mol Plant Microbe Interact. 2006;19: 350–358.
678 Jones RAC. Global plant virus disease pandemics and epidemics. Plants. 2021;10.

679 Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, *et al*. Efficient
680 control of population structure in model organism association mapping. Genetics.
681 2008;178: 1709–1723.
682 Kobayashi M, Seo S, Hirai K, Yamamoto-Katou Y, Katou S, Seto H, Meshi T, Mitsuhara
683 I, Ohashi Y. Silencing of WIPK and SIPK mitogen-activated protein kinases reduces
684 tobacco mosaic virus accumulation but permits systemic viral movement in tobacco
685 possessing the *N* resistance gene. Mol Plant-Microbe Interact. 2010;23: 1032–1041.
686 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable
687 and accurate long-read assembly via adaptive k-mer weighting and repeat separation.
688 Genome Res. 2017;27: 722–736.
689 Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q. A mixed-model approach for
690 genome-wide association studies of correlated traits in structured populations. Nat
691 Genet. 2012;44: 1066–1071.
692 Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment
693 of short DNA sequences to the human genome. Genome Biol. 2009;10: R25.
694 Lalić J, Agudelo-Romero P, Carrasco P, Elena SF. Adaptation of tobacco etch potyvirus
695 to a susceptible ecotype of *Arabidopsis thaliana* capacitates it for systemic infection
696 of resistant ecotypes. Phil Trans R Soc B 2010; 365:1997-2007.
697 Leister D. Tandem and segmental gene duplication and recombination in the evolution of
698 plant disease resistance genes. Trends Genet. 2004;20: 116–122.
699 Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37:
700 4572–4574.
701 Li X, Clarke JD, Zhang Y, Dong, X. Activation of an EDS1-mediated *R*-gene pathway in
702 the *snc1* mutant leads to constitutive, NPR1-independent pathogen resistance. Mol
703 Plant-Microbe Interact. 2001;14: 1131–1139.
704 Lippert C, Casale FP, Rakitsch B, Stegle O. LIMIX: genetic analysis of multiple traits.
705 bioRxiv. 2014; doi: 10.1101/003905.
706 Liu S, Chen M, Li R, Li, WX, Gal-On A, Jia Z, Ding SW. Identification of positive and
707 negative regulators of antiviral RNA interference in *Arabidopsis thaliana*. Nat
708 Commun. 2022;13: 2994.
709 Luan H, Niu H, Luo J, Zhi H. Soybean cytochrome b5 is a restriction factor for soybean
710 mosaic virus. Viruses. 2019;11: 546.

711 Marone D, Russo M, Laidò G, De Leonardis A, Mastrangelo A. Plant nucleotide binding
712 site-leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses.
713 Int J Mol Sci. 2013;14: 7302–7326.
714 McDonald BA, Stukenbrock EH. Rapid emergence of pathogens in agro-ecosystems:
715 global threats to agricultural sustainability and food security. Phil Trans R Soc B.
716 2016;371: 20160026.
717 Melero I, González R, Elena SF. Host developmental stages shape the evolution of a plant
718 RNA virus. Phil Trans R Soc B. 2023;378: 20220005.
719 Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of
720 NBS-LRR–encoding genes in Arabidopsis. Plant Cell. 2003;15: 809–834.
721 Monnot S, Desaint H, Mary-Huard T, Moreau L, Schurdi-Levraud V, Boissot N.
722 Deciphering the genetic architecture of plant virus resistance by GWAS, state of the
723 art and potential advances. Cells. 2021;10: 3080.
724 Montes N, Cobos A, Gil-Valle M, Caro E, Pagán I. *Arabidopsis thaliana* genes associated
725 with cucumber mosaic virus virulence and their link to virus seed transmission.
726 Microorganisms. 2021;9: 692.
727 Navarro R, Ambrós S, Butković A, Carrasco JL, González R, Martínez F, *et al*. Defects
728 in plant immunity modulate the rates and patterns of RNA virus evolution. Virus Evol.
729 2022;8: veac059.
730 Oerke EC. Crop losses to pests. J Agric Sci. 2006;144: 31–43.
731 Ouibrahim L, Caranta C. Exploitation of natural genetic diversity to study plant-virus
732 interactions: what can we learn from *Arabidopsis thaliana*? Natural virus resistance in
733 *Arabidopsis thaliana*. Mol Plant Pathol. 2013;14: 844–854.
734 Pagán I, Fraile A, Fernandez-Fueyo E, Montes N, Alonso-Blanco C, García-Arenal F.
735 *Arabidopsis thaliana* as a model for the study of plant–virus co-evolution. Phil Trans
736 R Soc B. 2010;365: 1983–1995.
737 R Core Team. R: A language and environment for statistical computing. R Foundation
738 for Statistical Computing; 2021.
739 Revers F, García JA. Molecular biology of potyviruses. Adv Virus Res. 2015;92: 101–
740 199.
741 RStudio Team. RStudio: Integrated development environment for R. Boston, MA:
742 RStudio, PBC; 2021.

743 Rubio B, Cosson P, Caballero M, Revers F, Bergelson J. Genome-wide association study
744 reveals new loci involved in *Arabidopsis thaliana* and turnip mosaic virus (TuMV)
745 interactions in the field. New Phytol. 2019;224: 2026–2038.
746 Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, *et al*. An efficient multi-
747 locus mixed-model approach for genome-wide association studies in structured
748 populations. Nat Genet. 2012;44: 825–830.
749 Simko I, Piepho HP. The area under the disease progress stairs: calculation, advantage,
750 and application. Phytopathology. 2012;102: 381–389.
751 Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M. Natural allelic variation
752 underlying a major fitness trade-off in *Arabidopsis thaliana*. Nature. 2010;465: 632–
753 636.
754 Tomlinson JA. Epidemiology and control of virus diseases of vegetables. Ann Appl Biol.
755 1987;110: 661–681.
756 Van de Weyer A-L, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, *et al*. A
757 species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. Cell.
758 2019;178: 1260-1272.e14.
759 Wu G, Cui X, Chen H, Renaud JB, Yu K. Dynamin-like proteins of endocytosis in plants
760 are coopted by potyviruses to enhance virus infection. J Virol. 2018;92: e01320–18.
761 Xu J, Wang X, Guo W. The cytochrome P450 superfamily: Key players in plant
762 development and defense. J Integ Agric. 2015;14: 1673–1686.
763 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, *et al.* Common
764 SNPs explain a large proportion of the heritability for human height. Nat Genet.
765 2010;42: 565–569.

Table 1. Significant SNPs and their corresponding genomic regions on chromosome 2 detected for necrosis in the multiple trait analysis for both viral isolates.

SNP	Locus	−log*P*	Function of significant genes	Reference
5,923,326	200 bp upstream of *AT2G14070*	20.55	Wound-responsive protein-like protein	-
5,928,864	*AT2G14080*	20.43	Disease-resistance protein of the TIR-NBS-LRR class lead to hypersensitive response and resistance	Meyers et al. (2003); Marone et al. (2013);
5,931,085	1.2 Kb upstream of *AT2G14095*	14.39	Hypothetical protein	-
5,933,529	*AT2G14095* and upstream of *AT2G14100*	8.42	A member of the cytochrome P450 family; plays roles in plant defense against pathogens and resistance.	Xu et al. (2015); Lua
5,952,757	*AT2G14110*	11.32	Haloacid dehalogenase-like hydrolase (HAD) proteins are involved in phosphate homeostasis	Caparrós-Martín et a
5,956,602	*AT2G14120*	14.62	Dynamin-related protein 3b suppresses TuMV replication and movement-	Wu et al. (2018)

5,975,892	200 bp downstream of *AT2G14160*	13.69	RNA-binding (RRM/RBD/RNP motifs) family protein; involved in inhibition of RNA virus replication, movement, and translation.	García-Moreno et al.

766

Table S1. Loci associated with necrosis, indicating whether the association was with a common response to both TuMV isolates, or with an isolate-specific response.

Effect	Chromosome	Position	Locus	Allele frequency	–log*P*	Function of significant genes
Isolate-specific	1	6,167,132	*AT1G17930*	0.036	10.96	*MAIN*, involved in silencing of transposable elements. Interacts and functions with *MAIL1* and *PP7L* in gene silencing
Isolate-specific	1	12,086,029	*AT1G33330*	0.031	8.70	Class I peptide chain release factor involved in translational termination
Isolate-specific	1	22,508,950	*AT1G61100*	0.310	9.61	Disease resistance protein of the TIR class
Common	2	5,927,469	*AT2G14080*	0.034	74.28	Disease resistance protein of the TIR-NBS-LRR family
Isolate-specific	2	8,198,266	*AT2G18917*	0.045	8.71	Anti-sense long-noncoding RNA overlaps with *AT2G18920* (hypothetical protein) and *ADAGIO2* (regulator of circadian rhythm)
Common	2	8,570,307	*AT2G19850*	0.054	7.68	Transcription repressor
Isolate-specific	3	10,661,084	-	0.031	10.16	Non-coding intergenic region
Isolate-specific	4	273,465	*AT4TE01410*	0.170	8.20	Transposable element in 5' of *AT4G00651* (Fe superoxide dismutase)
Isolate-specific	4	3,920,265	*AT4TE17125*	0.039	10.30	Gypsy-like retrotransposon family (Athila)
Common	4	7,917,355	*AT4TE34810*	0.070	8.01	Transposable element in 5' of *MATERNAL EFFECT EMBRIO ARREST 57*, a DNA methyltransferase regulated by Polycomb group complex during endosperm development
Common	4	16,556,249	*AT4G34690*	0.032	9.23	Transmembrane protein
Common	5	2,386,477	*GRP16*	0.037	7.93	*GLYCINE-RICH PROTEIN 16* is expressed in flowering during specific developmental stage
Isolate-specific	5	16,238,180	*AT5G40540*	0.035	8.21	Cytoplasmic protein kinase superfamily protein

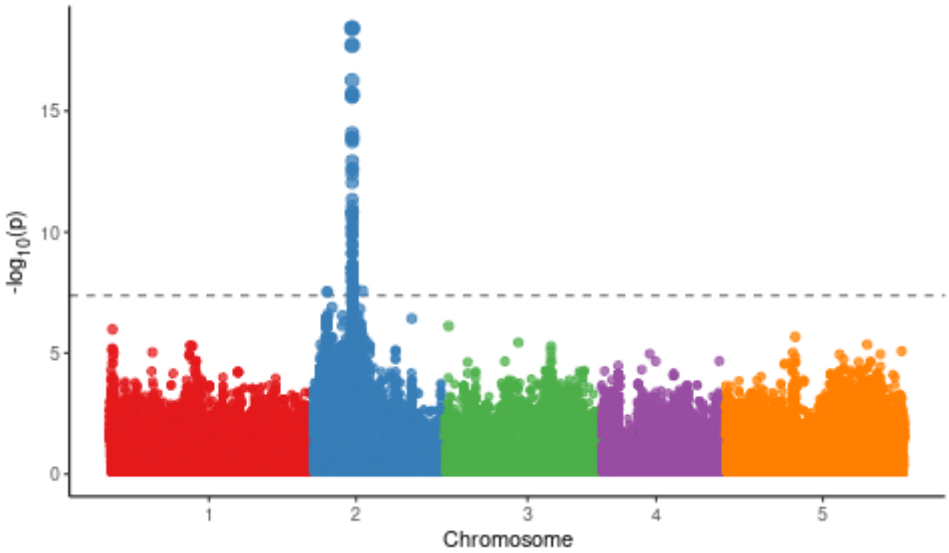


Figure S1. Manhattan plots for the association between SNPs and necrosis in the replicate dataset of 116 lines.

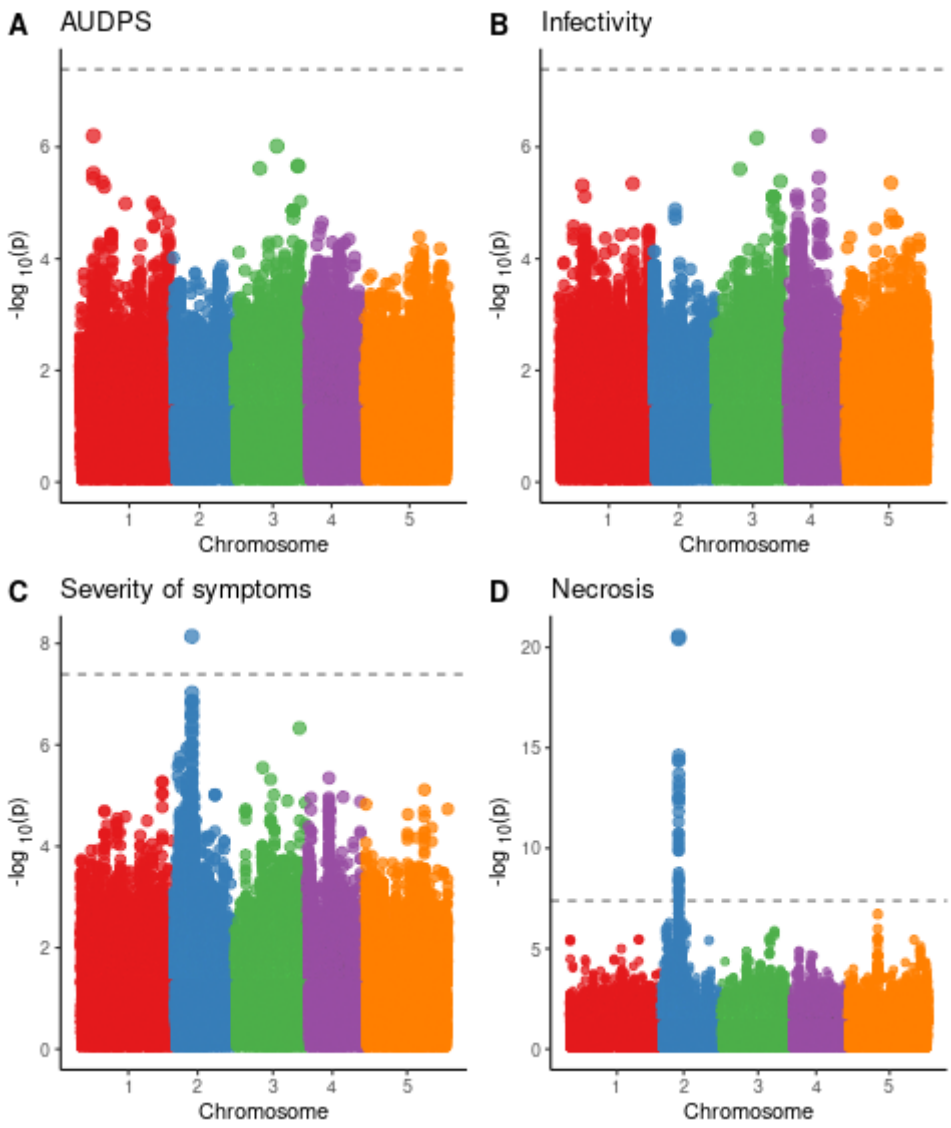


Figure S2. Manhattan plots for associations with common disease responses to both TuMV isolates.

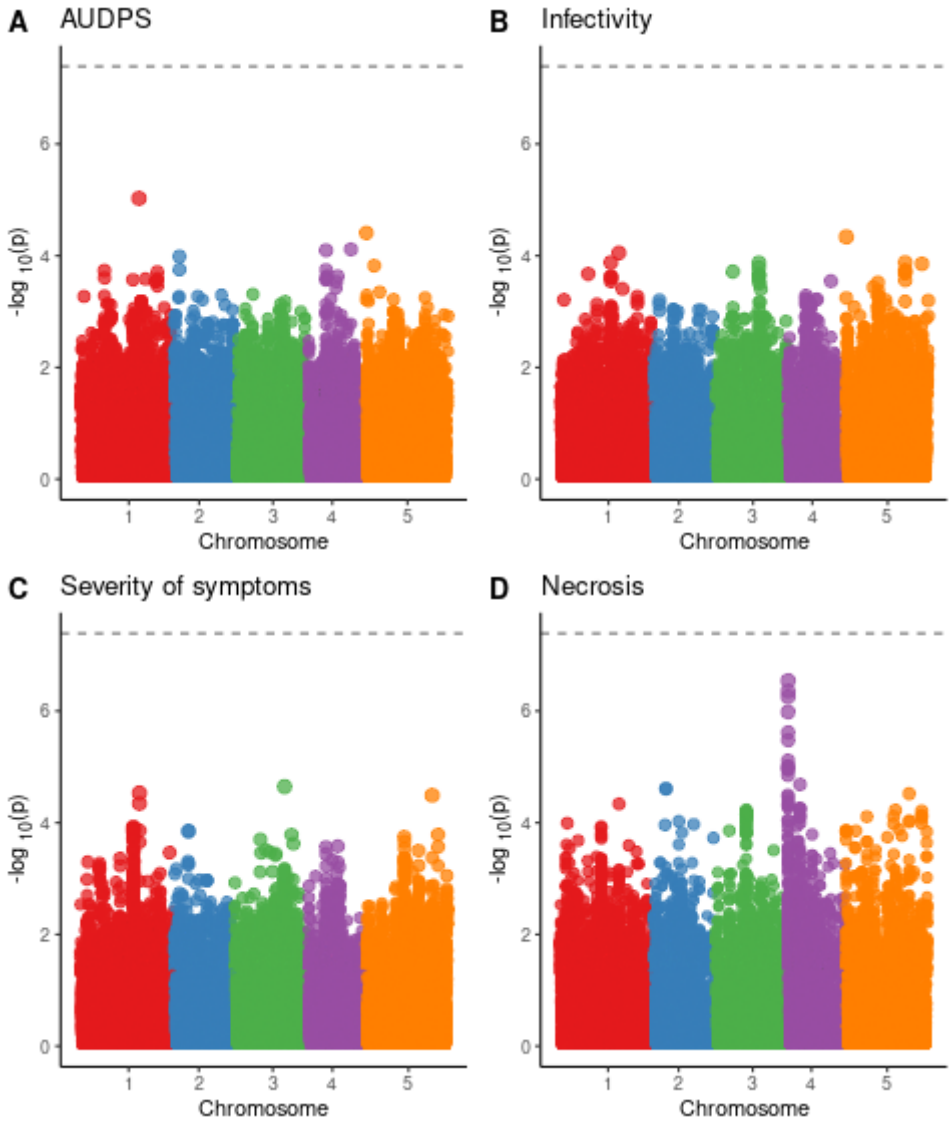


Figure S3. Manhattan plots for associations with disease responses specific to individual TuMV isolates.