1 A Multimodal Study of Wild Tomato Species Introgressions 2 Unravels the Genetic Basis of Transcriptome and Metabolome 3 Variation in Important Fruit Traits

4
5 Jędrzej Szymański^{1,2*+}, Samuel Bocobza^{1,3*}, Sayantan Panda^{1,4*}, Prashant Sonawane¹, Pablo D.
6 Cárdenas⁵, Justin Lashbrooke⁶, Avinash Kamble⁷, Nir Shachaf¹, Sagit Meir¹, Arnaud Bovy⁸, Jules
7 Beekwilder⁸, Yury Tikunov⁸, Irene Romero de la Fuente⁸, Dani Zamir⁹, Ilana Rogachev¹ and Asaph
8 Aharoni¹⁺

- 9 1. Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot, Israel
10 2. Current address: Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop
11 Plant Research, Gatersleben, Germany
12 3. Current address: Institute of Plant Sciences, Agricultural Research Organization, Institute of Plant
13 Sciences, Rishon LeTzion, Israel
14 4. Gilat Research Center, Agricultural Research Organization, Mobile Post Negev, 85280 Israel
15 5. Department of Plant and Environmental Sciences, University of Copenhagen, Denmark
16 6. Institute for Wine Biotechnology, Faculty of Agrisciences, Stellenbosch University, South Africa
17 7. Savitribai Phule Pune University, Department of Botany, Pune, India
18 8. Wageningen Plant Research, Department of Plant Breeding, Wageningen, Netherlands
19 9. Faculty of Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel

20

21 * These authors contributed equally to this study

22 + Correspondence should be addressed to A.A. (asaph.aharoni@weizmann.ac.il) and J.S.
23 (szymanski@ipk-gatersleben.de)

25 ABSTRACT

26 Wild tomato species represent a rich gene pool for numerous desirable traits lost during domestication.
27 Here, we exploited an introgression population representing wild desert-adapted species and a
28 domesticated cultivar to establish the genetic basis of gene expression and chemical variation
29 accompanying transfer of wild species-associated fruit traits. Transcriptome and metabolome analysis of
30 580 lines coupled to pathogen sensitivity assays resulted in the identification of genomic loci significantly
31 linked with levels of hundreds of transcripts and metabolites. These associations occurred in functionally
32 coherent hotspots that represent coordinated perturbation of metabolic pathways and ripening-related
33 processes. We demonstrated the efficacy of our approach by discovering yet undisclosed components of
34 the renowned *Solanum* alkaloids pathway, identifying new genes and metabolites involved in pathogen
35 defence and linking fungal resistance with changes in the fruit ripening regulatory network. Together, our
36 results outline a framework for understanding metabolism and pathogen resistance during tomato fruit
37 ripening and provide insights into key fruit quality traits.

38 INTRODUCTION

39 Domestication and breeding of tomato have mostly been driven by modification of fruit phenotypes to
40 satisfy marketing and economic desires¹. At the same time, continuous selection through breeding
41 resulted in reduced genetic diversity^{2,3} and recurrent elimination of significant fruit qualities; flavour,
42 aroma and pathogen resistance represent the most apparent lost traits⁴. Presently, researchers and breeders
43 use wild tomato relatives in search of a genetic pool required to reintroduce these traits into the
44 commercial cultivars⁵⁻⁷. Several recent Genome-Wide Association Studies (GWAS) used natural
45 populations for identification of genetic loci responsible for tomato fruit size, colour and flavour^{4,7-10}.
46 Other studies employed Quantitative Trait Loci (QTL) analysis to highlight loci associated with traits

47 such as soluble sugar content (*i.e.* Brix) and acyl-sugar accumulation^{11–13}. Following these studies, it
48 became apparent that more advanced mapping populations and deeper, multi-omics phenotyping are
49 required to identify genetic cues responsible for complex traits.

50 Recently, an advanced mapping population for tomato was developed using the wild Peruvian
51 accession of *Solanum pennellii* (LA0716 or PI246502) and the *Solanum lycopersicum* cultivar M82¹⁴.
52 This Backcrossed Inbred Line (BIL) population is complementary to the *S. pennellii* Introgression Lines¹⁵
53 (IL) population with which it shares parents, yet, it provides remarkably improved resolution of QTL
54 mapping^{14,16}. Using these introgression lines, we recently identified the enzyme responsible for steroidal
55 glycoalkaloids (SGAs) hydroxylation in tomato and potato plants¹⁷. At the same time, emerging
56 phenotyping and molecular profiling technologies enable high throughput quantification of an increasing
57 number of plant parameters and variants as shown in a GWAS mapping of fruit quality traits in tomato⁹.

58 In this study, we utilized advanced genetic resources together with multimodal molecular and
59 phenotype profiling to perform an integrative QTL analysis in tomato fruit. We focused on human
60 nutrition-associated secondary (specialized) metabolites and fruit resistance to pathogens. Altogether, 580
61 lines of combined BIL and IL mapping populations were phenotyped at the breaker and red stage of fruit
62 ripening through (i) 3' sequencing transcriptome analysis (at the breaker stage), (ii) metabolomics (at
63 both breaker and red stages) and (iii) pathogen sensitivity assays (at the red stage). By integrating genome
64 - transcript - metabolite - phenotype QTL analysis, we identified and characterized a yet undisclosed
65 enzymatic step in the predicted pathway wherein α -tomatine, the fundamental alkaloid in tomato fruit, is
66 converted to esculeosides and lycopersides during fruit development and ripening. This 'chemical shift'
67 is likely important for reducing the bitterness provided by α -tomatine and/or countering temporal needs
68 for defensive fruit metabolites. Moreover, loci and genes associated with accumulation of health-
69 promoting flavonoids in the fruit skin tissue were delineated. Finally, we analysed the molecular data

70 with respect to acquired fruit resistance to a generalist pathogen *Botrytis cinerea*, one of the most
71 significant fungal pathogens of tomato plants, and identified metabolites and genes associated with
72 increased *B. cinerea* resistance.

73 RESULTS

74 **Generating a multi-omics data resource for fruit of 580 introgression lines**

75 We first dissected skin tissue from fruit of 580 inbred lines (504 BILs and 76 ILs) at two fruit
76 developmental stages: breaker (~44 days post-anthesis; DPA) and red ripe (~48 DPA) (**Figure 1a**).
77 Samples from both stages were examined for metabolite composition in a non-targeted approach using
78 high-resolution mass spectrometry. Gene expression in the breaker stage samples was analysed using
79 TRANSeq 3'-end RNA sequencing¹⁸. In parallel, we performed pathogen susceptibility assays on the red,
80 ripe fruit by infecting the entire set of introgression lines with a generalist pathogen *B. cinerea* and
81 quantifying the progression of infection. *S. pennellii* introgressions in IL and BIL lines used in this study
82 were previously mapped using genetic markers¹⁹. Here, we also used the data obtained through TRANSeq
83 transcriptome analysis to identify genomic regions displaying reciprocal variations in the alignment of *S.*
84 *lycopersicum*- and *S. pennellii*-unique reads (see Supplemental Materials). We next used the introgression
85 mapping results to divide the tomato genome into narrow genomic regions (termed “mapping bins”); each
86 bin associated with a unique subset of lines carrying an introgression in that region. Altogether, we
87 defined 1296 mapping bins covered in median by 19 introgression lines (between 1 and 77 lines per bin;
88 **Figure 1b**). This resulted in a median resolution of nine annotated genes per bin ($\mu = 26.1$), and a median
89 of four expressed genes per bin ($\mu = 9.9$) (**Supplemental Figure S1**). As expressed, we defined here
90 11,992 genes that passed the quality thresholds of the TRANSeq data filtering (see Methods section).

91 Transcripts, metabolites and the *B. cinerea* post-infection susceptibility score were then statistically
92 associated with the occurrence of an introgression in a specific mapping bin using a linear model (with
93 significance thresholds of $\text{LOD} \geq 5$ and absolute fold change ≥ 2 ; **Figure 1c**). These QTLs were used to
94 infer a set of gene-transcript-metabolite-pathogen sensitivity interactions (**Figure 1d**).

95 **Gene regulatory interactions during fruit development uncovered through eQTL** 96 **analysis**

97 Transcriptomics of the population revealed significant associations between the presence of *S. pennellii*
98 introgressions and expression of particular genes (**Figure 2a**). Around 9% of the 11,992 quantified genes
99 showed at least one significant eQTL (1,100 transcripts with empirical false discovery rate (FDR) $\approx 10\%$,
100 logarithm of the odds (LOD) ≥ 5 and fold change ≥ 2 ; **Supplemental Table S1**). We found 582
101 transcripts possessing a significant *cis*-eQTL and 496 showing at least one significant *trans*-eQTL. This
102 relatively low number, as compared to the one reported for GWAS analysis⁹, is likely due to the fold
103 change threshold introduced in our analysis. eQTL filtering based on LOD values only, resulted in total
104 number of 5,549 eQTLs with 2,202 and 1,692 eQTLs in the *cis* and *trans* position, respectively. While the
105 number of *cis*-eQTLs per bin was proportional to the number of genes present in a given bin (Spearman
106 correlation $R_S = 0.43$, p-value $< 1e^{-10}$), the number of *trans*-eQTLs associated with a given bin varied
107 greatly and showed no such relationship ($R_S = 0.15$). The distribution of the number of *trans*-eQTLs per
108 mapping bin contrasted with the one obtained in a random permutation test, highlighting the presence of a
109 significant number of *trans*-eQTL hotspots (see Supplemental Materials).

110 *S. pennellii* introgressions lead to gene expression changes associated with the
111 ripening processes

112 The repertoire of transcripts being up-regulated in response to *S. pennellii* introgressions was significantly
113 enriched in those involved in secondary metabolism, photosynthesis, cellulose synthesis, cuticle
114 formation, fatty acid biosynthesis, cytokinin, brassinosteroid signalling and biotic stress response
115 (adjusted Fisher exact test p-value ≤ 0.05 ; **Figure 2a**; **Supplemental Table S2**). On the other hand, genes
116 that were down-regulated by the presence of introgressions were particularly related to starch synthesis,
117 ripening regulation, and phenylpropanoid metabolism. Remarkably, introgressions in the genomic loci of
118 key fruit ripening regulators such as *TOMATO AGAMOUS LIKE 1* (*TAGL1*), *FRUITFUL1* (*FUL1*) and
119 *FRUITFUL2* (*FUL2*), *AUXIN RESPONSE FACTOR 2A* (*ARF2A*) and *ETHYLENE RESPONSE FACTOR*
120 *1* (*ERF1*) were associated with the presence of clear *trans*-eQTL hotspots (**Figure 2a**). On the other hand,
121 in the case of *COLORLESS NON-RIPENING* (*CNR*), *RIPENING INHIBITOR* (*RIN*), *NON-RIPENING*
122 (*NOR*) and *NEVER RIPE* (*NR*) loci, such hotspots were not observed, indicating that only a subset of
123 known ripening regulators has been perturbed by *S. pennellii* introgressions. The eQTL hotspots in the
124 *FUL1* (**Figure 2b**) and *FUL2* loci included several previously reported target genes of these transcription
125 factors²⁰ (**Supplemental Figure S2a**). Four target genes of *FUL1*, including *CINNAMOYL-COA*
126 *REDUCTASE*, *HYDROXYCINNAMOYL TRANSFERASE* and *PECTATE LYASE* homologs were
127 differentially expressed, all of them being down-regulated presumably as a result of introgressions in the
128 *FUL1* locus, including the *FUL1* gene itself. A similar pattern was observed for the *FUL2* locus and its
129 target genes (**Supplemental Figure S2a**). Although the *RIN* introgression did not lead to the occurrence
130 of an eQTL hotspot, two of the *RIN*-activated genes, namely, *ACC SYNTHASE 4* (*ACS4*) and a putative
131 *ETHYLENE-RESPONSIVE PROTEINASE INHIBITOR 1* (*ERI*) mapped to the *RIN* locus (**Figure 2c**).
132 Interestingly, the majority of significant eQTLs of *FUL1*, *FUL2* and *RIN* direct targets was not co-

133 localized with the *FUL1*, *FUL2* and *RIN* loci. This observation illustrated that most of the observed
134 eQTLs are secondary downstream effects, resulting from the propagation of the gene perturbation
135 introduced by genetic introgressions through the gene regulatory network. Thus, only a subset of them
136 reflect the direct regulator-target interactions. The differential impact of *S. pennellii* introgression in the
137 *RIN*, *FUL1* and *FUL2* loci was further supported by correlation analysis. This showed that while the
138 expression of the *FUL1* and *FUL2* targets across all introgression lines is consistently correlated with the
139 expression of their respective regulator, this is not the case for the *RIN* target genes (**Supplemental**
140 **Figure S2b**).

141 **Metabolomics discloses enzymatic bottlenecks of secondary metabolism during fruit** 142 **ripening**

143 Fruit skin tissues of breaker and red ripening stages were subjected to metabolomics assays using liquid
144 chromatography coupled to high-resolution quadrupole-time-of-flight-mass spectrometry (LC-qTOF-
145 MS). This analysis resulted in estimated 526 and 1,019 semi-lipophilic metabolites detected in the breaker
146 and red stage samples, respectively. Seventy-nine of the breaker and 91 of the red stage metabolites were
147 annotated to various confidence levels (**Supplemental Table S3**). At the breaker stage, 229 metabolites
148 (43 annotated) exhibited at least one significant mQTL ($\text{LOD} \geq 5$, fold change ≥ 2), while at the red stage
149 the same was observed for 686 metabolites (83 annotated). As observed for eQTLs, mQTLs of multiple
150 metabolites appeared in hotspots (**Figure 3**). These were largely specific to one developmental stage and
151 representing altered accumulation of chemically-related metabolites (**Supplemental Figure S3**). These
152 observations highlighted the temporal dynamics of genetic associations changing concomitantly with the
153 metabolic transition from fruit development to the ripening stage. The observed mQTLs were often
154 associated with loci of key regulatory or enzymatic genes. For example, the *MYB12* transcription factor

155 locus localizes within the flavonoid QTL on chromosome 1^{21,22}. The *PHENYLALANINE AMMONIA*
156 *LYASE (PAL)* is positioned in the region of an mQTL hotspot that includes phenylalanine, while the
157 *CHALCONE ISOMERASE 1 (CHII)* locus was reported to affect the accumulation of flavonoids in bin-
158 424²³.

159 The mQTL in bin-424 was the largest hotspot of metabolic changes, displaying differential
160 accumulation of 249 metabolites. These included at least 22 individual flavonoids that were significantly
161 downregulated at the red fruit stage, while rutin and four kaempferol glycosides were up-regulated
162 (**Supplemental Figure S4a**). The up-regulation of rutin and kaempferol glycosides was also evident at
163 the breaker stage of the fruit development, indicating sustained effect of the locus through development
164 and ripening. Notably, *CHII* was among 42 candidate genes present in the 400 kb region corresponding
165 to bin-424 and among two candidate genes exhibiting a *cis*-eQTL in bin-424 (**Supplemental Figure**
166 **S4b**). In addition, five genes exhibited a *trans*-eQTL in bin-424 (**Supplemental Figure S4c**). None of
167 them however was directly related to the flavonoid metabolism, indicating that mQTL in bin-424 is a *cis*
168 effect. *CHII* expression is considered as a bottleneck in the biosynthesis of flavonoids during cultivated
169 tomato fruit ripening tomato, and a radical drop in its expression was observed in *S. lycopersicum* fruit
170 skin tissue between the breaker and orange ripening stage²³. Finally, a study on *Solanum chmielewskii*
171 introgression population reported similar flavonoid changes associated with the *CHII* gene²⁴, indicating
172 stability of the *CHII*-related flavonoid QTL across different mapping populations.

173 Our mQTL analysis confirmed known and led to the discovery of new gene functions; two of
174 them, detailed below. These are associated with the chemical shift in steroidal glycoalkaloid (SGA)
175 composition occurring in the course of tomato fruit ripening.

176 **Deciphering the chemical shift in defensive steroidal glycoalkaloids composition** 177 **during fruit ripening**

178 In tomato, the bitter tasting α -tomatine is the predominant SGA in leaves and green fruit. During the
179 transition from green to red fruit in tomato, the entire pool of α -tomatine is converted to esculeoside A
180 and its derivatives (**Figure 4a; Supplemental Figure S5**)^{25,26}. Metabolism of α -tomatine is required for
181 the reduction of bitterness in the late stages of development, contributing to formation of fruit adequate
182 for seed dispersal as well as human consumption. Yet, most genes and enzymes involved in the pathway
183 leading to this chemical shift remain elusive to date. Our mQTL data highlighted multiple loci associated
184 with changes in α -tomatine and its downstream derivatives (**Figure 4b**); in two cases, it pointed to the
185 function of key enzymes in the biosynthetic pathway generating SGAs during ripening. It included
186 GLYCOALKALOID METABOLISM 31 (GAME31), a 2-oxoglutarate dependent dioxygenase (2-ODD)
187 catalysing the hydroxylation of α -tomatine¹⁷, and GAME5, a uridine diphosphate (UDP)
188 glycosyltransferases (UGTs) family member mediating the formation of esculeoside A from acetox-
189 hydroxytomatine.

190 *S. pennellii* introgression in the genomic region of *GAME31* gene (bin-87, chromosome 2)
191 resulted in an increase in hydroxytomatine together with a corresponding decrease in α -tomatine levels
192 (**Figure 4b**). Among 29 genes positioned in bin-87, we detected two tandem genes encoding full-length
193 2-oxoglutarate-dependent dioxygenases (2-ODDs), *GAME31* and *GAME31-like*, and three truncated 2-
194 ODDs. This observation in fruit was in accordance with a parallel study of tomato leaf tissue¹⁷, reporting
195 GAME31 as a 2-ODD enzyme responsible for hydroxylation of α -tomatine to hydroxytomatine
196 (**Supplemental Figure S5**). Virus-induced gene silencing (VIGS) assays performed here in fruit (and not
197 in leaves as reported in the parallel study¹⁷), supported GAME31 activity (**Figure 4c, Supplemental**
198 **Figure S6a and S6b**). In order to test if the QTL originates from altered *GAME31* expression or

199 differential enzymatic activity of its *S. lycopersicum* and *S. pennellii* variants, we conducted activity
200 assays with the recombinant *S. pennellii* GAME31 enzyme. Multiple steroidal alkaloid (SA) aglycones
201 and SGA substrates were tested including ones from tomato (tomatidine, α -tomatine and
202 dehydrotomatine), potato (solanidine, α -solanine and α -chaconine) and eggplant (solasodine and α -
203 solamargine). Incubation of GAME31 with α -tomatine or dehydrotomatine as substrates (in the presence
204 of α -ketoglutarate, ascorbate and Fe^{2+}) yielded the hydroxylated products, hydroxytomatine
205 (**Supplemental Figure S6c**) and hydroxy-dehydrotomatine (**Supplemental Figure S6d**), respectively.
206 The recombinant GAME31 enzyme was also able to produce hydroxy-solasodine (**Supplemental Figure**
207 **S6e**) and hydroxy-solamargine (**Supplemental Figure S6f**) when provided with, solasodine and α -
208 solamargine as substrates, respectively. However, the GAME31 enzyme did not hydroxylate any of the
209 tested potato SA or SGA substrates. The enzyme activity assays suggested that changes in SGAs levels in
210 the QTL were likely a result of changes in *GAME31* gene expression (possibly in the promoter region);
211 yet this has to be investigated in more detail.

212 A different mQTL underlined the accumulation of esculeoside A, a major SGA in red ripe tomato
213 fruit. Thirty-six BIL and two IL lines possessing an introgression in the bin-1006 genomic region
214 (chromosome 10) displayed more than a 10-fold decrease in esculeoside A levels in the red ripe stage
215 (**Supplemental Figure S7a**). During tomato fruit maturation and ripening, acetoxy-hydroxytomatine is
216 glycosylated to form esculeoside A (**Figure 4a**). Out of total 30 genes annotated in bin-1006, we detected
217 three putative UDP-glycosyltransferases (UGTs) and among them only *GAME5* appeared highly up-
218 regulated between the breaker and orange stages (**Supplemental Figure S7b**). While VIGS-based
219 silencing of *GAME5* in red ripe fruit resulted in mild but not significant reduction in esculeoside A, levels
220 of acetoxy-hydroxytomatine (esculeoside A precursor) were significantly increased (**Figure 4d**). This
221 indicated the involvement of *GAME5* in the conversion of acetoxy-hydroxytomatine to esculeoside A. To

222 establish its glycosyltransferase activity, we expressed *GAME5* in *E. coli* and performed recombinant
223 enzyme assay with acetoxyl-hydroxytomatine and UDP-glucose. This assay resulted in the formation of
224 esculentoside A providing strong evidence for the activity of *GAME5* as acetoxyl-hydroxytomatine UDP-
225 glucosyltransferase (**Figure 4e**).

226 **Revealing genetic associations of sensitivity to *Botrytis cinerea* in ripe fruit**

227 The entire BIL and IL collections were scored for susceptibility to *B. cinerea* by measuring the diameter
228 of the lesion area following fungal inoculation of red ripe stage fruit (**Supplemental Data S2**). The
229 results were translated to *Botrytis* Resistance Score (BRS; see Methods section) describing the impact of
230 each mapping bin on *B. cinerea* infection. QTLs associated with increased resistance included genomic
231 regions of bin-542 (spanning 88 annotated genes) on chromosome 6, and a large region between bin-820
232 and bin-862 on chromosome 9 with two significance maxima at bin-827 and bin-859 (encompassing
233 further 289 annotated genes). Conversely, two QTLs, encompassing 1669 annotated genes in the region
234 of bins 18 and 19 and 58 genes in the region of bins 23 and 24, correlated with increased susceptibility to
235 *B. cinerea* infection.

236 **Integration of whole population fungal resistance data reveals new elements in fruit** 237 **pathogen association**

238 We next applied the multi-omics approach to predict the resistance of BIL and IL lines to *B. cinerea*
239 infection. We estimated the total contribution of each dataset to the prediction of *B. cinerea* resistance
240 using regularized regression analysis²⁷ (**Figure 5a**). The result indicated the transcriptome data as a major
241 predictor of *B. cinerea* resistance (~33% prediction contribution; **Figure 5b**), followed by the metabolite
242 levels in the red stage (~7%). Metabolite levels in the breaker stage as well as the introgression mapping

243 data contributed to a limited extent (~5% and 3%, respectively; described in details in Supplemental
244 Materials).

245 In the next step, we followed a classical ‘guilt by association’ principle²⁸ using weighted
246 correlation analysis and constructed a network of transcripts and metabolites that share significant QTLs
247 and exhibit correlated changes with *Botrytis* Resistance Score (BRS; **Figure 6a**). The candidate
248 transcripts and metabolites formed two tightly co-expressed network communities. The fungal resistance-
249 related community included 110 transcripts, the predominant tomato flavonoid naringenin chalcone and
250 two flavonoid glycosides 3',5'-di-C-glucosylphloretin and phloretin-trihexose, as well as six unidentified
251 metabolites. While the drop in naringenin chalcone levels was associated with increased susceptibility to
252 pathogen infection among lines carrying an introgression in chromosome 1 (bin-19 and bin-25), the
253 elevated levels of the glycosylated phytoalexin phloretin²⁹, 3',5'-di-C-glucosylphloretin and phloretin-
254 trihexose, correlated with increased resistance of those lines having introgressions in chromosome 9
255 (between bin-820 and bin-862; **Figures 6b-d**). Similarly, separate sets of transcripts showed correlation
256 with resistance changes acquired by introgressions in different regions of the genome. This indicated that
257 more than one metabolic and regulatory mechanism might be involved in the observed fungal resistance-
258 related changes. The *B. cinerea* susceptibility-related network community included 99 transcripts and
259 seven metabolites, one of them annotated as 3;4-dihydroxybenzoyl-D-xylopyranoside.

260 **Accumulation of pantothenic acid in tomato fruit skin is significant for resistance to** 261 ***B. cinerea***

262 The six metabolites found in the fungal resistance-related network community, represented potentially
263 novel elements of the *B. cinerea* resistance mechanism. Therefore, fruit extracts of the selected
264 introgression lines, that accumulated them to the highest degree, were re-analysed using LC-MS together

265 with the parent *S. pennellii* and *S. lycopersicum* samples (see Materials & Methods). One candidate
266 antifungal metabolite appeared to be pantothenic acid (PA; vitamin B₅; **Figure 6a**) and its identity was
267 confirmed by comparison to an authentic standard compound (**Supplemental Figure S8a**). Interestingly,
268 PA was present only in the *S. pennellii* and respective BIL-2387 but was absent in *S. lycopersicum* fruit
269 (**Supplemental Figure S8b**). Infection inhibition assays using varied PA concentrations validated its
270 effect on *B. cinerea* growth, showing 10% to 40% reduction in fungal growth at 10 μ M and 100 μ M PA,
271 respectively (**Supplemental Figure S8c**).

272 **Silencing of selected candidate genes results in increased susceptibility to *B. cinerea***

273 Both the resistance and susceptibility subnetworks contained a set of transcripts putatively involved in
274 fruit pathogen response and defence (**Supplemental Table S4**). To further characterize the resistance and
275 susceptibility-associated genes we analysed their tissue and developmental stage-specific expression as
276 well as expression changes in the *ripening inhibitor* (*rin*) and *non-ripening* (*nor*) ripening mutants, and
277 upon treatment with the ethylene-signalling inhibitor 1-methylcyclopropene (MCP) (**Figure 7**). The
278 results revealed the presence of four major gene expression clusters. The ‘ripening’ cluster included genes
279 up-regulated in the orange and red ripe stage of fruit development. Two green fruit-related clusters
280 exhibited predominantly green stage specific expression, differing between each other only in the mutant
281 *nor*. The fourth cluster included transcripts displaying various expression patterns, most of them highly
282 up-regulated in the red, ripe stage. The clustering analysis indicated that transcripts belonging to the
283 ‘resistance’ subnetwork were predominantly expressed in green fruit tissues (*i.e.* belonged to the green
284 fruit clusters), both in skin and flesh tissues and up-regulated in *nor* fruit (Wilcoxon rank sum test p-value
285 $< 1e-7$, mean $\log_2$ fold difference ~ 1 for all expressed transcripts). Nevertheless, most of the
286 ‘susceptibility’-related subnetwork genes were part of the ‘ripening’ cluster, being up-regulated in the

orange and red ripe stages, and predominantly down-regulated in the *rin* and *nor* mutants (Wilcoxon rank sum test p-value < 1e-7, mean log2 fold difference ~ 0.8 in respect to all expressed transcripts).

We next selected three potential resistant candidate genes including *ACC OXIDASE5* (*ACO5*), *ACCELERATED CELL DEATH2* (*ACD2*) and *4-COUMARATE:COA LIGASE-LIKE* (*4CL-Like*), for silencing in tomato. Gene silencing was based on the use of a visual marker genes encoding MAGNESIUM CHELATASE SUBUNIT H (*ChlH*) in a VIGS assay. The effectiveness of VIGS with the *ChlH* marker was established by the appearance of a bleaching phenotype in the *pTRV-ChlH*-infiltrated leaves and fruit and monitoring the relative transcript levels of *ACO5*, *ACD2* and *4CL-Like* genes (**Supplemental Figure S9a-d**). At 3 days after fungal inoculation of detached fruit, the necrotic lesions of *pTRV:ChlH:ACO5*-, *pTRV:ChlH:ACD2*- and *pTRV:ChlH:4CL-Like*- fruit increased 42.8%, 36.2% and 47.4%, in size, respectively (compared to the control, *pTRV:ChlH* fruit) (**Supplemental Figure S9e,f**). The results suggested that the three genes identified through network analysis are likely required for fungal resistance in ripe tomato fruit.

DISCUSSION

Wild species domestication and subsequent intensive breeding in tomato and additional fruit crops had a major impact on fruit ripening and the myriad of metabolic processes accompanying it. Changes in metabolism affected many archetypal fruit characteristics with economical relevance including texture, pigmentation, flavour, and resistance to pathogens. Exploring the trajectory from wild to cultivated fruit is indispensable for a comprehensive understanding of fruit metabolism and the impact of human selection on both positive and negative fruit quality traits. In this study, we completed a multi-omics investigation of variation in gene expression, metabolites and pathogen sensitivity introduced to the cultivated species' genetic background through the introgression of genomic regions from the wild

tomato. We focused on the breaker and red ripe stages representing the largest metabolic variation between fruit of wild and cultivated species^{25,30} and capturing the relationship between the accumulation of metabolite precursors in the breaker stage and their end products during fruit ripening²⁵. This relationship led to our choice to conduct transcriptomic analysis of breaker samples rather than the ones from the red stage and has to be considered when interpreting the results of correlating transcript and metabolite data from different developmental stages.

While inbred lines populations have been widely utilized for fine mapping, particularly in tomato¹⁹, our study interrogated such a population in higher resolution providing a unique dataset integrating multi-omics phenotyping data. Several previous studies indicated benefits of integration of transcript, protein and metabolite data in elucidating metabolic genes function in plants^{31–33}, however, only few utilized it in the genomic context^{9,34}. In the current study, we showed that the appearance of mQTLs is accompanied by extensive alteration in gene expression. Transcriptome data assisted in the elucidation of mQTLs by significantly reducing the number of candidates through selection of genes not only based on their genomic location but also according to their expression pattern.

The power of this approach was demonstrated by specific examples predominantly the discovery of GAME31 and GAME5 as core enzymes of the SGAs pathway in ripe tomato fruit. Several groups suggested possible pathway intermediates and enzyme activities through which the large pool of α -tomatine is converted to esculeosides and lycopersides in the course of tomato fruit maturation^{30,35,36}. Thus, GAME 5 and GAME31 discovered here add substantial knowledge regarding this postulated pathway. The chemical shift occurring in the course of ripening in tomato was likely selected for in the course of fruit evolution and domestication to circumvent high α -tomatine accumulation in ripe fruit that results in bitter flavour³⁷ as well as reducing the toxicity of fruit for frugivores consumption³⁸.

Our integrative approach also facilitated the identification of genes and metabolites putatively promoting *B. cinerea* resistance, and highlighted the role of phenylpropanoids in fruit defence. While the function of phenylpropanoids in tomato fruit response to fungal infection has been previously reported^{39,40-41,42}, the results here suggest that a very specific set of flavonoids is associated with resistance. The naringenin-7-O-glucoside, reported previously by Troncoso-Rojas et al.⁴¹ as highly induced upon powdery mildew infection, accumulated in the sensitive lines, while its aglycone naringenin was abundant in the resistant lines. Rutin on the other hand, one of the most abundant flavonoids in tomato reported earlier to accumulate in response to *Alternaria solani* infection⁴², did not correlate with resistance to *B. cinerea* infection in our study. The results of this work highlight the role of 3',5'-glucosylphloretin and pantothenic acid in fruit resistance to *B. cinerea*. Up to our knowledge, an antifungal role of both molecules was not reported in any plant species to date. Pantothenate is an important precursor of coenzyme A and holo-acyl carrier proteins (holo-ACPs)⁴³, and reducing its levels could impact a range of downstream resistance-related metabolic pathways^{44,45}. Alternatively, the direct impact of PA on *B. cinerea* growth shown here suggested that PA acts as a protective metabolite itself.

B. cinerea resistance was also associated with coordinated changes in transcripts involved in pathogen response. Genes of the resistance subnetwork were expressed mostly in green fruit and up-regulated in the *nor* mutant corroborating previous reports showing increased resistance to *B. cinerea* infection of *nor* but not the *rin* mutant⁴⁶. Many of these genes were previously reported to be associated with pathogen resistance. Functional characterization of selected candidates, ACD2, ACO5 and 4CL-Like, revealed their role in fruit resistance against *B. cinerea*. ACO5, is among the five tomato paralogs of the ACC OXIDASE (ACO) gene family. It was shown previously to be induced in tomato roots during compatible interaction with *Verticillium dahliae* and *Pseudomonas fluorescens*^{47,48}. ACO participates in the final step of ethylene production in plants and has been reported to play a crucial role in pathogen

defence and the appearance of necrotrophic symptoms during the development of a fungal infection^{49,50}. Another characterized gene, *ACD2*, encodes a Red Chlorophyll Catabolite Reductase (RCCR). In *Arabidopsis*, overexpression of *ACD2* reduces the necrotrophic disease symptoms caused by *Pseudomonas syringae*⁵¹. During *B. cinerea* infection, spreading of necrotic lesions occurs due to cell death and further accelerated in mutants from the *acd* class⁵². Previous reports showed positive correlation between the expression of the same *4CL-Like* gene and the ripe tomato fruit turgidity and skin resistance to damage⁵³. Moreover, 4CL is one of the major enzymes of the phenylpropanoid pathway that were shown to be up-regulated during the early defence response in grapes to *B. cinerea* infection⁵⁴.

The large data set generated here is a unique resource for the research community. We organized it in an easily accessible form in our online data browser (<https://szymanskilab.shinyapps.io/kilbil/>). Users can query loci, transcripts and metabolites, filter their associations according to LOD and fold change thresholds and visualize them on tomato chromosomes. While we were able to characterize in-depth only a few candidate genes and metabolites, the dataset can potentially be mined for tens if not hundreds more candidates and could be integrated with wealth of published phenomics data available for the same introgression lines. We anticipate that the genotype-phenotype associations inferred through this study will be a significant contribution to the current molecular breeding efforts to counter the recurrent elimination of key fruit quality traits such as flavour and pathogen resistance.

METHODS

Plant material and sampling

Fruit at the breaker and red ripe stages of ILs and BILs were harvested from the field at the Western Galilee Experimental Station in Akko, Israel. Plants were grown in a randomized design with 4 plants per

375 m². Fruits were brought to the lab and the skin tissue at the breaker and red, ripe stages was separated
376 from the flesh and immediately frozen in liquid nitrogen. The frozen tissue was subsequently grinded and
377 100 mg (fresh weight) was used for metabolite analysis and a similar amount for RNA extraction.

378 **RNA library preparation and sequencing**

379 RNA was extracted from skin samples using a standard tri-reagent protocol and sequencing libraries
380 prepared as described previously⁵⁵. Sequencing and data processing were performed as described by¹⁸. In
381 brief; pre-processing of the data included removal of reads with high A or T content, trimming of GGG
382 bases at the start of the read, and trimming of Illumina adapters and bases with quality <10 using
383 Cutadapt⁵⁶. The reads were mapped against a combined file that included the genomes from *S.*
384 *lycopersicum* (ITAG 3.0^{57,58,59} and *S. pennellii* v2.0^{57,58}). Mapping was performed with STAR 2.4.2a⁶⁰.
385 The alignments in the resulting SAM file were divided into genomes based on the alignment scores
386 (described in details in Supplemental Materials). Counting was performed on the TES (Transcript End
387 Site) region of each gene defined as 200 bp downstream from the STOP codon and at most 500 bases
388 upstream to the STOP codon. The reads were counted using htseq-count⁶¹, with parameters –stranded yes
389 and –mode union.

390 **Mapping of the BIL and IL populations using TRANSeq read alignment**

391 Genotyping was based on differential alignment of the reads to the *S. lycopersicum* and *S. pennellii*
392 genomes, resulting in parent-specific distribution of the read counts aligned to the 3'UTR of each gene.
393 Change Point Model (CPM) on the *S. lycopersicum*- and *S. pennellii*-specific read counts was calculated
394 using the ecp library function 'processStream'⁶² for a training set of 40 manually selected BIL lines with
395 no unannotated genomic regions (according to 10K SOL-CAP SNP-CHIP mapping of Ofner et al.¹⁴. The

396 estimated optimal parameter set reproducing DNA-sequencing results was used to detect the significant
397 change points related to the borders of tested introgressions in all remaining lines using ANOVA post-hoc
398 analysis with the $FDR \leq 0.05$ (**Supplemental Data S1**; described in detail in Supplemental Materials).

399 **TRANSeq data normalization and integration**

400 The number of mapped reads for samples used in eQTL analysis is provided in **Supplemental Table S5**.
401 The read counts were normalized to the effective library size factor using edgeR package⁶³. Differential
402 expression analysis using negative binomial distribution model and quasi-likelihood F-test with genotype
403 data as a design matrix. Negative $\log_{10}$ of the obtained Benjamini-Hochberg adjusted p-values⁶⁴ was used
404 to estimate the logarithm of the odds (LOD) values. Significant eQTLs have been identified in three
405 steps: first, for a given transcript its peaks of significant deviation from the average have been identified
406 using $LOD \geq 5$ as a threshold; second, maximums of such peaks were identified, pointing at the mapping
407 bin with the most significant deviation of the gene expression; finally, the maxima exhibiting fold-change
408 ≥ 2 were retrieved as significant eQTLs.

409 **Defining *cis*- and *trans*- expression QTLs**

410 According to the relative genomic location of the mapping bin and its significantly associated genes, we
411 defined two basic types of eQTLs: the *cis*-type, where a replacement of the parent *S. lycopersicum* gene x
412 by its syntenic ortholog leads to significantly differential expression of gene x; and the *trans*-type, where
413 a differentially expressed gene y was located outside the mapping bin of interest and may be the
414 consequence of a differentially expressed gene x inside the mapping bin. This is an analogous situation to
415 QTL analysis using GWAS populations⁶⁵. Due to the genetic structure of the BIL and IL populations, the
416 border between the *cis*- and *trans*-effects is not sharp and might be expressed as a percentage, depending

on the number of introgression lines sharing the genetic background at each genomic location. We referred to this percentage as mapping bin similarity and to the mapping bin containing a gene as its *cis*-bin. According to this definition, a *cis*-eQTL displayed a 100% mapping bin similarity and is described as a peak of differential expression in the *cis*-bin of a gene.

Metabolomics analysis of the BILs and ILs at the breaker and red stages

For LC-MS analysis, tomato fruit samples were prepared as follows: frozen powder of plant tissue (100 mg) was extracted with 80% methanol, followed by brief vortex and sonication for 20 min at room temperature. Finally, the extracts were centrifuged for 15 min at 14,000 g and filtered through 0.22 µm PTFE filters. Quality control (QC) samples were prepared separately for the red and the breaker ripening stages by pooling and homogenizing frozen powder from 60 random tomato lines. For each measurement batch the QC samples were extracted in parallel with the experiment samples. A consistent and equal (1:1) extract mixture of Arabidopsis leaves and tomato (*S. lycopersicum*; cv. M82) fruit red skin (i.e. Superpool) was prepared as described previously⁶⁶. Tomato extracts were analysed using a high-resolution UPLC/qTOF system comprised of a UPLC (Waters Acquity) which was connected online to a photodiode array (PDA) and qTOF detectors (tandem quadrupole/time-of-flight mass spectrometer, Xyvo G2-S, Waters). Separation of metabolites was performed on a 100 x 2.1-mm i.d., 1.7- µm UPLC BEH C18 column (Waters Acquity) and the mobile phase consisted of 0.1% formic acid in acetonitrile:water (5:95, v/v; phase A) and 0.1% formic acid in acetonitrile (phase B). The linear gradient was as follows: from 100 to 72% phase A over 22 min, from 72 to 60% phase A over 0.5 min, from 60 to 0% phase A over 0.5 min, then held at 100% phase B for further 3.5min; and then returned to the initial conditions (100% phase A) in 0.5 min and conditioning at 100% phase A for 1.0 min. The flow rate was 0.3 mL/min, the column temperature was kept at 35°C with an injection volume of 1 µL. A divert valve

(Rheodyne) excluded the first 1.8 and last 4.5 min from the injection. ESI was used in positive and negative ionization mode at an m/z range from 50 to 1600 Da, and PDA was set to a range from 200 to 700 nm. Masses of the eluted compounds were detected with the following settings: capillary voltage - 1 kV; cone voltage - 27 V; source temperature was set to 140 °C, desolvation to 500 °C, desolvation gas flow to 800 L/h and Argon used as the collision gas. Data acquisition was performed in the MS^E mode with energy ramp that records an exact mass precursor and fragments ion information from every detectable component in the sample. MS^E mode rapidly alternates between two functions: the first acquiring low-energy exact mass precursor ion spectra and the second acquiring elevated-energy exact mass fragment ion spectra. The collision energy was set to 4 eV for the low-energy (MS1) function and to a range of 10 to 30 eV ramp for the high-energy (MS2) function in the positive ionization mode (15 to 35 eV in the negative ionization mode). Scan time for each function was set to 0.15 s. The MS system was calibrated with sodium formate and Leu enkephalin was used as the lock mass. The MassLynx software version 4.1 (Waters) was utilized to control the instrument and calculate accurate masses and elemental composition. Sample injection was randomized, with QC samples injected in 15 samples interval within each analytical set consisting of (45-60 experiment samples per set. QC samples were later used for intra and inter-batch correction of peak areas during data analysis. In addition, a mixture of 15 standard metabolites was injected next to QCs for the quality control of analytical instrument performance. Superpool samples were injected twice in the beginning and end of each analytical set and were later used for retention time adjustment during library searches.

Data acquisition and normalization

Raw data files produced as described above were converted from the vendor's format to the netCDF open source format using the Databridge software (Waters Inc.). The netCDF files, corresponding with LC-MS

data in the positive and negative ionization modes, were subsequently processed by the open source software packages XCMS⁶⁷ and CAMERA⁶⁸ which process individual raw data signals into an aligned matrix of features (i.e. a sequence of m/z – RT tuples with the integrated peak areas across the measured samples). Due to the large number of samples analysed, the memory-intensive procedure of peak picking was performed on subsets of 100 samples each, and then merged into one XCMS object to perform the subsequent steps of peaks grouping, alignment and clustering. XCMS peak picking parameters were as follows: snthresh 5, fwhm 10, mzdifff 0.05, max 200; XCMS peak grouping and alignment parameters were as follows: bw 20, mzwid 0.05, minsamp 4, max 200, smooth "loess", span 0.75, missing 120, extra 20; CAMERA clustering and isotope/adduct annotation parameters were as follows: ppm 20, multiplier 2, quick=F, cor_eic_th = 0.85, calcCiS=T, calcIso=T. The experimental artifact of inter-batch differences (related to the large number of measured samples and thus several measurement batches), were corrected using the ComBat function from the sva package⁶⁹ separately for each ionization mode and each developmental stage. The inter-batch differences (related to long sample runs) were corrected using a 'loess' model fitted to a series of QC-pool measurements. Peak intensities were further normalized column-wise to the median of mass intensities and log normalized.

Annotation of metabolites

The measured features of m/z -RT tuples with their intensity profiles were annotated using a combination of in-house software scripts and a previously published software package⁶⁶. The reference MS libraries were, accordingly, a combination of reference LC-MS peaks related to tomato species plus the Weizmass library of plant metabolites, resulting in a series of annotated features corresponding with the accepted metabolite annotation confidence levels^{70,71}. Thus, software feature annotations corresponding with confidence levels one to three were manually validated by comparing the measured m/z -RT tuples of the

483 molecular ion and the related adduct and fragment ions to the values registered in the respective reference
484 library. Feature annotations corresponding to confidence level four were verified by elemental
485 composition using the measured masses of naturally occurring heavy isotope of the putative molecular
486 ion using the MassLynx software (Waters inc.) and by elaborate examination of coeluting peaks, to
487 exclude the option of falsely annotating the mass fragment of a larger metabolite as the molecular ion.
488 When possible, chemical class assignments were carried out manually using the annotated molecular
489 structure. See Results section and **Supplementary Table S3** for a detailed list of metabolites putatively
490 identified in this study. Identification of pantothenic acid (PA) in tomato skin extracts of *S. lycopersicum*
491 (M82), *S. pennellii* and BIL-2387 involved the following modifications of the above-mentioned
492 procedures. A divert valve excluded the first 1.1 and last 4.5 min from the injection. Analysis was done in
493 the positive ionization mode. The MS settings were modified to: capillary voltage - 3 kV; cone voltage -
494 80 V. The collision energy was set to 1 eV for the low-energy (MS1) function and to a range of 10 to 45
495 eV ramp for the high-energy (MS2) function. RT and fragmentation pattern of the putative metabolite
496 predicted to be PA was compared to the PA standard (Sigma, USA), injected at the concentration of
497 5ug/ml in the same analytical batch using the same diluent and the same LCMS parameters.

498 **Assays with the necrotrophic fungal pathogen *Botrytis cinerea***

499 Fungal infection of the introgression line fruits was carried out according to Flors et al.⁷² with minor
500 modifications. In brief, conidia of *B. cinerea* were collected from 15-day-old PDA plates with sterile
501 water containing 0.01% (v/v) Tween-20, 10 mM sucrose and 10 mM KH₂PO₄. The suspension was
502 filtered, adjusted to an appropriate concentration of 1×10⁶ spores ml⁻¹ by using a haemocytometer, and
503 was stored for 2 h in the dark with no shaking. Each red ripe tomato fruit was pricked on the surface at
504 four sites by a 1 mm syringe needle and 5µl droplet of spore suspension was applied. The droplets were

505 allowed to dry for 2 h, and later were sprinkled with sterile water and kept in a chamber with high
506 humidity at 20°C. Fungal mycelium grew concentrically at the site of spore application, causing a visible
507 necrosis 3 days after inoculation. Disease scoring was done by measuring the diameter of the necrotic
508 lesion after 5 days of spore application. Disease scoring was carried out on an average of 11 fruits per
509 introgression line with 4 lesions per fruit. Results obtained from the 20 top-scoring resistant lines were
510 validated in an independent experiment using the same procedure (**Supplemental Data S2**). *B. cinerea*
511 susceptibility assays in the fruit VIGS experiments were performed as described for introgression lines,
512 however, due to faster development of the fungal mycelium, disease scoring was done after 3 days of
513 spore application. Disease scoring was carried out by measuring the diameter of two necrotic lesions per
514 fruit using 8 fruits per line. The experiment was repeated twice independently. To examine fungal growth
515 inhibitory effect of PA, *B. cinerea* was maintained on potato dextrose agar (PDA) plate supplemented
516 with two concentrations (10 and 100 µM) of methanol diluted Pantothenic acid (PA) standard (Sigma,
517 USA). Plates supplemented with methanol only were used as controls and the experiment performed with
518 eight biological replicates. Reduction in *B. cinerea* growth (as % inhibition) was recorded after 3 days of
519 fungal inoculation.

520 **Virus induced gene silencing (VIGS) assays**

521 Tomato (*S. lycopersicum*) cv. Micro Tom plants (E8:Del/Ros MT) expressing Del/Ros1 transcription
522 factors under fruit specific promoter E8 were grown in a climate-controlled greenhouse at 24°C during
523 the day and 18°C during the night, with natural light. The *GAME31* VIGS construct was generated by
524 introducing a 298 bp fragment (primer oligonucleotides provided in **Supplemental Table S6**) to
525 pCR8/GW/TOPO (by ECORI) and further cloning to the pTRV2 binary vector using the Gateway LR
526 Clonase II enzyme mix (Invitrogen). VIGS experiment was performed on mature green stage fruits

(where purple pigments begin to accumulate) of E8:Del/Ros Micro Tom tomato plants as described previously⁷³ except final O.D. of agrobacterium was kept 0.3 in current VIGS experiment. Tomato plants infected with Agrobacterium, containing the pTRV2 vector harbouring only Del/Ros sequences were used as control. Due to the co-silencing of the desired gene (here *GAME31*) and Del/Ros transcription factors, clear visible red sectors appeared on the background of purple fruit upon maturation. Silenced red patched fruit tissues were collected 3 or 4-weeks post infection and analysed for SGAs profile by LC-MS. For VIGS assays of *GAME5* in tomato, detached tomato fruit (*S. lycopersicum* L. cv. Maxifort) were used for this experiment. The silencing of the tomato *GAME5* gene was performed as described in Romero et al.⁷⁴. Oligonucleotides used for silencing the *GAME5* next to the *PDS* gene (reporter gene) are provided in **Supplemental Table S6**. In order to quantify endogenous phytoene desaturase (PDS) transcript abundances in pTRV1/2-PDS and pTRV1/2-PDS-*GAME5* tomato infiltrated fruits and pTRV1/2 control fruits, as well as for quantification of the *GAME5* silencing area, qRT-PCR analysis was used (oligonucleotides provided in **Supplemental Table S6**). The procedure for subsequent RNA extraction and qRT-PCR data calculation was performed as described in Romero et al.⁷⁴. VIGS assays were also used for investigating candidate genes associated with the *Botrytis* resistance network. For most of the reports on TRV-derived VIGS system to study fruit development in tomato, transgenic over expression of Del/Ros1 has been used as visual marker that leads to development of purple fruits due to anthocyanin accumulation^{73,75,76}. However, induced anthocyanin level alter the spreading of the ROS burst during fungal infection that reduces the susceptibility against *B. cinerea*⁷⁷. To overcome this problem, another visual marker *Magnesium Chelatase Subunit H (ChlH)* has been chosen⁷⁸. First, ~300 bp of *ChlH* was inserted into an EcoRI digested pTRV2 vector to serve as a silencing marker. For the silencing of candidate genes in tomato, all gene sequences (~200~400 bp) were amplified from tomato fruit cDNA template (for sequences of primers used to amplify DNA fragments see **Supplemental Table**

550 **S6).** Secondly, purified amplicons of candidate genes were inserted into modified pTRV2 vector digested
551 with BamHI. The use of this pTRV2 plasmid results in co-silencing of two genes –*ChlH* (which results in
552 chlorotic phenotype and serves as a silencing marker) and the candidate gene. This approach allowed
553 visual identification of tissue in which the candidate genes were silenced. Fruit with severe phenotypes
554 have been marked in green stage (unripe) and were used in red stage (ripe) for fungal infection assay.

555 **GAME5 and GAME31 enzyme assays**

556 For the GAME5 enzyme activity assay, its open reading frame was amplified from tomato cDNA
557 (oligonucleotides provided in **Supplemental Table 5**). The amplification product was restriction digested
558 with *Bgl*III and *Pst*I and cloned into pACYCDUET-1. The resulting plasmid, pACYC-GAME5, was
559 transformed to *Escherichia coli* BL21 DE3 (Novagen). An empty pACYCDUET-1 vector, was also
560 transformed to this strain and served as negative control in enzyme assays. Fresh overnight cultures were
561 diluted 1:100 in 25 mL yeast-tryptone medium with 50 µg/mL chloramphenicol and incubated at 37°C
562 and 250 rpm until an OD600 of 0.5 was reached. Subsequently, isopropyl-b-D-thio-galactoside (IPTG)
563 was added to a concentration of 1 mM, and the incubation continued overnight at 18°C and 250 rpm. The
564 next day, cells were harvested by centrifugation and pellets resuspended in 2 mL of 50 mM Tris HCl, pH
565 7.5, with 8 mM β-mercaptoethanol. After breaking the cells by sonication, insoluble material was
566 removed by centrifugation, and the soluble fraction was treated with Ni-nitrilotriacetic acid spin columns
567 (Qiagen) according to the manufacturer's instructions for purification of the recombinant protein. The
568 solution containing purified protein (200 µL) was dialyzed in the presence of 2 liters of 50 mM Tris HCl,
569 pH 7.5, with 8 mM β-mercaptoethanol at 4 °C. The substrate acetoxyl-hydroxytomatine was dissolved at 1
570 mM in DMSO. Enzyme assays were performed in 50 mM Tris-HCl, pH = 7.5, with 8 mM β-
571 mercaptoethanol, using 5 mg/mL enzyme, 8 mM UDP-Glucoside (Sigma, St Louis), and 0.1 mM

572 substrate in a volume of 200 μ L. After 2 h of incubation under agitation at 30°C, reactions were stopped
573 by addition of 500 μ L methanol and 0.1% formic acid. Samples were prepared by sonication for 10min.
574 Then, the extracts were centrifuged for 5 min at 20,000g. Enzyme assays were analysed by LC-MS. For
575 chromatographic separation, a Luna C18(2) pre-column (2.0 $\times$ 4 mm) and an analytical column (2.0 $\times$
576 150 mm, 100 Å, particle size 3 μ m) from Phenomenex (Torrance, CA, USA) were used at 40°C. Five μ L
577 of each sample were injected into the system for LC-MS analysis using formic acid:water (1:1000, v/v;
578 eluent A) and formic acid:acetonitrile (1:1000, v/v; eluent B) as elution solvents. Flow was set at 0.19
579 mL/min with the gradient from 95% eluent A and 5% eluent B to 30% Eluent A and 70% eluent B across
580 a period of 45 min. ESI-MS analysis was performed using a LTQOrbitrap XL hybrid MS system (Waters)
581 operating in positive electrospray ionization mode heated at 300 °C with a source voltage of 4.5 kV for
582 full-scan LC–MS in the m/z range 100–1500. Cloning, *E. coli* expression and purification of GAME31
583 from *S. pennellii* was performed similarly as described by Cárdenas et al.¹⁷. In vitro recombinant
584 GAME31 enzyme (5 μ g) assay with the steroidal alkaloid aglycones (i.e., tomatidine, dehydrotomatidine,
585 solanidine and solasodine) and steroidal glycoalkaloids (i.e., α -tomatine, dehydrotomatine, α -solanine, α -
586 chaconine and α -solamargine) substrates along with control reactions were performed under standard
587 assay conditions, as described previously¹⁷. GAME31 enzyme assay products (hydroxylated SGAs) were
588 putatively identified by comparing their retention times, elemental composition and fragmentation pattern
589 with those described earlier¹⁷.

590 **Prediction of *B. cinerea* resistance from multi-omic data**

591 Predictive features for each dataset, including introgression design matrix, transcript, metabolite breaker
592 and metabolite red datasets, were identified by LASSO-regularization of the fitted linear models using the
593 ‘glmnet’ package in R⁷⁹; R Team, 2015). The lambda penalty parameters were estimated using 10-fold

cross-validation within the training set, and the ' $\lambda_{\min} + 1\text{SE}$ '-rule was applied to choose the model parameters for each dataset separately. The prediction power of each dataset was estimated in Monte Carlo cross-validation test with 100 samples with 90% and 10% of samples taken as a training and test set respectively. All resulting features with non-zero regression coefficients were included in a combined dataset, for which new penalty parameters have been estimated. Contribution of each dataset to the *B. cinerea* resistance prediction has been estimated by comparison of the R^2 of the combined dataset with the R^2 of the dataset combined of all datasets except the one in question. The model's performance has been again estimated using 100 Monte Carlo cross-validation samples as described above.

***B. cinerea* resistance network reconstruction**

In the first step, mapping bins associated with significantly increased susceptibility and resistance to the pathogen infection were identified using threshold of absolute BRS > 6. Secondly, sets of transcripts and metabolites exhibiting a significant QTL in any of these bins were identified using LOD > 5 and fold-change > 2. In the third step, *Botrytis* fold-change vector (vector of *Botrytis*-resistance scores averaged for each mapping bin) and the fold-change vectors of selected transcripts and metabolites were used in weighted correlation analysis. For each correlated pair xy , vectors of weights $w_x = 1 - p_x$ and $w_y = 1 - p_y$ have been used to calculate weighted correlation coefficients R_{xy} and R_{yx} , where p_x and p_y were vectors of p -values of the association of x and y level with each mapping bin. Finally, mutual criterion of $R_{xy}^2 > 0.5 \cap R_{yx}^2 > 0.5$ were used to identify significant interactions. The resulting network was divided into two clusters based on the weighted correlation sign with the *Botrytis* fold-change vector. The clusters represented transcripts and metabolites associated with resistance and susceptibility to *B. cinerea* infection, respectively.

615 **Tissue-specific gene expression analysis**

616 *S. lycopersicum* plants (cv Micro Tom) were grown and sampled for transcriptomics analysis as described
617 by Cárdenas et al.⁸⁰. 1-methylcyclopropene (MCP) treatment has been performed as described by Itkin et
618 al.²⁵. Briefly, Illumina RNAseq transcriptome data were obtained for from 13 tomato samples, including
619 skin and flesh samples in five fruit developmental stages (immature green, mature green, breaker, orange
620 and red ripe; 10, 35, 38, 41, and 44 days after anthesis, respectively), whole fruit samples for *rin* and *nor*
621 mutants, as well as for the MCP ethylene receptor inhibitor treatment. Clustering of the expression data
622 was performed using hierarchical clustering and the number of clusters was estimated using GAP
623 statistics ('cluster' R package⁸¹). Complete linkage hierarchical clustering and row-wise Z-normalization
624 were used to construct the heat map. All calculations, data processing and visualization were performed
625 using R and Bioconductor⁸².

626 **Online data browser**

627 For the purpose of this study we provide an online browser that enables visualization and downloading
628 results of our experiments (<https://szymanski-lab.shinyapps.io/kilbil/> temporarily closed access available
629 for reviewers and editors; login: szymanski@ipk-gatersleben.de, password: revieweraccess2019). The
630 browser allows searching genes and genomic regions of interest and obtaining lists of transcripts and
631 metabolites exhibiting a significant QTL in their genomic vicinity. The generated tables and plots might
632 be downloaded and directly used for presentation. Additional metadata, including functional annotation of
633 transcripts and metabolite annotation confidence, is also included.

634

635 ACKNOWLEDGMENTS

636 This project has received funding from the European Research Council (ERC) grant agreement No.
637 204575-SAMIT and the EU Framework Program FP7/2007–2013 grant agreement No. 613692-TriForC
638 and the Israel Science Foundation (ISF Grant No. 1805/15). We thank the Adelis Foundation, Leona M.
639 and Harry B. Helmsley Charitable Trust, Jeanne and Joseph Nissim Foundation for Life Sciences, Tom
640 and Sondra Rykoff Family Foundation Research and the Raymond Burton Plant Genome Research Fund
641 for supporting the AA lab activity. Avinash Kamble is thankful for a short-term EMBO fellowship
642 (EMBO-ASTF-146-2014). We are thankful to the AA laboratory members that assisted in harvesting,
643 handling, peeling, freezing and grinding the thousands of fruit tissue samples used in this study. We also
644 thank Yoko Iijima and Koh Aoki for providing purified acetoxyl-hydroxytomatine. AA is the incumbent
645 of the Peter J. Cohn Professorial Chair.

646 AUTHOR CONTRIBUTIONS

647 JS performed the integrative analysis and wrote the manuscript. SB conducted the RNAseq library
648 preparation, performed metabolite extraction and contributed in writing the manuscript. PS, PC, SP and
649 AK performed the *B. cinerea* resistance experiment. SP contributed in RNAseq library preparation and
650 functional characterization of candidate genes for fungal resistance. PS and PC characterized the
651 *GAME31* gene function. AB, JB, YT and IRF characterized the *GAME5* gene function. NS, SM and IR
652 performed the metabolomics analysis and metabolite annotation. DZ provided BIL lines and supervised
653 the field experiment. AA planned and supervised the study. SB, PS, SP, JL and NS supported the study
654 with their expertise and contributed in writing selected sections of the manuscript.

655 COMPETING INTERESTS STATEMENT

656 The authors declare no competing financial interests.

657 FIGURE LEGENDS

658 **Figure 1**

659 **A multimodal study of wild tomato species introgressions.** (a) Schematic representation of the tissue
660 sampling time points, and datasets generated and collected in the study. (b) Graphical representation of
661 transcriptome-based mapping results of 504 BILs and 76 ILs sorted according to the genomic location of
662 the first detected introgression (lower panel) together with a respective bar plot describing the number of
663 lines having an *S. pennellii* introgression at each genomic region (upper panel). (c) Toy-example
664 representing key concepts of the analysis. A single chromosome for six theoretical BIL lines is depicted.
665 Localization of the *S. pennellii* introgressions in these six lines (red), divides the chromosome into 21
666 mapping bins. For each bin, values of the measured variables (transcripts, metabolites and phenotypes)
667 are compared between the BILs carrying an introgression in the genomic range of the bin and those
668 exhibiting *S. lycopersicum* genotype in that region. Here, bin-11 is shown as an example. The differential
669 variables point at QTLs in the genomic region of the bin. (d) Scheme of possible relationships between
670 tomato genes and all measured variables.

671 **Figure 2**

672 **Mapping of expression QTLs across the entire introgression line population.** (a) Distribution of the
673 number of differentially expressed genes (*trans*-eQTL only) per mapping bin according to the $\text{LOD} > 5$

and fold-change > 2 thresholds. The up- and down-regulated genes are displayed symmetrically on the top and the bottom panel, respectively. Loci of the major known ripening regulators mentioned in the main text are labelled by their gene name abbreviation. eQTL hotspots enriched in particular functional categories (MapMan bins), according to the adjusted Fisher exact test p-value ≤ 0.05 , are highlighted. (b, c) Genomic localization and sign of eQTLs of all targets of FUL1 and RIN identified in ChIPSeq and mutant analysis by Fujisawa et al.²⁰. CNR – Colourless Non-Ripening; FUL1, Fruitful-like 1; FUL2, Fruitful-like 2; ARF2A, Auxin Response Factor 2A; RIN, Ripening Inhibitor; ERF1, Ethylene Response Factor 1; TAGL1, Tomato Agamous-like 1; NR, Ethylene Receptor 3 (Never Ripe); NOR, Non-Ripening.

Figure 3

Mapping of metabolic QTLs across the entire introgression line population. Distribution of the number of differentially accumulated metabolites per mapping bin according to the genomic location of the bins. For mQTL hotspots, the annotated metabolites are highlighted together with their developmental stage specificity. The accumulated and depleted metabolites are displayed symmetrically on the top and the bottom panel, respectively. Loci of the genes pointed out in the main text are labelled by their gene name abbreviation. MYB12, SIMYB12 transcription factor; FPS, farnesyl pyrophosphate synthase; CCMT, caffeoyl CoA O-methyltransferase; GAME31 (GlycoAlkaloid MEtabolism 31), 2-oxoglutarate dependent dioxygenase; CHI1, naringenin chalcone isomerase 1; ACAT, 3-ketoacyl-CoA thiolase 1; ETR6, ethylene receptor 6; GAME5 (GlycoAlkaloid MEtabolism 5), UDP glycosyltransferase; PAL, phenylalanine ammonia lyase.

693 **Figure 4**

694 **Discovery of two enzymes mediating the ‘chemical shift’ towards ripening associated steroidal**
695 **glycoalkaloids in the course of tomato fruit ripening.** (a) The proposed metabolic steps in the steroidal
696 glycoalkaloids (SGAs) biosynthetic pathway catalysed by GAME31 (green colour) and GAME5 (red
697 colour). (b) Logarithm of Odds (LOD) scores corresponding to the accumulation of the metabolites and
698 for gene expression of *GAME5* and *GAME31*. Up-regulation and down-regulation represented as positive
699 and negative values, respectively. (c) Levels of SGAs in *GAME31*-silenced red, ripe tomato fruit [Virus-
700 Induced Gene Silencing (VIGS)] as compared to control (plants infected with the pTRV2 vector
701 harbouring only *Del/Ros* sequences). Values represent mean $\pm$ standard error (n=7 for silenced and n=3
702 for control plants). (d) Levels of the SGAs acetoxyl-hydroxytomatine and esculeoside A in red, ripe fruit
703 of *GAME5*-silenced tomato plants compared to control [phytoene desaturase (*PDS*)-silenced plants].
704 Values represent mean $\pm$ standard error (n=5). (e) Enzyme assay of the purified recombinant tomato
705 GAME5 produced in *E. coli* BL21 (DE3) cells. Extracted ion chromatograms of empty pACYCDUET-1
706 vector control assay reaction with acetoxyl-hydroxytomatine (m/z 1108.55; upper panel), tomato GAME5
707 enzymatic reaction using acetoxyl-hydroxytomatine as substrate (middle panel) and esculeoside A
708 standard (m/z 1270.61; lower panel). Enzyme assay reactions were analysed by LC-MS. *E. coli*:
709 *Escherichia coli*, m/z : mass to charge. Asterisks (panel c and d) indicate significant changes from control
710 plants as calculated by a Student’s t-test (*P-value < 0.05; **P-value < 0.01; ***P-value < 0.001).

711 **Figure 5**

712 **Prediction of *B. cinerea* resistance from multi-omic data.** (a) *B. cinerea* resistance prediction
713 expressed as a boxplot of R^2 values obtained for a LASSO-regularized regression model. (b)

Contributions of each level of measured molecular features to the *B. cinerea* resistance expressed as percentage of the dataset-unique variance predictive towards *B. cinerea* resistance.

Figure 6

Network analysis reveals *B. cinerea* resistance-associated genes and metabolites. (a) Overview at the integrated transcript-metabolite network with transcripts and metabolites represented as nodes. Pantothenic acid has been additionally highlighted, as it was annotated and tested in a follow-up experiment as described in the main text. (b) Normalized average fold-change of the *B. cinerea* resistance across all mapping bins. (c) Normalized average fold-change of each transcript and metabolite of the resistance subnetwork across all mapping bins. (d) Normalized average fold-change of each transcript and metabolite of the susceptibility subnetwork across all mapping bins. Lines in panels b-c are colour-coded according to the legend in panel a.

Figure 7

Gene expression patterns of *B. cinerea* resistance-associated genes. Expression patterns of genes of the resistance and susceptibility subnetworks in fruit flesh and skin during fruit development and ripening and in the breaker stage of *rin* and *nor* ripening mutants' fruit and upon 1-methylcyclopropene (MCP; ethylene receptor inhibitor) treatment. The developmental series include: Immature Green (IG), Mature Green (MG), Breaker (Br), Orange (Or), and Red ripe (R) stage of fruit development. Four clusters, representing major expression patterns were highlighted. The labelled transcripts have described function in pathogen response and defence and some of them are discussed in the main text. A complete list of genes including Solyc IDs appears in **Supplemental Table 4**.

735 SUPPLEMENTAL MATERIALS

736 Supplemental Figure S1

737 **Number of annotated genes and expressed transcripts per mapping bin.** (a) Histogram of annotated
738 gene count per mapping bin. (b) Histogram of annotated gene count per mapping bin, for which expressed
739 transcripts have been quantified in transcriptome analysis.

740 Supplemental Figure S2

741 **Correlation between the known *FUL2* regulatory loci and its downstream target genes.** (a) eQTLs of
742 *FUL2* target genes identified by ChIP-seq and mutant analysis²⁰. The boxplot compares *FUL2* expression
743 in BILs carrying an introgression in bin 223 with its expression in background lines (Sly) and across all
744 BIL lines (BILs). (b) Distribution of Pearson correlation coefficients between *RIN*, *FUL1* and *FUL2*
745 expression and their respective activated (red line) and repressed (blue line) direct targets identified in the
746 ChIP-seq analysis and mutant studies. The grey area represents the distribution of Pearson correlation
747 coefficients between *RIN*, *FUL1* and *FUL2* and all measured transcripts.

748 Supplemental Figure S3

749 **Distribution of metabolic QTLs across the genome at the level of chemical classes.** (a) Distribution of
750 the number of mQTLs of each chemical class per mapping bin in the breaker stage. (b) Distribution of the
751 number of mQTLs of each chemical class per mapping bin in the red stage. The colour legend also
752 provides a summary of the mQTL count per metabolite class.

753 **Supplemental Figure S4**

754 **The metabolic QTL at bin-424 represents a major hub for flavonoids.** (a) Metabolic changes
755 resulting from *S. pennellii* introgression in bin-424. (b) Expression of genes located in the largest mQTL
756 hotspot, bin-424. (c) Gene expression changes for genes located outside bin-424 (i.e. *trans*-eQTL) in
757 response to *S. pennellii* introgression in bin-424.

758 **Supplemental Figure S5**

759 **A detailed scheme of the tomato steroidal glycoalkaloid biosynthetic pathway in developing tomato**
760 **fruit.** The characterized enzymatic steps are noted with the respective protein names.

761 **Supplemental Figure S6**

762 **Characterization of the GAME31 2-oxoglutarate-dependent dioxygenase.** The GAME31 gene was
763 silenced in E8:Del/Ros *S. lycopersicum* by the use of VIGS. (a) Relative expression level of GAME31
764 gene in VIGS-silenced red fruit tissue as compared to control (plants infected with the pTRV2 vector
765 harboring Del/Ros sequences). (b) Levels of α -tomatine, dehydrotomatine, acetoxyl-hydroxytomatine and
766 esculeoside A in red ripe fruit of GAME31-silenced tomato plants compared to control plants. Values
767 (panel a and b) represent mean $\pm$ standard error (n = 7 for silenced and n = 3 for control plants). Asterisks
768 indicate statistically significant differences as compared to respective controls determined by two-tailed t-
769 test (* p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001). (c-f) Hydroxylation of α -tomatine (c) and
770 dehydrotomatine (d) by recombinant *S. pennellii* GAME31 produced in *E. coli* cells (marked in red).
771 Recombinant *S. pennellii* GAME31 enzyme also converts solasodine (e) and α -solamargine (f) to
772 hydroxy-solasodine and hydroxy-solamargine, respectively (marked in red). Extracted ion
773 chromatograms are presented. The control reaction (shown in black) was performed with the respective

774 substrates using extracts from *E. coli* cells transformed with an empty pET28 vector. For each substrate
775 and hydroxylated product, m/z is shown. Enzyme assay reactions were analysed by LC-MS. m/z ; mass to
776 charge. Green asterisks in panel c and d indicate additional hydroxytomatine or hydroxy-dehydrotomatine
777 isomers formed in respective GAME31 assay reaction.

778 **Supplemental Figure S7**

779 **Metabolic and gene expression changes associated with mapping bin-1006.** (a) Boxplots representing
780 accumulation of putative metabolites significantly associated with the genomic region of bin-1006. The
781 coloured boxes represent metabolite accumulation in lines carrying an introgression in the genomic
782 region of bin-1006. A single annotated metabolite, esculeoside A, is highlighted in bold font. Grey boxes
783 represent accumulation of respective metabolites in all the remaining BILs. (b) Expression of genes
784 located in bin-1006, coloured accordingly. Three glycosyltransferases, including a truncated gene next to
785 *GAME5* are highlighted with bold text.

786 **Supplemental Figure S8**

787 **Identification of pantothenic acid as a potential antifungal metabolite in tomato fruit.** (a) Extracted-
788 ion chromatograms (XIC) from LCMS analysis at $m/z = 220.11$ Da; dashed line at RT = 2.5 min indicates
789 the position of Pantothenic Acid (PA); Y-axis are linked. (b) Comparison between fragments of the
790 compound eluted at RT = 2.5 min obtained from chromatograms of BIL-2387 and the PA standard. (c)
791 Inhibitory effect of PA (10 μ M and 100 μ M) on mycelium growth of *B. cinerea* after 3 days of
792 inoculation on Potato Dextrose Agar (PDA) plates.

793 **Supplemental Figure S9**

794 ***ACO5*, *ACD2* and *4CL-Like* genes are required for resistance to *B. cinerea*.** (a) VIGS-based silencing
795 of *ChlH* leads to development of a yellowish bleached phenotype, visible in green tissues (leaf and green
796 fruit). (b) Quantitative real time PCR of Mature Green (MG) and Red Ripe (RR) tomato fruit showing
797 *ChlH* transcript level. (c) Quantitative real time PCR of *pTRV:EV* and *pTRV:ChlH* infiltrated MG fruit
798 showing *ChlH* transcript level. (d) Quantitative real time PCR of *pTRV:ChlH* and *pTRV:ACO5* or
799 *pTRV:ACD2* or *pTRV:4CLL* infiltrated RR fruit showing transcript abundance of *ACO5*, *ACD2* and *4CL*-
800 *Like* genes, respectively. (e) *pTRV:ACO5* or *pTRV:ACD2* or *pTRV:4CL-Like* fruit show increased
801 susceptibility to *B. cinerea* (infected area marked in circle). (f) Necrotic lesion size in *B. cinerea*
802 inoculated fruit at 3 dpi (experiment repeated twice independently with >10 biological replicates). The
803 data in bars represents mean $\pm$ s.d. of at least three biological replicates, and data points denote individual
804 biological replicates. Asterisks indicate statistically significant differences as compared to respective
805 controls determined by two-tailed t-test (* p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001).

806 **Supplemental Figure S10**

807 **Transcriptome read counts obtained for three exemplary BIL lines in order of genomic location.** *S.*
808 *lycopersicum*-unique reads (*lyc*-unique; orange) are compared with the reads indistinguishable between *S.*
809 *lycopersicum* and *S. pennellii* (shared; blue). The resulting introgression mapping for the BIL is compared
810 with the one achieved by previously published Illumina 10k SNP-Chip genotyping¹⁴. The colour code
811 represents *S. lycopersicum* background genome (grey), *S. pennellii* introgression (red), heterozygous
812 region (orchid), and unknown (green). The displayed examples represent three cases: (a) improvement of
813 annotation confidence, (b) detection of mislabelled or wrongly annotated lines and (c) improvement of
814 annotation precision.

815 **Supplemental Figure S11**

816 **Empirical False Discovery Rate (FDR) estimation.** An empirical FDR estimate for different values of
817 LOD threshold with constant fold-change threshold ≥ 2 estimated in 1000 random permutation test.

818 **Supplemental Figure S12**

819 **Relationship between FDR and BILs coverage.** Empirical FDR estimate for LOD ≥ 5 and fold-change
820 threshold ≥ 2 for each mapping bin (estimated in 1000 random permutation test) with respect to BILs
821 coverage of that mapping bin.

822 **Supplemental Figure S13**

823 **Distance between the expected and observed eQTL.** Simulated gene expression data with the variance
824 structure and effect sizes equal to those estimated for the real experimental data was used to evaluate the
825 efficacy of the method to identify the genomic location of the eQTLs. Presented histogram reflects
826 distribution of observed distances in mapping bins between expected eQTLs (where the source of
827 variation was simulated) and observed eQTLs (recovered as for the real experimental data using LOD ≥ 5
828 and fold-change threshold ≥ 2 and the peak of significance method).

829 **Supplemental Figure S14**

830 **Distance in mapping bins between loci of genes and their closest eQTL observed in the data.** For
831 each gene exhibiting at least one significant eQTL on the same chromosome as its locus (LOD ≥ 5 and
832 fold-change threshold ≥ 2), the distance in mapping bins between the gene loci and its closest eQTL has
833 been recorded. The histogram shows the distribution of observed distances, with the distance of 0
834 signifying a cis-eQTL.

835 **Supplemental Figure S15**

836 **Analysis of the observed eQTL distribution.** Here we checked how many mapping bins (y axis) are
837 significantly associated with a specific number of eQTLs (x axis). Frequencies observed in our data (red
838 line) are compared with the frequencies observed in a random permutation test (box plot from results of
839 1000 random permutations). The difference in the observed distribution indicates biological origin of the
840 eQTL hotspot pattern described in the manuscript. Panels (a) and (b) differ only in the y-axis scale range
841 to highlight both the high and the low frequency cases.

842 **Supplemental Figure S16**

843 **Functional coherence of the genes sets associated with each mapping bin.** The positive eQTLs,
844 representing up-regulated genes (a) were analysed separately from the negative eQTLs, representing the
845 down-regulated genes (b). The functional coherence of each co-localized eQTL set has been calculated
846 (y-axis). Observed median and 95% confidence intervals are denoted by the red solid and dashed lines,
847 respectively. (c) Genomic location of the eQTL gene sets exhibiting a significant functional coherence
848 with $p\text{-value} \leq 0.05$.

849 **Supplemental Figure S17**

850 **Prediction of metabolite levels.** The heatmap presents R^2 values of the fitted LASSO models for the 56
851 metabolites with masses annotated or putatively annotated in the breaker and red stage samples (see
852 Supplemental Materials). Columns represent the following: prediction of metabolite levels in the breaker
853 stage from genomic location of *S. pennellii* introgressions (Genome – Metabolites Breaker); prediction of
854 metabolite levels in the red stage from genomic location of *S. pennellii* introgressions (Genome –
855 Metabolites Red); prediction of metabolite levels in the breaker stage from transcript levels in the breaker

stage (Transcriptome – Metabolites Breaker); prediction of metabolite levels in the red stage from transcript levels in the breaker stage (Transcriptome – Metabolites Red); prediction of metabolite levels in the red stage from metabolite levels in the breaker stage (Metabolites Breaker – Metabolites Red).

Supplemental Table S1

All significant eQTLs with the significance threshold of $LOD \geq 5$ and fold change ≥ 2 . The gene annotation is based on ITAG 3.0. Cis/trans value reflects the proximity of the eQTL to the gene locus of the eQTL transcript. Value of 1 indicates that all BILs pointing at a given eQTL carry also a *S. pennellii* introgression in the locus of that eQTL gene, and thus it's a cis-eQTL. Conversely, value of 0 indicates that no BILs pointing at a given eQTL have a *S. pennellii* introgression in the locus of that eQTL gene, and thus it's a trans-eQTL. Intermediate values indicate the ratio of BILs having a *S. pennellii* introgression in the locus of an eQTL gene in respect to all BILs pointing at a given eQTL.

Supplemental Table S2

Functional enrichment of MapMan bin categories in eQTL hotspots. All transcripts (Sheet 1), functional enrichment among all up and down-regulated genes observed in the analysed introgression lines. Bin wise up-regulated (Sheet 2), functional enrichment for genes up-regulated in response to *S. pennellii* introgressions in each consecutive mapping bin. Bin wise down-regulated (Sheet 3), functional enrichment for genes down-regulated in response to *S. pennellii* introgressions in each consecutive mapping bin.

875 **Supplemental Table S3**

876 **Metabolite annotation and confidence.** Annotation (Sheet 1); metabolites annotated with high
877 confidence (validated by manual inspection of the MS spectra) together with their retention time, m/z
878 values of the mass used for identification, ionization mode and information about library hit. Provided
879 citations refer to the bibliography in the main text. mQTL count and LASSO regression (Sheet 2), count
880 of mQTLs in breaker and red stage for annotated metabolites with detectable masses in both
881 developmental stages. R^2 of the metabolite variance predicted by the LASSO model from each respective
882 dataset is also visualized in Supplemental Figure S16.

883 **Supplemental Table S4**

884 **Expression data for transcript of the *B. cinerea* resistance and susceptibility subnetworks.** Provided
885 gene annotation is based on ITAG 3.0. Log-normalized and scaled gene expression data in five fruit
886 developmental stages (Immature Green, Mature Green, Breaker, Orange and Red Ripe; 10, 35, 38, 41,
887 and 44 days after anthesis, respectively) in fruit flesh and skin as described by Cárdenas et al.¹⁷. The *rin*
888 and *nor* mutant data as well as the MCP treatment are fold-changes in respect to control plants in the red
889 ripe stage as described by Itkin et al.²⁵.

890 **Supplemental Table S5**

891 **Statistics of the TRANSeq data.** The total number of processed reads for each sample followed by the
892 read number after trimming and the number of uniquely mapped reads.

893 **Supplemental Table S6**

894 **Oligonucleotides sequences used in VIGS experiments and in qRT-PCR analysis.**

895 **Supplemental Data S1**

896 **Mapping of the introgressions used in the study.** Introgressions were mapped as described in
897 Supplemental Materials. The table denotes background *S. lycopersicum* genes as 0 and genes within *S.*
898 *pennellii* introgressions as 1. The genome was divided into mapping bins as described in Materials and
899 Methods. Bin-0 contains all chromosome 0 genes.

900 **Supplemental Data S2**

901 **Results of *B. cinerea* susceptibility assays.** BIL and IL lines pathogen susceptibility assays (Sheet 1),
902 scoring of BIL and IL lines susceptibility to *B. cinerea* infection as lesion diameter in mm 5 days upon
903 spore application (R = replicate; X = not available), as described in Materials and Methods. Independent
904 validation (Sheet 2), pathogen susceptibility assays on selected BIL lines in an independent experiment.

906 **REFERENCES**

- 907 1. Gross, B. L. & Olsen, K. M. Genetic perspectives on crop domestication. *Trends Plant Sci.* **15**, 529–537
908 (2010).
- 909 2. Dixon, M. S. *et al.* The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-
910 rich repeat proteins. *Cell* **84**, 451–459 (1996).
- 911 3. Albert, E. *et al.* Association mapping reveals the genetic architecture of tomato response to water deficit:
912 focus on major fruit quality traits. *J. Exp. Bot.* **67**, 6413–6430 (2016).
- 913 4. Tieman, D. *et al.* A chemical genetic roadmap to improved tomato flavor. *Science* **355**, 391–394 (2017).
- 914 5. Causse, M., Damidaux, R. & Rousselle, P. Traditional and enhanced breeding for quality traits in tomato.
915 *Tomato. Science Publisher* **2**, 153–192 (2007).

- 916 6. Klee, H. J. & Tieman, D. M. The genetics of fruit flavour preferences. *Nat. Rev. Genet.* **19**, 347–356 (2018).
- 917 7. Lin, T. *et al.* Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* **46**, 1220–
- 918 1226 (2014).
- 919 8. Soyk, S. *et al.* Bypassing Negative Epistasis on Yield in Tomato Imposed by a Domestication Gene. *Cell* **169**,
- 920 1142–1155.e12 (2017).
- 921 9. Zhu, G. *et al.* Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell* **172**, 249–261.e12 (2018).
- 922 10. Gao, L. *et al.* The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.*
- 923 **51**, 1044–1051 (2019).
- 924 11. Alseekh, S. *et al.* Canalization of Tomato Fruit Metabolism. *Plant Cell* **29**, 2753–2765 (2017).
- 925 12. Ning, J. *et al.* A Feedback-Insensitive Isopropylmalate Synthase Affects Acylsugar Composition in Cultivated
- 926 and Wild Tomato. *Plant Physiol.* **169**, 1821–1835 (2015).
- 927 13. Fridman, E., Pleban, T. & Zamir, D. A recombination hotspot delimits a wild-species quantitative trait locus
- 928 for tomato sugar content to 484 bp within an invertase gene. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4718–4723
- 929 (2000).
- 930 14. Ofner, I., Lashbrooke, J., Pleban, T., Aharoni, A. & Zamir, D. *Solanum pennellii* backcross inbred lines (BILs)
- 931 link small genomic bins with tomato traits. *Plant J.* **87**, 151–160 (2016).
- 932 15. Eshed, Y. & Zamir, D. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato
- 933 enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**, 1147–1162 (1995).
- 934 16. Müller, N. A. *et al.* Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nat.*
- 935 *Genet.* **48**, 89–93 (2016).
- 936 17. Cárdenas, P. D. *et al.* Pathways to defense metabolites and evading fruit bitterness in genus *Solanum* evolved
- 937 through 2-oxoglutarate-dependent dioxygenases. *Nat. Commun.* **10**, 5169 (2019).
- 938 18. Tzfadia, O. *et al.* The ‘TranSeq’3'-end sequencing method for high-throughput transcriptomics and gene space
- 939 refinement in plant genomes. *Plant J.* **96**, 223–232 (2018).
- 940 19. Fulop, D. *et al.* A New Advanced Backcross Tomato Population Enables High Resolution Leaf QTL Mapping
- 941 and Gene Identification. *G3: Genes, Genomes, Genetics*, **6**, 3169–3184 (2016)..

- 942 20. Fujisawa, M. *et al.* Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and
943 associated MADS box proteins. *Plant Cell* **26**, 89–101 (2014).
- 944 21. Adato, A. *et al.* Fruit-surface flavonoid accumulation in tomato is controlled by a SIMYB12-regulated
945 transcriptional network. *PLoS Genet.* **5**, e1000777 (2009).
- 946 22. Ballester, A.-R. *et al.* Biochemical and Molecular Analysis of Pink Tomatoes: Deregulated Expression of the
947 Gene Encoding Transcription Factor SIMYB12 Leads to Pink Tomato Fruit Color. *Plant Physiology*. **152**, 71–
948 84 (2010).
- 949 23. Muir, S. R. *et al.* Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased
950 levels of flavonols. *Nat. Biotechnol.* **19**, 470–474 (2001).
- 951 24. Ballester, A.-R. *et al.* Identification of loci affecting accumulation of secondary metabolites in tomato fruit of
952 a *Solanum lycopersicum* × *Solanum chmielewskii* introgression line population. *Front. Plant Sci.* **7**, 1428
953 (2016).
- 954 25. Itkin, M. *et al.* GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and
955 prevention of phytotoxicity in tomato. *Plant Cell* **23**, 4507–4525 (2011).
- 956 26. Yamanaka, T. *et al.* C22 isomerization in alpha-tomatine-to-esculeoside A conversion during tomato ripening
957 is driven by C27 hydroxylation of triterpenoidal skeleton. *J. Agric. Food Chem.* **57**, 3786–3791 (2009).
- 958 27. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* **58**,
959 267–288 (1996).
- 960 28. Oliver, S. Guilt-by-association goes global. *Nature*. **403**, 601–603 (2000).
- 961 29. Burse, A., Weingart, H. & Ullrich, M. S. The phytoalexin-inducible multidrug efflux pump AcrAB contributes
962 to virulence in the fire blight pathogen, *Erwinia amylovora*. *Mol. Plant. Microbe. Interact.* **17**, 43–54 (2004).
- 963 30. Moco, S. *et al.* Tissue specialization at the metabolite level is perceived during the development of tomato
964 fruit. *J. Exp. Bot.* **58**, 4131–4146 (2007).
- 965 31. Bylesjö, M. *et al.* Integrated Analysis of Transcript, Protein and Metabolite Data To Study Lignin
966 Biosynthesis in Hybrid Aspen. *J. Proteome Res.* **8**, 199–210 (2009).
- 967 32. Srivastava, V. *et al.* OnPLS integration of transcriptomic, proteomic and metabolomic data shows multi-level

- oxidative stress responses in the cambium of transgenic hipI- superoxide dismutase *Populus* plants. *BMC Genomics* **14**, 893 (2013).
33. Szymanski, J., Brotman, Y., Willmitzer, L. & Cuadros-Inostroza, Á. Linking gene expression and membrane lipid composition of *Arabidopsis*. *Plant Cell* **26**, 915–928 (2014).
34. Wentzell, A. M. *et al.* Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. *PLoS Genet.* **3**, 1687–1701 (2007).
35. Iijima, Y. *et al.* Metabolite annotations based on the integration of mass spectral information. *Plant J.* **54**, 949–962 (2008).
36. Mintz-Oron, S. *et al.* Gene expression and metabolism in tomato fruit surface tissues. *Plant Physiol.* **147**, 823–851 (2008).
37. Rick, C. M., Uhlig, J. W. & Jones, A. D. High alpha-tomatine content in ripe fruit of Andean *Lycopersicon esculentum* var. cerasiforme: developmental and genetic aspects. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 12877–12881 (1994).
38. Sonawane, P. D. *et al.* Short-chain dehydrogenase/reductase governs steroidal specialized metabolites structural diversity and toxicity in the genus *Solanum*. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5419–E5428 (2018).
39. Mansfield, J. W. Antimicrobial Compounds and Resistance. in *Mechanisms of Resistance to Plant Diseases* 325–370 (Springer, Dordrecht, 2000).
40. Naoumkina, M. A. *et al.* Genome-wide analysis of phenylpropanoid defence pathways. *Mol. Plant Pathol.* **11**, 829–846 (2010).
41. Troncoso-Rojas, R. *et al.* A fungal elicitor enhances the resistance of tomato fruit to *Fusarium oxysporum* infection by activating the phenylpropanoid metabolic pathway. *Phytoparasitica* **41**, 133–142 (2013).
42. Shinde, B. A. *et al.* Dynamic metabolic reprogramming of steroidal glycol-alkaloid and phenylpropanoid biosynthesis may impart early blight resistance in wild tomato (*Solanum arcanum* Peralta). *Plant Mol. Biol.* **95**, 411–423 (2017).
43. Webb, M. E. & Smith, A. G. Pantothenate Biosynthesis in Higher Plants. in *Advances in Botanical Research*

(eds. Rébeillé, F. & Douce, R.) vol. 58 203–255 (Academic Press, 2011).

44. Lewandowska, M., Keyl, A. & Feussner, I. Wax biosynthesis upon danger: its regulation upon abiotic and biotic stress. *New Phytol.* (2020) doi:10.1111/nph.16571.

45. Lim, G.-H., Singhal, R., Kachroo, A. & Kachroo, P. Fatty acid– and lipid-mediated signaling in plant defense. *Annual Review of Phytopathology* **55**, 505–536 (2017).

46. Cantu, D. *et al.* Ripening-regulated susceptibility of tomato fruit to *Botrytis cinerea* requires NOR but not RIN or ethylene. *Plant Physiol.* **150**, 1434–1449 (2009).

47. Fu, X., Li, C., Zhou, X., Liu, S. & Wu, F. Physiological response and sulfur metabolism of the *V. dahliae*-infected tomato plants in tomato/potato onion companion cropping. *Sci. Rep.* **6**, 36445 (2016).

48. Scotti, R., D’Agostino, N. & Zaccardelli, M. Gene expression profiling of tomato roots interacting with *Pseudomonas fluorescens* unravels the molecular reprogramming that occurs during the early phases of colonization. *Symbiosis* **78**, 177–192 (2019).

49. Ruduś, I., Sasiak, M. & Kępczyński, J. Regulation of ethylene biosynthesis at the level of 1-aminocyclopropane-1-carboxylate oxidase (ACO) gene. *Acta Physiol. Plant* **35**, 295–307 (2013).

50. Cristescu, S. M., De Martinis, D., Te Lintel Hekkert, S., Parker, D. H. & Harren, F. J. M. Ethylene production by *Botrytis cinerea* in vitro and in tomatoes. *Appl. Environ. Microbiol.* **68**, 5342–5350 (2002).

51. Mach, J. M., Castillo, A. R., Hoogstraten, R. & Greenberg, J. T. The *Arabidopsis*-accelerated cell death gene *ACD2* encodes red chlorophyll catabolite reductase and suppresses the spread of disease symptoms. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 771–776 (2001).

52. van Baarlen, P., Woltering, E. J., Staats, M. & van Kan, J. A. L. Histochemical and genetic analysis of host and non-host interactions of *Arabidopsis* with three *Botrytis* species: an important role for cell death control. *Mol. Plant Pathol.* **8**, 41–54 (2007).

53. D’Esposito, D. *et al.* Unraveling the complexity of transcriptomic, metabolomic and quality environmental response of tomato fruit. *BMC Plant Biol.* **17**, 66 (2017).

54. Xu, D. *et al.* Fulvic acid-induced disease resistance to *Botrytis cinerea* in table grapes may be mediated by regulating phenylpropanoid metabolism. *Food Chem.* **286**, 226–233 (2019).

- 1020 55. Jaitin, D., Amit, I. & Keren-Shaul, H. High throughput transcriptome analysis. *World Patent* (2014).
- 1021 56. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**,
- 1022 10–12 (2011).
- 1023 57. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution.
- 1024 *Nature* **485**, 635–641 (2012).
- 1025 58. Bolger, A. *et al.* The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* **46**,
- 1026 1034–1038 (2014).
- 1027 59. Mueller, L. A. *et al.* The SOL Genomics Network: a comparative resource for Solanaceae biology and
- 1028 beyond. *Plant Physiol.* **138**, 1310–1317 (2005).
- 1029 60. Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinformatics* **51**, 11.14.1–
- 1030 19 (2015).
- 1031 61. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing
- 1032 data. *Bioinformatics* **31**, 166–169 (2015).
- 1033 62. James, N. A. & Matteson, D. S. ecp: An R Package for Nonparametric Multiple Change Point Analysis of
- 1034 Multivariate Data. *arXiv [stat.CO]* (2013).
- 1035 63. Dimont, E., Shi, J., Kirchner, R. & Hide, W. edgeRun: an R package for sensitive, functionally relevant
- 1036 differential expression discovery using an unconditional exact test. *Bioinformatics* **31**, 2589–2590 (2015).
- 1037 64. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to
- 1038 Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
- 1039 65. Michaelson, J. J., Loguercio, S. & Beyer, A. Detection and interpretation of expression quantitative trait loci
- 1040 (eQTL). *Methods* **48**, 265–276 (2009).
- 1041 66. Shahaf, N. *et al.* The WEIZMASS spectral library for high-confidence metabolite identification. *Nat.*
- 1042 *Commun.* **7**, 12423 (2016).
- 1043 67. Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry
- 1044 data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**,
- 1045 779–787 (2006).

- 1046 68. Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R. & Neumann, S. CAMERA: an integrated strategy for
1047 compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal.*
1048 *Chem.* **84**, 283–289 (2012).
- 1049 69. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. SVA: Surrogate variable analysis. *R*
1050 *package version 3*, (2013).
- 1051 70. Schymanski, E. L. *et al.* Identifying small molecules via high resolution mass spectrometry: communicating
1052 confidence. *Environ. Sci. Technol.* **48**, 2097–2098 (2014).
- 1053 71. Salek, R. M. *et al.* Embedding standards in metabolomics: the Metabolomics Society data standards task
1054 group. *Metabolomics* **11**, 782–783 (2015).
- 1055 72. Flors, V. *et al.* Absence of the endo- β -1, 4-glucanases Cel1 and Cel2 reduces susceptibility to *Botrytis cinerea*
1056 in tomato. *Plant J.* **52**, 1027–1040 (2007).
- 1057 73. Orzaez, D. *et al.* A visual reporter system for virus-induced gene silencing in tomato fruit based on
1058 anthocyanin accumulation. *Plant Physiol.* **150**, 1122–1134 (2009).
- 1059 74. Romero, I., Tikunov, Y. & Bovy, A. Virus-induced gene silencing in detached tomatoes and biochemical
1060 effects of phytoene desaturase gene silencing. *J. Plant Physiol.* **168**, 1129–1135 (2011).
- 1061 75. Fantini, E., Falcone, G., Frusciante, S., Giliberto, L. & Giuliano, G. Dissection of tomato lycopene
1062 biosynthesis through virus-induced gene silencing. *Plant Physiol.* **163**, 986–998 (2013).
- 1063 76. Kiferle, C. *et al.* Tomato R2R3-MYB Proteins SLANT1 and SLAN2: Same Protein Activity, Different Roles.
1064 *PLoS One* **10**, e0136365 (2015).
- 1065 77. Zhang, Y. *et al.* Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing
1066 susceptibility to gray mold. *Curr. Biol.* **23**, 1094–1100 (2013).
- 1067 78. Cai, X. *et al.* Efficient gene silencing induction in tomato by a viral satellite DNA vector. *Virus Res.* **125**, 169–
1068 175 (2007).
- 1069 79. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate
1070 descent. *J. Stat. Softw.* **33**, 1 (2010).
- 1071 80. Cárdenas, P. D. *et al.* GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the

1072 plant mevalonate pathway. *Nat. Commun.* **7**, 10654 (2016).

1073 81. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. Cluster: cluster analysis basics and
1074 extensions. *R package version 1.15.1*. (2014).

1075 82. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and
1076 bioinformatics. *Genome Biol.* **5**, R80 (2004).

1077

Figure 1

Figure 2

Figure 3

Figure 4

a *B. cinerea* resistance prediction

b Contribution to phenotype prediction

Figure 5

Figure 6

Figure 7