Expansion of a core regulon by transposable elements promotes Arabidopsis chemical diversity and pathogen defense

Brenden Barco* (1)(3), Yoseph Kim (2), and Nicole K. Clay (1)

*Correspondence should be addressed to B. Barco (brenden.barco@syngenta.com)

Affiliations:

(1) Department of Molecular, Cellular & Developmental Biology, Yale University, Kline Biology Tower 734, 219 Prospect St., New Haven, CT 06511

(2) Hopkins School, 986 Forest Rd, New Haven, CT 06515

Present address:

(3) Seeds Research, Syngenta Crop Protection, 9 Davis Dr., Durham, NC 27703

Abstract

Plants synthesize hundreds of thousands of ecologically specialized, lineage-specific metabolites through biosynthetic gene duplication and functional specialization. However, the rewiring of duplicated genes into existing regulatory networks remains unclear. We show that the duplicated gene *CYP82C2* was recruited into the WRKY33 regulon and indole-3-carbonylnitrile (ICN) biosynthetic pathway through exaptation of a retroduplicated LINE retrotransposon (*EPCOT3*) into a novel enhancer. The stepwise development of a chromatin-accessible WRKY33-binding site on *EPCOT3* potentiated the regulatory neofunctionalization of *CYP82C2* and the evolution of inducible defense metabolite 4-hydroxy-ICN in *Arabidopsis thaliana.* Transposable elements (TEs) have long been recognized to have the potential to rewire regulatory networks; these results establish a more complete understanding of how duplicated genes and TEs contribute in concert to chemical diversity and pathogen defense.

Plant secondary or specialized metabolites are essential for plant survival in co-evolving biotic and fluctuating abiotic environments. The evolutionary process of chemical innovation resulted in the collective synthesis of hundreds of thousands of ecologically specialized, mostly lineage-specific metabolites (Chae et al., 2014; Weng et al., 2012; Dixon and Strack, 2003; Wink, 2003). Plant specialized metabolic enzymes are ultimately produced from primary metabolic enzymes through gene duplication and subsequent functional divergence of one or both paralogs to produce enzymes with altered expression patterns and/or protein functions (Ohno, 1970; Force *et al*., 1999;

36 Weng *et al*., 2012). They are also often organized into transcription factor (TF) regulons
37 of co-regulated genes for optimal timing, amplitude, and tissue-specific pathway gene
38 expression and subsequent metabolite accumulation (Grotewold, 2005; Hartmann,
39 2007; Martin *et al*., 2010; Tohge & Fernie, 2012; Omranian et al., 2015).
40
41 Changes in *cis*-regulatory modules such as enhancers and promoters can accelerate
42 the capture of duplicated genes into regulons, thus driving phenotypic diversity (Levine
43 and Davidson, 2005; Prud'homme *et al*., 2007; Wray, 2007; Wittkopp & Kalay, 2012;
44 Rogers *et al*., 2013). Enhancers consist of transcription factor binding sites (TFBSs) and
45 are derived either through mutation or co-option of a TFBS-carrying transposable
46 element (TE) (Spitz & Furlong, 2012; Wittkopp & Kalay, 2012). TE exaptations have
47 been hypothesized to be responsible for the rapid transcriptional rewiring of gene
48 regulatory networks in ancient lineages of vertebrates (Feschotte 2008; Bourque 2009;
49 Lynch *et al*., 2011; de Souza *et al*., 2013; Chuong et al., 2016) and plants (Hénaff *et al*.,
50 2014), but general understandings of the physiological significance of this rewiring are
51 greatly limited.
52
53 Bacteria elicit two primary immune defense modes in plants, pattern- and effector-
54 triggered immunity (PTI and ETI) (Jones & Dangl, 2006). Pathogenic bacteria
55 additionally compromise PTI via specific virulence effector proteins (effector-triggered
56 susceptibility, ETS; Jones & Dangl, 2006). PTI involves the extracellular perception of
57 conserved molecules known as microbe-associated molecular patterns (MAMPs),

58 whereas ETI involves the cytosolic perception of effectors. Although ETI results in the
59 formation of more rapid and robust pathogen-specific response including the
60 hypersensitive response (HR), a form of programmed cell death (Jones & Dangl, 2006),
61 both result in the ability of naïve host cells to generate, through non-self perception and
62 subsequent transcriptional reprogramming, pathogen-inducible specialized metabolites
63 necessary for defense (Hammerschmidt, 1999; Mansfield, 2000; Clay *et al.*, 2009).
64
65 Three pathogen-inducible tryptophan (Trp)-derived defense metabolites – camalexin, 4-
66 methoxyindol-3-ylmethylgucosinolate (4M-I3M), and 4-hydroxyindole-3-carbonylnitrile
67 (4OH-ICN) – have been shown to expand innate immunity in *Arabidopsis thaliana*
68 (Bednarek *et al.,* 2009; Clay *et al.,* 2009; Thomma *et al.,* 1999; Tsuji *et al.,* 1992;
69 Rajniak *et al.,* 2015). The three biosynthetic pathways share an early step, which is the
70 conversion of Trp to indole-3-acetaldoxime (IAOx) via the genetically redundant P450
71 monooxygenases CYP79B2 and CYP79B3 (Fig. 1a) (Hull *et al.*, 2000; Mikkelsen *et al.*,
72 2000; Bak *et al.*, 2001; Glawischnig *et al.*, 2004; Klein *et al.*, 2013; Rajniak *et al.*, 2015).
73 The camalexin and 4OH-ICN pathways additionally share the conversion of IAOx to
74 indole-3-cyanohydrin (ICY) by partially redundant P450s CYP71A12 and CYP71A13
75 (Fig. 1a) (Nafisi *et al.*, 2007; Klein *et al.*, 2013; Rajniak *et al.*, 2015). CYP71A13 and
76 CYP71B15/PAD3 catalyze further reactions, leading to camalexin production, whereas
77 the flavin-dependent oxidase FOX1/AtBBE3 and P450 CYP82C2 convert ICY to 4OH-
78 ICN (Fig. 1a) (Nafisi *et al.*, 2007; Böttcher *et al.*, 2009; Rajniak *et al.*, 2015). 4M-I3M is
79 widely distributed across the mustard family (Brassicaceae), whereas camalexin is

80 restricted to the Camelineae tribe of Brassicaceae (Bednarek *et al.*, 2011). The

81 evolutionary conservation of 4OH-ICN has not yet been investigated.

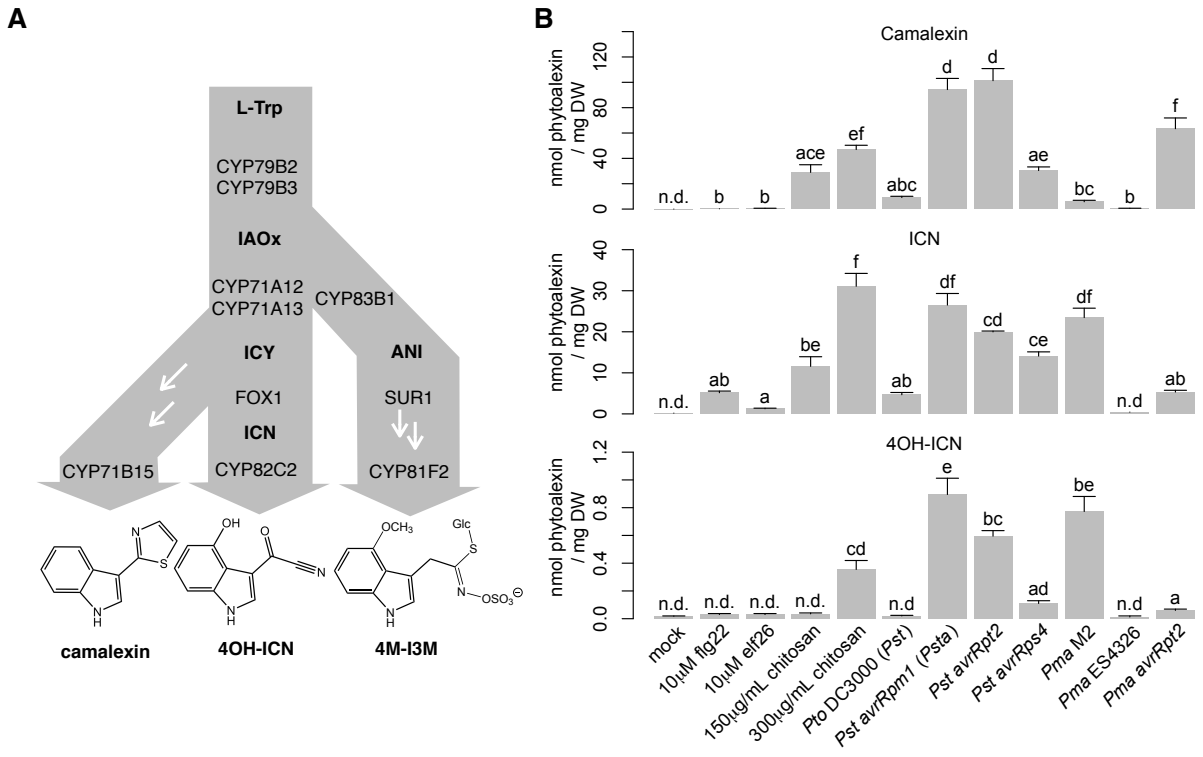


Figure 1. 4OH-ICN is synthesized under ETI-like responses. (**a**). Schematic of tryptophan (L-Trp)-derived specialized metabolism in *A. thaliana*. White arrows denote the presence of additional enzymes. ICY, indole cyanohydrin; ANI, *aci*-nitro indole. (**b**). LC-DAD-FLD-MS analysis of camalexin (top), ICN (middle), and 4OH-ICN (bottom) in seedlings elicited with indicated MAMPs and bacterial strains for 27 hr. Data represent mean ± SE of 3-4 biological replicates. Different letters denote statistically significant differences ($P < 0.05$, one-factor ANOVA coupled to Tukey's test). ICA-ME and 4OH-ICA-ME are methanolic degradation products of ICN and 4OH-ICN, respectively. 4OH-ICA is an aqueous degradation product of 4OH-ICN. For (**b**), source data are provided as a Source Data file.

93 The TF WRKY33 has been shown to regulate the pathogen-inducible biosynthesis of
94 camalexin in *A. thaliana* and its orthologs regulate numerous unrelated specialized
95 metabolites in other flowering plant lineages (Qiu *et al.,* 2005; Liu *et al.,* 2015; Birkenbihl
96 *et al.,* 2017; Schluttenhofer & Yuan, 2015). The group I class of WRKYs to which
97 WRKY33 belongs is an ancient clade of regulators; orthologs in the green alga
98 *Chlamydomonas reinhardtii* may be ancestral to all higher plant WRKYs (Rinerson *et*
99 *al*., 2015; Schluttenhofer & Yuan, 2015). While all WRKY TFs bind to the W-box core
100 sequence [TTGAC(T/C)], WRKY33 preferentially binds W-boxes that are within 500 nt
101 of the 'WRKY33-specific' motif [(T/G)TTGAAT]) (Rushton *et al.,* 2010; Liu *et al*., 2015).
102
103 Here, we show that a recent, lineage-specific TE exaptation resulted in the expansion of
104 a core regulon within the framework of Arabidopsis Trp-derived defense metabolism.
105 Specifically, the LINE retrotransposon *EPCOT3* retroduplicated from a WRKY33-TFBS-
106 carrying progenitor and inserted upstream of the newly duplicated gene *CYP82C2*.
107 Subsequent chromatin remodeling in *A. thaliana* led *EPCOT3* to become a *bona fide*
108 enhancer with demonstrated biochemical, regulatory, physiological, and fitness-
109 promoting characteristics by way of WRKY33-binding and pathogen-responsive
110 *CYP82C2* transcription, 4OH-ICN biosynthesis, and antibacterial defense.
111
112 **Results**
113 **4OH-ICN requires ETI-like responses.** To identify the major Trp-derived specialized
114 metabolites synthesized in ETI in *A. thaliana*, we compared host transcriptional and

115 metabolic responses to the PTI-eliciting bacterial MAMPs flg22, elf26, and fungal MAMP
116 chitosan; the PTI/ETS-eliciting pathogens *Pseudomonas syringae* pv. *tomato* DC3000
117 (*Pto* DC3000 or *Pst*); *Pseudomonas syringae* pv. *maculicola* ES4326 (*Pma*); and the
118 ETI-eliciting pathogens *Pst avrRpm1* (*Psta*), *Pst avrRpt2*, *Pst avrRps4*, *Pma* M2, and
119 *Pma avrRpt2* under similar conditions as those of previous studies (Denoux *et al*., 2008;
120 Clay *et al*., 2009). *Psm* M2 is an ETI-eliciting strain from which the *avrRpm1* gene was
121 originally isolated (Debener *et al.*, 1991). Both flg22 and *Psta* induced genes involved in
122 4OH-ICN, camalexin and 4M-I3M biosynthesis, with 4OH-ICN and camalexin
123 biosynthetic genes having a higher level of induction than those of 4M-I3M in *Psta*-
124 inoculated plants (Supplementary Fig. 1a; Denoux *et al*., 2008). In contrast to the
125 quantitative differences observed in transcriptional responses between PTI and ETI
126 (Tao *et al*., 2003; Navarro *et al*., 2004), the metabolite responses between PTI and ETI
127 differed largely qualitatively. 4M-I3M and the 4M-I3M precursor molecule 4OH-I3M were
128 present in uninfected plants and accumulated to modest levels at the expense of parent
129 metabolite I3M in flg22- and *Psta*-inoculated plants (Supplementary Fig. 1b) (Clay *et al*.,
130 2009). By comparison, ICN, 4OH-ICN, and camalexin were absent in uninfected plants
131 and at low-to-undetectable levels in plants treated with saturating concentrations of the
132 bacterial MAMPs flg22 and elf26 (10 μM; Felix *et al*., 1999; Zipfel *et al*., 2006). In
133 contrast, ICN, 4OH-ICN and camalexin accumulated to high levels upon inoculation with
134 ETI-inducing pathogens (Fig. 1b; Supplementary Fig. 1c). Furthermore, camalexin, ICN,
135 and 4OH-ICN metabolism was greatly diminished, and indole glucosinolate levels were
136 mostly unchanged in the *rpm1* mutant, which is impaired in ETI recognition of *Psta*

137 (Bisgrove *et al*., 1994) (Supplementary Fig. 1b-c). By contrast, camalexin and ICN were
138 absent in uninfected plants and largely at low-to-undetectable levels in plants treated
139 with MAMPs and PTI/ETS-eliciting pathogens, with 4OH-ICN not detected in most
140 cases. One exception was the fungal MAMP chitosan. 150 μg/mL chitosan induced high
141 levels of camalexin and detectable levels of ICN, consistent with previous observations
142 of camalexin biosynthetic genes upregulation (Fig. 1b) (Povero *et al*., 2011). Higher
143 chitosan concentrations (≥200 μg/mL) have been shown to induce HR-like cell death in
144 Arabidopsis (Cabrera et al., 2006), a phenomenon commonly observed for ETI (Jones
145 and Dangl, 2006). To our surprise, 300 μg/mL chitosan additionally induced detectable
146 levels of 4OH-ICN (Fig. 1b). These results suggest that 4OH-I3M, 4M-I3M, camalexin,
147 and ICN are synthesized in response to multiple PTI elicitors, whereas 4OH-ICN
148 biosynthesis is specific to ETI-like responses.
149
150 ***WRKY33* is required to activate 4OH-ICN in response to *Psta***. 4OH-ICN biosynthetic
151 genes are highly co-expressed with each other (Rajniak *et al*., 2015) and with camalexin
152 biosynthetic genes (Supplementary Fig. 1d), which are in the WRKY33 regulon (Qiu *et*
153 *al*., 2008; Birkenbihl *et al*., 2012). To determine whether 4OH-ICN biosynthetic genes
154 are also in the WRKY33 regulon, we compared camalexin, ICN and 4OH-ICN levels
155 between wild-type and a *wrky33* loss-of-function mutant that encodes two differently
156 truncated proteins (Fig. 2a; Zheng *et al*., 2006). Consistent with a previous report (Qiu
157 *et al*., 2008), *wrky33* was impaired in camalexin biosynthesis in response to *Psta* and
158 *Pst avrRps4* (Fig. 2b; Supplementary Fig. 2a). The *wrky33* mutant was similarly

159 impaired in 4OH-ICN biosynthesis (Fig. 2b; Supplementary Fig. 2a). These results
160 indicate that WRKY33 is required for camalexin and 4OH-ICN biosynthesis in response
161 to multiple ETI elicitors.
162

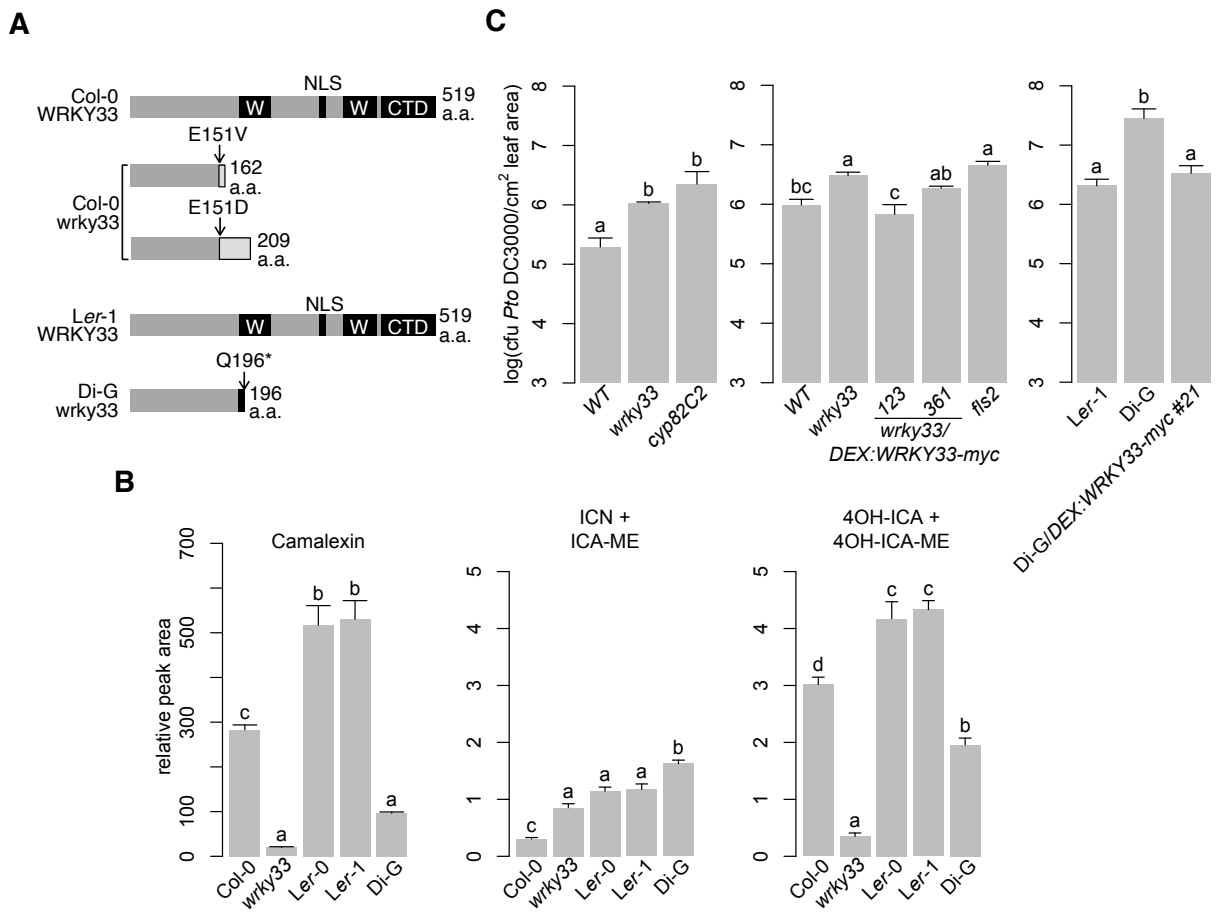


Figure 2. Intraspecific variation in *WRKY33* affects 4OH-ICN and immunity. (**a**) Schematic of WRKY33 proteins in Col-0, Col-0 *wrky33*, L*er*-1 and Di-G. Black boxes denote WRKY domains (W), nuclear localization signal (NLS), or C-terminal domain (CTD). (**b**) LC-DAD-MS analysis of camalexin, ICN, and 4OH-ICN in seedlings inoculated with *Psta* for 24 hr. Data represent mean ± SE of four replicates. (**c**) Bacterial growth analysis of *Pst* in surface-inoculated leaves. Middle and right panels were pre-treated with 20 μM dex for 6-8 hr. Data represent mean ± SE of 4 (left), 6-11 (middle), and 6-8 (right) biological replicates. CFU, colony-forming units. Different letters in (**b-c**) denote statistically significant differences ($P < 0.05$, one-factor ANOVA coupled to

173 Tukey's test). Experiments in (**b-c**) were performed at least twice, producing similar

174 results. For (**b-c**), source data are provided as a Source Data file.

175

176

177

178 To confirm that *WRKY33* is required to activate the 4OH-ICN pathway, we used a two-
179 component glucocorticoid-inducible system to generate *wrky33* plants that in the
180 presence of the glucocorticoid hormone dexamethasone (dex) express a wild-type copy
181 of WRKY33 with a C-terminal fusion to 1x flag epitope (*wrky33/DEX:WRKY33-flag*;
182 Supplementary Fig. 2b-c). Induced expression of *WRKY33-flag* restored camalexin and
183 4OH-ICN biosynthesis in *Psta*-challenged *wrky33* plants to greater than wild-type levels
184 (Supplementary Fig. 2d). These results indicate that *WRKY33* is required to activate
185 camalexin and 4OH-ICN biosynthesis in response to *Psta*.
186
187 **Intraspecific variation in *WRKY33* affects 4OH-ICN synthesis and pathogen**
188 **defense**. Intraspecific variation in TFs can contribute to gain or loss of phenotypes,
189 such as branching in maize (Studer *et al*., 2011) or pelvic loss in three-spined
190 stickleback fish (Chan *et al*., 2010). In addition, the wide variation in camalexin
191 biosynthesis reported among natural accessions of *A. thaliana* (Kagan &
192 Hammerschmidt, 2002) suggests that a similar variation in 4OH-ICN biosynthesis may
193 exist. To identify additional transcriptional activators of 4OH-ICN biosynthesis that
194 otherwise might be refractory to traditional genetic approaches, we compared
195 intraspecific variation in *Psta*-induced camalexin, ICN and 4OH-ICN among 35 re-
196 sequenced accessions and *wrky33* (Col-0 accession). We found camalexin and 4OH-
197 ICN levels to be positively correlated among accessions (R^2 = 0.37**;** Supplementary Fig.
198 3a), lending further support to their co-regulation by WRKY33. Accession Dijon-G (Di-G)
199 was identified to produce less camalexin and 4OH-ICN and more ICN than its near-

200 isogenic relatives, the Landsberg accessions L*er*-0 and L*er*-1 (Fig. 2b; Supplementary
201 Fig. 3a-b). In addition, differences observed in the metabolite response between
202 Landsberg accessions and Di-G most closely resembled those between Col-0 and
203 *wrky33* mutant (Fig. 2b; Supplementary Fig. 3a). These results led us to hypothesize
204 that genetic variation in a regulatory gene, as opposed to an immune signaling gene, is
205 responsible for the metabolite phenotypes observed in Di-G. To test this hypothesis,
206 genetic variation between Di-G and three sequenced Landsberg accessions (La-0, L*er*-
207 0, and L*er*-1) were used to identify 354 genes that were differentially mutated to high
208 effect in Di-G (Supplementary Fig. 3c). Twenty-eight of these mutated Di-G genes were
209 annotated by Gene Ontology to have roles in defense, including *WRKY33*
210 (Supplementary Table 1). We confirmed by Sanger sequencing that Di-G *WRKY33*
211 harbors a nonsense mutation early in the N-terminal DNA-binding motif (Fig. 2a), likely
212 abolishing protein function. Our findings indicate that camalexin and 4OH-ICN are
213 sensitive to intraspecific variation in *WRKY33*.
214
215 Camalexin and 4OH-ICN promote plant fitness by contributing non-redundantly to
216 pathogen defense against the fitness-reducing *Pst* (Kover & Scaal, 2002; Rajniak *et al.*,
217 2015). To confirm that disease resistance to *Pst* is also sensitive to intraspecific
218 variation in *WRKY33*, we measured bacterial growth in adult leaves of *wrky33* and Di-G
219 and their respective (near-)isogenic accessions Col-0 and L*er*-1. *wrky33* and Di-G were
220 more susceptible to *Pst* than their (near)isogenic relatives and comparable to the 4OH-
221 ICN biosynthetic mutant *cyp82C2* (Fig. 2c; Rajniak *et al.*, 2015).

222

223 We additionally generated *wrky33* plants that in the presence of dex express a wild-type
224 copy of WRKY33 with a C-terminal fusion to a larger 6x myc epitope
225 (*wrky33/DEX:WRKY33-myc*; Supplementary Fig. 4a-c). Induced expression of
226 WRKY33-myc complemented *wrky33* and Di-G to Col-0 and L*er*-1 levels of resistance
227 to *Pst*, respectively (Fig. 2c). Additionally camalexin and ICN levels complemented
228 and/or exceeded Col-0 and L*er*-1 levels in *Psta*-challenged *wrky33/DEX:WRKY33-myc*
229 and *Di-G/DEX:WRKY33-myc* plants, respectively (Supplementary Fig. 4d-e). Together,
230 our results support a role of *WRKY33* in pathogen defense as an activator of Trp-
231 derived specialized metabolism.

232

233 **WRKY33 activates 4OH-ICN biosynthesis**. To confirm that the 4OH-ICN biosynthetic
234 pathway is in the WRKY33 regulon, we first compared *WRKY33*, *CYP71A13,*
235 *CYP71B15, FOX1* and *CYP82C2* transcript levels among WT, *wrky33*,
236 *wrky33/DEX:WRKY33-flag*, and *wrky33/DEX:WRKY33-myc*. Consistent with previous
237 reports (Qiu *et al*., 2008), *CYP71A13*, *CYP71B15*, and *FOX1* expression was down-
238 regulated in *wrky33* plants in response to *Psta* and upregulated in both
239 *wrky33/DEX:WRKY33-flag* and *wrky33/DEX:WRKY33-myc* (Fig. 3a) (Supplementary
240 Fig. 4f, 5a). Interestingly, *CYP82C2* expression and 4OH-ICN production were restored
241 in *wrky33/DEX:WRKY33-flag* but not *wrky33/DEX:WRKY33-myc* or Di-
242 G*/DEX:WRKY33-myc* plants (Fig. 2d, 3a) (Supplementary Fig. 4d-f), likely due to the
243 interference of the larger myc tag with the WRKY33 C-terminus, a region previously

244 linked with transactivation activity (Zhou *et al.*, 2015). These transcriptional and

245 metabolic findings indicate that WRKY33 mediates camalexin and 4OH-ICN

246 biosynthesis in response to pathogen effectors.

247

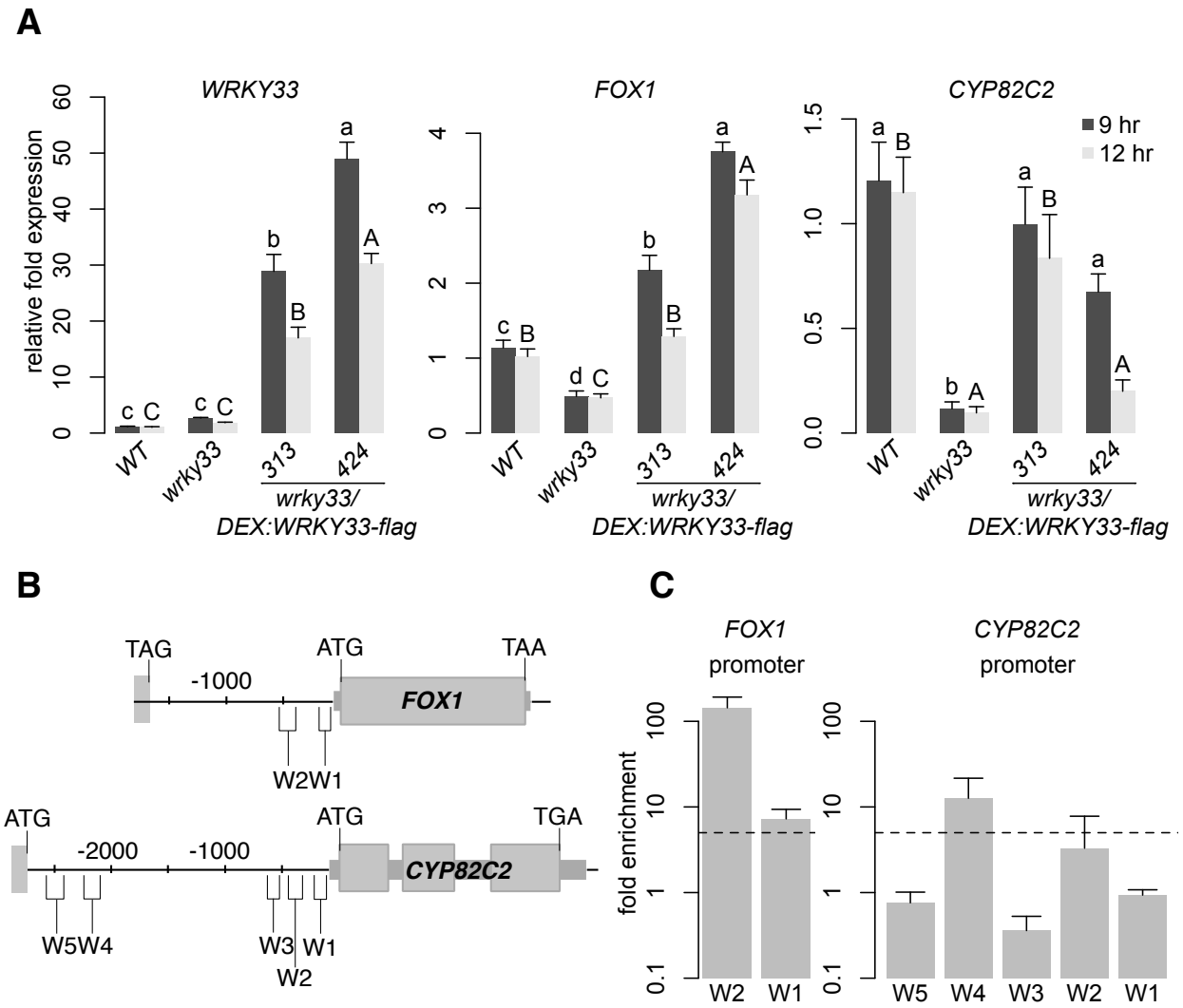


Figure 3. WRKY33 directly activates 4OH-ICN biosynthetic genes. (**a**) qPCR analysis of 4OH-ICN regulatory and biosynthetic genes in seedlings inoculated with 20 *μ*M dex and *Psta* for 9 and 12 hr. Different letters denote statistically significant differences ($P < 0.05$, one-factor ANOVA coupled to Tukey's test). Lowercase and uppercase letters denote comparisons across 9 and 12 hr timepoints, respectively. Data represent mean ± SE of 4-6 replicates. (**b**) Schematic of *FOX1* and *CYP82C2* loci, indicating nt positions of W-box-containing regions (W). (**c**) ChIP-PCR analysis of W-box-containing regions upstream of *FOX1* and *CYP82C2* in *wrky33/DEX:WRKY33-flag*

257 plants co-treated with 20 μM dex (D) or mock solution (M) and *Psta* for 9 hr. Dashed line
258 represents the 5-fold cutoff between weak and strong TF-DNA interactions. Data
259 represent mean ± SE of four replicates. For (**a**,**c**), source data are provided as a Source
260 Data file.
261
262 We then tested for WRKY33-binding to W-box-containing regions upstream of
263 camalexin and 4OH-ICN biosynthetic genes in dex-treated and *Psta*-infected
264 *wrky33/DEX:WRKY33-flag* seedlings by chromatin immunoprecipitation (ChIP)-PCR.
265 WRKY33 has been shown to bind to a W-box region upstream of *CYP71A12* (Birkenbihl
266 *et al*., 2017), a region that also contains three WRKY33-specific motifs and is consistent
267 with WRKY33's reported binding site preference (Liu *et al*., 2015). We additionally
268 observed that *Psta*-induced WRKY33 bound strongly (greater than 5-fold enrichment) to
269 a single W-box region upstream of *FOX1* and *CYP82C2* (W2 and W4, respectively; Fig.
270 3b-c; Supplementary Fig. 5b). Both regions also contain one to three WRKY33-specific
271 motifs. Together with our expression analysis, our findings indicate that WRKY33 uses
272 preferred WRKY33-binding sites to directly activate 4OH-ICN biosynthetic genes in
273 response to pathogen effectors.
274
275 Interestingly, *Psta*-induced WRKY33 did not bind to the W5 region upstream of
276 *CYP82C2* (Fig. 3c), a W-box region that does not contain any WRKY33-specific motifs
277 and is just upstream of neighboring gene of unknown function *At4g31960* (Fig. 3b).
278 WRKY33 reportedly binds to W5 in response to flg22 and *B. cinerea* (Liu *et al*., 2015;

279 Birkenbihl *et al*., 2017). By contrast, *Psta*-induced WRKY33 bound strongly to W1
280 region upstream of *CYP71B15* (Supplementary Fig. 5c-d), a W-box region that also
281 does not contain any WRKY33-specific motifs. WRKY33 reportedly binds to a region
282 encompassing W1 in response to flg22 and *Psta* (Qiu *et al*., 2008; Birkenbihl *et al*.,
283 2012). These findings suggest that WRKY33 may use W-box extended motifs or novel
284 specificity motifs to target camalexin biosynthetic genes in response to pathogen
285 effectors, or 4OH-ICN biosynthetic genes in response to MAMPs or fungal pathogens.
286
287 ***CYP82C2* underwent regulatory neofunctionalization**. CYP82C2 catalyzes the last
288 step in 4OH-ICN biosynthesis, hydroxylating ICN to form 4OH-ICN (Rajniak *et al*.,
289 2015), and likely was the last 4OH-ICN pathway gene to be recruited to the WRKY33
290 regulon in *A. thaliana*. To explore the phylogenetic distribution pattern of 4OH-ICN
291 biosynthesis, we profiled ICN and 4OH-ICN metabolites in close and distant relatives of
292 *A. thaliana* in response to *Psta*. While ICN biosynthesis was observed across multiple
293 close relatives, 4OH-ICN was only detected in *A. thaliana* (Fig. 4a; Supplementary Fig.
294 6a). This result suggests that 4OH-ICN manifests a species-specific diversification of
295 pathogen-inducible Trp-derived metabolism in the mustard family.
296

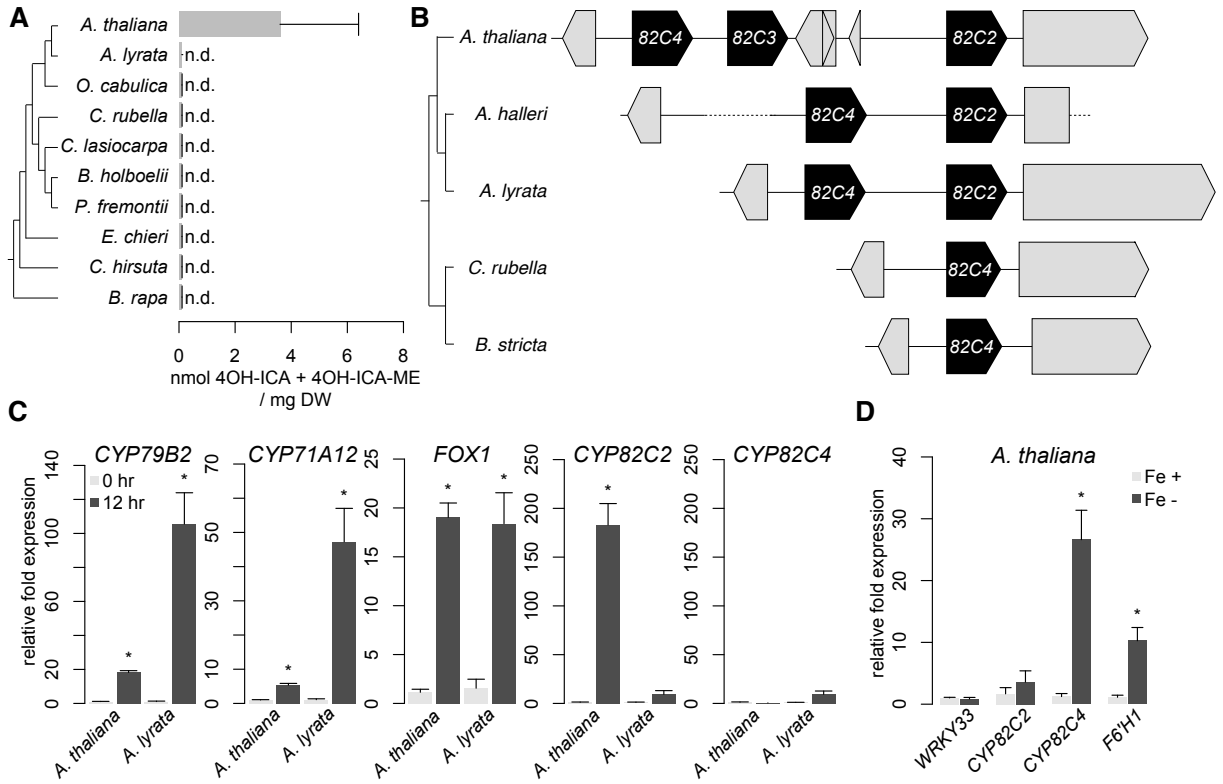


Figure 4. Regulatory neofunctionalization of *CYP82C2*. (**a**) (Left) Phylogenetic species tree. (Right) HPLC-DAD analysis of 4OH-ICN in seedlings inoculated with *Psta* for 30 hr. Data represent mean ± SE of three independent experiments (n = 4 biological replicates), each with *A. thaliana* as a positive control. 4OH-ICA and 4OH-ICA-ME are aqueous and methanolic degradations products of 4OH-ICN, respectively. DW, dry weight; n.d., not detected. (**b**) (Left) phylogenetic species tree. (Right) Synteny map of the *CYP82C* genes. Grey arrows or rectangles represent non-*CYP82C* genes. Grey dotted lines represent large (>500 nt) sequence gaps. (**c-d**) qPCR analysis of 4OH-ICN and sideretin biosynthetic genes in seedlings inoculated with *Psta* (**c**) or grown in iron-deficient medium (**d**). Data represents the mean ± SE of four biological replicates. Asterisks denote statistically significant differences of stress-treated relative to

309 untreated samples ($P < 0.05$, two-tailed t-test). For (**a**,**c-d**), source data are provided as

310 a Source Data file.

311 In *A. thaliana*, *CYP82C2* resides in a near-tandem cluster with paralogs *CYP82C3* and
312 *CYP82C4* (Fig. 4b). We performed phylogenetic and syntenic analyses to identify
313 putative *CYP82C2* orthologs in ICN-synthesizing species. All identified homologs are
314 syntenic to *CYP82C2* or *CYP82C4*, and encode proteins with >88% identity to one
315 another (Fig. 4b; Supplementary Fig. 6b-c). *CYP82C3* is present only in *A. thaliana*, and
316 although more similar to *CYP82C2* than *CYP82C4* in sequence, it is not functionally
317 redundant with *CYP82C2* (Fig. 4b; Supplementary Fig. 6b; Rajniak *et al*., 2015).
318 *CYP82C4* is required for the biosynthesis of sideretin, a widely conserved,
319 phenylalanine-derived metabolite required for iron acquisition (Rajniak *et al*., 2018).
320 *CYP82C4* has syntenic orthologs in the mustard family, correlating with the distribution
321 of sideretin biosynthesis (Fig. 4b; Supplementary Fig. 6b**;** Rajniak *et al*., 2018). By
322 contrast, *CYP82C2* has syntenic orthologs only within the *Arabidopsis* genus (Fig. 4b;
323 Supplementary Fig. 6b). These results suggest that *CYP82C2* duplicated from
324 *CYP82C4* prior to the formation of the *Arabidopsis* genus and then acquired a new
325 expression pattern and/or catalytic function prior to *A. thaliana* speciation approx. 2
326 million years later (Hu *et al*., 2011; Hohmann *et al*., 2015).
327
328 CYP82C2 and CYP82C4 were previously characterized to 5-hydroxylate with equal
329 efficiency the specialized metabolite 8-methoxypsoralen, a molecule structurally
330 reminiscent of ICN and sideretin (Kruse *et al*., 2008). The apparent similarities in
331 substrate specificity and catalytic function suggest that *CYP82C2* may have diverged
332 from *CYP82C4* in expression but not protein function. To test this, we first compared the

333 expression of *CYP82C2* and *CYP82C4* in *A. lyrata* and *A. thaliana* in response to *Psta*.
334 4OH-ICN biosynthetic genes *CYP79B2, CYP71A12* and *FOX1* were upregulated in both
335 species, consistent with the common presence of ICN (Fig. 4a and c). By contrast,
336 *CYP82C2* levels were respectively upregulated and unchanged in *A. thaliana* and *A.*
337 *lyrata*, correlating with the distribution of 4OH-ICN in these species (Fig. 4a and c).
338 *CYP82C4* expression was unchanged in both species (Fig. 4c). These results indicate
339 that 4OH-ICN biosynthesis is linked with pathogen-induced expression of *CYP82C2.*
340
341 We then compared the aligned upstream sequences of *CYP82C2* and *CYP82C4* in *A.*
342 *lyrata* and *A. thaliana* and observed good sequence conservation among orthologs but
343 poor conservation among paralogs (Supplementary Fig. 6d), indicating that sequences
344 upstream of *CYP82C4* and *CYP82C2* were independently derived. We performed
345 expression analysis in *A. thaliana* to confirm that *CYP82C2* and *CYP82C4* have
346 different expression patterns. *CYP82C2* expression is upregulated in response to *Psta*
347 and unchanged under iron deficiency (Fig. 4c-d; Supplementary Fig. 1a**;** Rajniak *et al.*,
348 2015). Conversely, *CYP82C4* is upregulated under iron deficiency and unchanged in
349 response to *Psta* (Fig. 4c-d; Murgia *et al.*, 2011; Rajniak *et al.*, 2018). Finally, *CYP82C4*
350 was unchanged in *Psta*-challenged *wrky33* and *wrky33/DEX:WRKY33-flag*
351 (Supplementary Fig. 6e). Our findings suggest that *CYP82C2* diverged from *CYP82C4*
352 by acquiring WRKY33 regulation for its pathogen-induced expression.
353

354 We next assessed dN/dS ratios along branches of the CYP82C phylogenetic tree
355 (Supplementary Fig. 6b) and found good support for purifying selection acting on
356 CYP82C enzymes (ω=0.21), and no support for positive selection acting on CYP82C2/3
357 enzymes (Supplementary Table 2). Lastly, we identified non-conserved amino acid
358 residues among CYP82C homologs and mapped this information onto a homology
359 model of CYP82C2. The protein inner core, which encompasses the active site and
360 substrate channel, is highly conserved among CYP82C homologs (Supplementary Fig.
361 6f), and is consistent with CYP82C2 and CYP82C4's reportedly redundant catalytic
362 functions (Kruse *et al*., 2008). Altogether, our findings suggest that *CYP82C2*
363 underwent regulatory neofunctionalization (Moore & Purugganan, 2005), diverging from
364 *CYP82C4* in expression but not protein function.
365
366 **TE *EPCOT3* is a *CYP82C2* enhancer**. WRKY33 regulation of *CYP82C2* is mediated by
367 a WRKY33-TFBS in the W4 region (Figs. 3 and 5a; Supplementary Fig. 5c). Preferential
368 WRKY33-binding at this region should also be influenced by chromatin features
369 associated with *cis*-regulatory elements like enhancers and basal promoters (Slattery *et*
370 *al*., 2014). To investigate how *CYP82C2* acquired WRKY33-binding for its pathogen-
371 induced expression, we compared the aligned upstream sequences of *CYP82C*
372 homologs in ICN-synthesizing species. We observed three large upstream sequences
373 specific to *A. thaliana CYP82C2*, hereafter named ***E**ighty-two-C2 **P**romoter **C**ontained*
374 ***O**nly in A. **T**haliana1-3* (*EPCOT1–3*; Fig. 5a). *EPCOT3* in particular is a 240nt region
375 that completely encompasses W4 (Fig. 5a), indicating that the WRKY33's regulation of

376 *CYP82C2* in response to *Psta* may be species-specific. Further bioinformatics analysis
377 revealed that *EPCOT3* has the epigenetic signature of an active enhancer (Roudier *et*
378 *al*., 2011; Liu *et al*., 2018). Relative to neighboring sequences, *EPCOT3* is enriched with
379 activating histone mark H3K4me2 and lacks the repressive histone mark H3K27me3
380 (Fig. 5b) (Heintzman *et al*., 2007; Hoffman *et al*., 2010; Roudier *et al*., 2011; Bonn *et al*.,
381 2012; Wang *et al*., 2014). Our findings suggest that *EPCOT3* functions as an enhancer
382 that mediates WRKY33-binding and activation of *CYP82C2* in response to pathogen
383 effectors.
384

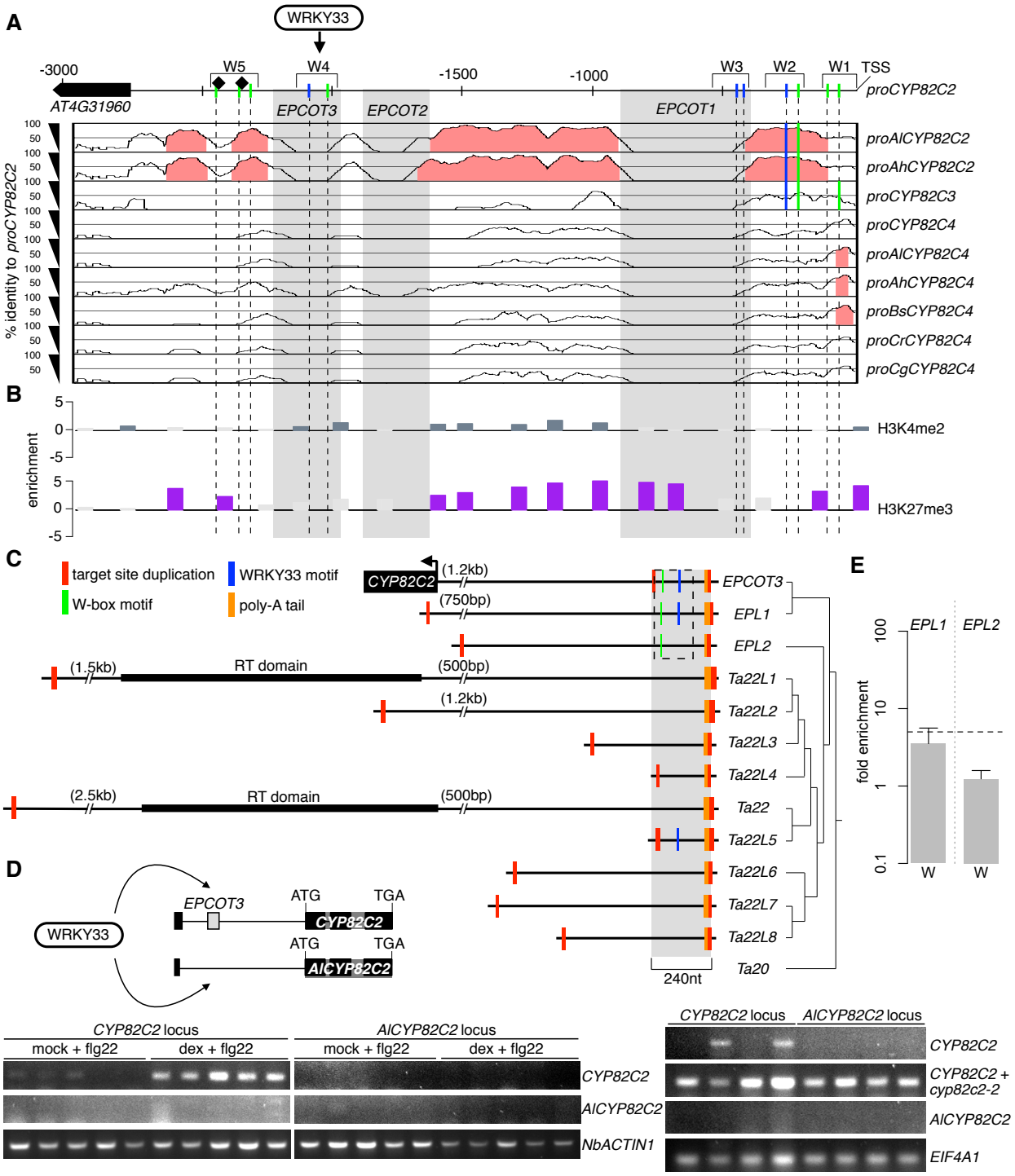


Figure 5. TE *EPCOT3* is a *CYP82C2* enhancer. (**a**) mVISTA plot of *CYP82C2* upstream sequence, indicating nt positions of unique (*EPCOT1–3;* gray boxes) and conserved regions (≥70% sequence identity; pink) among homologous sequences. Also

389 indicated are positions of W-boxes (green) and WRKY33-specific motifs (blue) that are
390 present (solid lines) or absent (dashed lines) in each homologous sequence, previously
391 known WRKY33-TFBSs (diamonds) and ChIP-tested regions (W1-5). TSS,
392 transciptional start site; *Al, Arabidopsis lyrata; Ah, Arabidopsis halleri*; *Cr, Capsella*
393 *rubella*; *Bs, Boechera stricta; Cg, Capsella grandiflora*. (**b**) Epigenetic map of *CYP82C2*
394 upstream sequence, indicating nt positions of significant amounts of H3K4me2 (blue-
395 gray bars), and H3K27me3 (purple bars). (**c**) (Left) Schematic of *EPCOT3* and related
396 LINE retrotransposons in *A. thaliana* drawn to scale, indicating nt positions of *CYP82C2*
397 and reverse transcriptase (RT) domain. A more detailed tree is available as
398 Supplementary Text 1. (Right) Phylogenetic maximum likelihood tree. Dashed box
399 represent region containing W-boxes (green lines) and/or WRKY33-binding motifs (blue
400 lines) within *EPCOT3, EPL1* and *EPL2*. (**d**) (Upper left) Schematic of *CYP82C2* and
401 *AlCYP82C2* transgenic loci used for WRKY33 transactivation experiments. Black box at
402 far left represents coding sequences of neighboring genes (*AT4G31960* at *CYP82C2*
403 locus, *AlCYP82C4* at *AlCYP82C2* locus). (Lower left) RT-PCR images of *CYP82C2*,
404 *AlCYP82C2*, and *NbACTIN1* in *N. benthamiana* leaves co-transfected with
405 *DEX:WRKY33-flag* and the *CYP82C2* or *AlCYP82C2* locus and incubated with 1 μM
406 flg22 and either mock solution (0.5% DMSO) or 20 μM dex for 30 hr (*CYP82C2* and
407 *AlCYP82C2*) or 24 hr (*NbACTIN1*). Data represents five biological replicates. (Lower
408 right) RT-PCR images of *CYP82C2*, *AlCYP82C2*, and *EIF4A1* in *A. thaliana cyp82C2-2*
409 mutant protoplasts transfected with the wild-type *CYP82C2* or *AlCYP82C2* locus and
410 elicited with 1 μM flg22 for 6 hr. As transcription is present downstream of the T-DNA

411 insertion in the *cyp82C2-2* mutant, primers for RT-PCR were selected corresponding to
412 a downstream region (*CYP82C2* + *cyp82c2-2*, second row) as well as a region flanking
413 the *cyp82C2-2* insertion (*CYP82C2*, first row). See Supplementary Table 6 for more
414 information. Data represents four biological replicates. (**e**) ChIP-PCR analysis of W-box-
415 containing regions (W) within *EPL1* and *EPL2* in *wrky33/DEX:WRKY33-flag* plants co-
416 treated with 20 μM dex (D) or mock solution (M) and *Psta* for 9 hr. Data represent mean
417 ± SE of four replicates. Dashed line represents the 5-fold cutoff between weak and
418 strong TF-DNA interactions. For (**d-e**), source data are provided as a Source Data file.
419

420 *EPCOT3* contains a 3' poly-A tail and is flanked by variable-length target site
421 duplications (Fig. 5c; Supplementary Fig. 7a), which are hallmarks of eukaryotic LINE
422 retrotransposons (Malik *et al*., 1999). LINE retrotransposition (reverse transcription and
423 integration) results in frequent 5'-truncation of retrocopies (Luan *et al*., 1993). We
424 identified eleven variably truncated retrocopies similar to *EPCOT3* throughout the
425 genome, including *Ta22*, one of the first LINEs characterized in *A. thaliana* (Fig. 5c;
426 Supplementary Fig. 7a-b, Supplementary Table 3; Wright *et al*., 1996). *EPCOT3*-related
427 LINEs were sorted into two groups roughly correspondent to their phylogenetic
428 placement: *EPCOT3-LIKE* (*EPL*) for those with high identity (>65%) to *EPCOT3* and
429 *Ta22* or *Ta22-LIKE* (*Ta22L*) for the remainder (Supplementary Fig. 7a; Supplementary
430 Table 3). Only *Ta22* and *Ta22L1* are full-length LINEs (Fig. 5c), presumably encoding
431 the proteins necessary for their own transposition and for the transposition of
432 nonautonomous family members like *EPCOT3*. We also identified two syntenic species-
433 specific *Ta22L*s, but no *EPLs*, in *A. lyrata* (Supplementary Table 3). To confirm the
434 involvement of *EPCOT3* in species-specific expression of *CYP82C2*, we introduced
435 WRKY33 into *N. benthamiana* leaves and *A. thaliana cyp82C2* protoplasts transfected
436 with either the *A. thaliana* and *A. lyrata CYP82C2* loci (including genes plus 3000 bp
437 upstream sequence, Fig 5d). We observed transactivation by WRKY33 of the *A.*
438 *thaliana* sequence, but not that of *A. lyrata* (Fig. 5d; Supplementary Fig. 7d). Altogether,
439 these data indicate that *EPCOT3* and *EPLs* arose from retrotransposition following the
440 speciation of *A. thaliana* and that the *EPCOT3*-containing *A. thaliana CYP82C2*
441 promoter is sufficient to confer WRKY33-mediated transcription of *CYP82C2*.

442

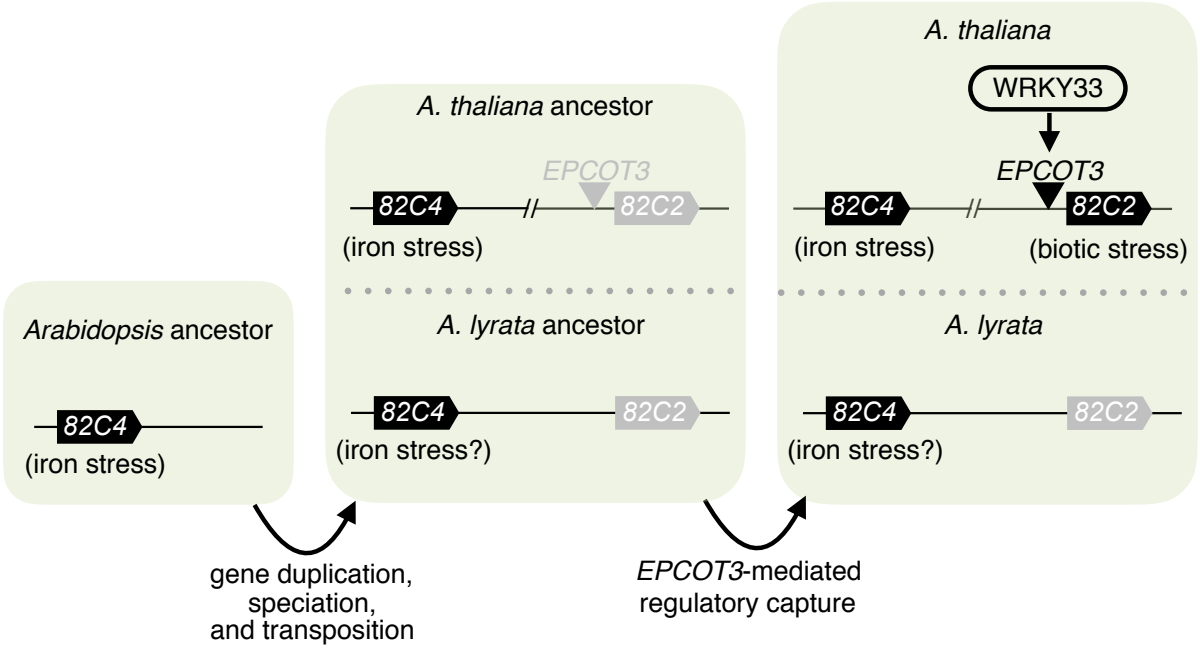
443 Of the *EPL* retrocopies, *EPL1* is most similar to *EPCOT3* (85.4% identity), sharing the
444 W-box and WRKY33-specific motif, whereas *EPL2* is less similar (67%) and lacks the
445 WRKY33-specific motif (Fig. 5c; Supplementary Table 3, Supplementary Fig. 7a). *EPL1*
446 and *EPL2* are much less truncated than *EPCOT3* (Fig. 5c), and lack epigenetic
447 signatures typical of *cis*-regulatory sequences (Supplementary Fig. 7c) (Roudier *et al*.,
448 2011; Liu *et al*., 2018). To investigate whether the sequence information and chromatin
449 features associated with *EPL*s are sufficient for WRKY33-binding, we tested for
450 WRKY33-binding to *EPL* sequences homologous to the W4 region of *EPCOT3* in dex-
451 treated, *Psta*-infected *wrky33/DEX:WRKY33-flag* plants by ChIP-(q)PCR. Compared to
452 *EPCOT3* (Fig. 3c), WRKY33 respectively bound weakly or not at all to *EPL1* and *EPL2*
453 (Fig. 5e; Supplementary Fig. 7e). Our findings suggest the following history: (1) *EPL1*
454 likely retroduplicated from *EPL2* or its progenitor, which already contained a W-box; (2)
455 *EPL1* then acquired a WRKY33-specific motif by mutation; (3) *EPCOT3* retroduplicated
456 from *EPL1* and then acquired epigenetic signatures of an enhancer, thereby allowing
457 selection to act on standing variation rather than *de novo* mutation for *CYP82C2*
458 recruitment into the 4OH-ICN biosynthetic pathway.

459

460 **Discussion**

461 TEs were originally conceived to act as “controlling elements” of several loci in the
462 genome (McClintock, 1956), and exaptation of TEs into *cis*-regulatory modules has
463 been hypothesized to be responsible for the rapid transcriptional rewiring in more

464 ancient lineages of vertebrates (Feschotte 2008; Bourque 2009; de Souza et al., 2013).
465 However, few (if any) evolutionarily recent TE exaptation events in vertebrates and
466 higher plants have been demonstrated to have biochemical, regulatory, physiological
467 and fitness-promoting functions (de Souza et al., 2013). With well over a dozen
468 genomes available including the genetic model *A. thaliana*, the mustard family presents
469 an excellent system for examining such events. In this study, we show that *EPCOT3* is
470 a TE-derived enhancer that mediates WRKY33-binding, pathogen-responsive
471 transcription of *CYP82C2,* synthesis of the species-specific metabolite 4OH-ICN, and
472 pathogen defense (Fig. 6). These results provide the first instance of a recent TE
473 exaptation responsible for the rewiring of a new gene into an ancient regulon, ultimately
474 leading to a positive effect on fitness.
475



476

477 **Figure 6. Model of regulatory neofunctionalization of *CYP82C2*.**

478 Although the *EPL1/EPCOT3* progenitor retrotransposed a preferred WRKY33-TFBS in
479 the form of *EPCOT3* upstream of *CYP82C2*, a further series of epigenetic modifications
480 were needed to facilitate optimal access of *EPCOT3* by WRKY33 (Fig. 6). *EPL1* exists
481 in a silenced heterochromatin state (Supplementary Fig. 7c), typical for TEs (Slotkin &
482 Martienssen, 2007), and is bound weakly by WRKY33 (Fig. 5e), whereas *EPCOT3* is in
483 an open chromatin state (Fig. 5b; Roudier *et al.*, 2011; Liu *et al.*, 2018) and bound
484 strongly by WRKY33 (Fig. 3c). The more severe 5'-truncation of *EPCOT3* could account
485 for its release from TE silencing mechanisms, and the initially weak WRKY33-binding
486 could provide a 'seed' for chromatin remodelers to drive the exaptation of newly
487 retrotransposed *EPCOT3* into a *bona fide* enhancer. Further epigenomic sampling
488 within *Arabidopsis* is needed to better clarify the epigenetic transformations underlying
489 the *EPCOT3* exaptation event.
490
491 Compared to closely-related Landsberg accessions (Supplementary Fig. 3; Hardtke *et*
492 *al.*, 1996), Di-G synthesizes less camalexin and 4OH-ICN (Fig. 2b; Kagan &
493 Hammerschmidt, 2002), is more susceptible to a range of bacterial and fungal
494 pathogens (Fig. 2c) (Hugouvieux *et al.*, 1998; Kagan & Hammerschmidt, 2002;
495 Mukherjee *et al.*, 2009), and is more sensitive to the phytohormone ethylene (Chatfield
496 *et al.*, 2008). WRKY33 has been implicated in camalexin biosynthesis (Qiu *et al.*, 2008),
497 antifungal defense (Zheng *et al.*, 2006), and ethylene biosynthesis (Li *et al.*, 2012). We
498 identified WRKY33 as causal for some if not all of these phenotypes in Di-G. This is the
499 first report of WRKY33's involvement in antibacterial defense and is consistent with the

500 contribution of camalexin and 4OH-ICN towards antibacterial defense (Rajniak *et al.*,
501 2015).
502
503 WRKY33 is an ancient transcription factor responsible for many fitness-promoting traits
504 in plants, thus it is unexpected that an *A. thaliana* accession would have a naturally
505 occurring *wrky33* mutation (C536T transversion). Di-G is the sole member of 1,135
506 sequenced accessions to have a high-effect single nucleotide polymorphism (SNP) in
507 *WRKY33* (1001 Genomes Consortium, 2016). Di-G and L*er*-0 have long been models
508 for studies in mutagenesis (Rédei, 1962, Müller, 1966), and thus a possibility exists that
509 Di-G may have originated from an ethyl methanesulfonate (EMS) mutagenesis screen
510 of L*er*-0. Historical EMS mutagenesis experiments generated upwards of tens of
511 thousands of mutations per cell (Müller 1966; Rédei & Koncz, 1993; Camara *et al.*,
512 2000), well within the range of ~25,000 SNPs that are not concordant between Di-G and
513 L*er*-0 (Supplementary Fig. 2f). However, features of EMS mutations (i.e. transversion
514 mutations) or X-ray mutations (i.e. indels) are not enriched in the Di-G pseudogenome
515 relative to related pseudogenomes (Supplementary Table 4). These findings suggest
516 that the wrky33 Di-G mutation is naturally derived.
517
518 **Methods**
519 **Plant materials and growth conditions.** For qPCR and HPLC-DAD analyses, surface-
520 sterilized *Arabidopsis thaliana* seeds were sown in 12-well microtiter plates sealed with
521 Micropore tape (3M, St. Paul, MN), each well containing ~15 seeds and 1 mL filter-

522 sterilized 1X Murashige and Skoog (MS; Murashige & Skoog, 1962) media (pH 5.7–5.8)
523 (4.43 g/L MS basal medium with vitamins [Phytotechnology Laboratories, Shawnee
524 Missions, KS], 0.05% MES hydrate, 0.5% sucrose). Iron deficient media was made as
525 previously described by Rajniak et al (2018). For *Polyctenium fremontii*, surface-
526 sterilized seeds were sown on MS agar plates. On day 9, seedlings were transferred to
527 6-well microtiter plates, each well containing ~15 seeds and 3 mL MS media. For all
528 other species, surface-sterilized seeds were sown in 6-well plates, each well containing
529 ~15 seeds and 3 mL MS media. On day 9, media were refreshed prior to bacterial
530 elicitation. Microtiter plates were placed on grid-like shelves over water trays on a
531 Floralight cart (Toronto, Canada), and plants were grown at 21°C with 60% humidity
532 under a 16-hr light cycle (70-80 µE m-2 s-1 light intensity). For chromatin
533 immunoprecipitation analyses, approximately 200 seeds were sown in a 100mm x
534 15mm petri plate containing 20mL of 1X MS media. Media were exchanged for fresh
535 media on day 9. Microtiter plates were placed on grid-like shelves over water trays on a
536 Floralight cart (Toronto, Canada), and plants were grown at 21°C with 60% humidity
537 under a 16-hr light cycle (70-80 µE m-2 s-1 light intensity). For bacterial infection
538 assays, seedlings were transferred to and grown on soil [3:1 mix of Farfard Growing Mix
539 2 (Sun Gro Horticulture, Vancouver, Canada) to vermiculite (Scotts, Marysville, OH)] at
540 22˚C daytime/18˚C nighttime with 60% humidity under a 12-hr light cycle [50
541 (dawn/dusk) and 100 (midday) µE m-2 s-1 light intensity].
542 Seed stock information is shown in Supplementary Table 5**.**
543

544 **Vector construction and transformation.** To generate the *DEX:WRKY33-flag*
545 construct, *WRKY33* was PCR-amplified from genomic DNA using the primers
546 WRKY33gXhoF (5'-AACTCGAGAAGAACAAGAACCATCAC-3'), and W33flgSpeIR (5'-
547 CGACTAGTCTACTTGTCGTCATCGTCTTTGTAGTCGGGCATAAACGAATCGAAA-3')
548 and subcloned into the *Xho*I/*Spe*I sites of pTA7002 vector (Aoyama and Chua, 1997;
549 McNellis et al., 1998). To generate the *DEX:WRKY33-myc* construct, *WRKY33* was
550 PCR-amplified using the primers WRKY33gXhoF and WRKY33gStuR (5'-
551 AAGGCCTGGCATAAACGAATCGAAAAATG-3') and subcloned into the *Xho*I/*Stu*I sites
552 of a version of pTA7002 modified to contain 6 tandem copies of the c-Myc epitope
553 downstream of the StuI site (Chezem *et al.*, 2017). The constructs were introduced into
554 *Arabidopsis thaliana wrky33* plants via *Agrobacterium*-mediated floral dip method
555 (Clough and Bent, 1998), and transformants were selected on agar media containing 15
556 µg/mL hygromycin B (Invitrogen, Carlsbad, CA).
557
558 To generate the *CYP82C2pro:CYP82C2* DNA construct, the *CYP82C2* promoter (pro)
559 and coding sequences were PCR-amplified from *A. thaliana* genomic DNA using the
560 primers At82C2proXbaF (5-GCTCTAGAAGCTTCCAATAAAACATTC-3') and
561 At82C2proBamR (5'-GCGGATCCAGTGGTTTGAGCGTGCAAA-3'), and
562 At82C2geneBamF (5-GCGGATCCATGGATACTTCCCTCTTTTC-3') and
563 At82C2geneSmaR (5'-
564 TTCCCGGGCTACTTGTCGTCATCGTCTTTGTAGTCCACATAAAGCCCTTCCTTAAG-
565 3'), and both sequences were subcloned into the *XbaI/SmaI* sites of pBI101 vector

566 (Jefferson *et al.,* 1987). To generate the *AlCYP82C2pro:AlCYP82C2* DNA construct, the
567 *AlCYP82C2* pro and coding sequences were PCR-amplified from *A.lyrata* genomic DNA
568 using the primers Al82C2proSalF (5'-CGGTCGACTATTCCAGGAGCATACAA-3') and
569 Al82C2proBglIIR (5'-GGAGATCTAATGTTTTAAAAGTGCAAAAGAG-3'), and
570 Al82C2geneBamHF (5'-GCGGATCCATGGATACATCCCTCTTTTC-3') and
571 Al82C2geneSmaR (5'-TTCCCGGGctacttgtcgtcatcgtctttgtagtccacaaaaagttcttccttaagac-
572 3'), and subcloned into the *SalI/SmaI* sites of pBI101 vector. Transient expression of
573 constructs in *Nicotiana benthamiana* leaves was performed as previously described
574 (Rajniak *et al.,* 2015) with the following modification: leaves were infiltrated with
575 transformed *Agrobacterium* strains that were grown in LB medium supplemented with
576 30 μg/mL gentamycin and 50 μg/mL kanamycin to an OD_{600} of 0.7. 16 h post-*Agro*-
577 infiltration, leaves were sprayed with 20 μM dexamethasone and 0.1% Tween-20 to
578 induce transgene expression for 0.5, 15, 24 and 30 h. Transient expression of
579 constructs in *Arabidopsis thaliana* T-DNA insertion mutant *cyp82C2-2* (Arabidopsis
580 Biological Resource Center; GABI_261D12) was performed as previously described
581 (Sheen, 2002) with the following modification: 2.5 x 10^5 protoplasts were transfected
582 with 3 μg of construct for 20 min, recovered in 2.5x volume of W5 solution and elicited
583 with 1 μM flg22 (QRLSTGSRINSAKDDAAGLQIA; Genscript, Nanjing, China) in 1 mL
584 W5 solution for 6 h.
585
586 **Bacterial infection and MAMP elicitation.** A single colony of *Pseudomonas syringae*
587 pv. *maculicola* (*Pma*) M2 (containing *avrRpm1*, but not *avrRps4* or *avrRpt2*), *Pma*

588 ES4326 (containing no aforementioned effectors), *Pma* ES4326 *avrRpt2*, *Pseudomonas*
589 *syringae* pv. *tomato* DC3000 (*Pto* DC3000 or *Pst,* containing no aforementioned
590 effectors), *Pst avrRpm1*, *Pst avrRps4*, and *Pst avrRpt2* from a freshly streaked 3-day-
591 old agar plate were used to inoculate 2 mL of LB containing appropriate antibiotics.
592 Strains were grown to log phase, washed in sterile water twice, resuspended in water to
593 OD_{600} of 0.2, and incubated at room temperature with no agitation for 3-6 and prior to
594 infection. Chitosan (90% deacetylated chitin; Spectrum Chemical Mfg Corp, New
595 Brunswich, NJ) was prepared in 0.1 N acetic acid and neutralized with 0.1 N NaOH to
596 pH 5.8 to a stock concentration of 1.2 mg/mL.
597
598 Hydroponically grown 9-day-old seedlings were inoculated with bacterial strains to
599 OD_{600} of 0.013 or treated with 10 µM flg22, 10 µM elf26 (ac-
600 SKEKFERTKPHVNVGTIGHVDHGKTT; Genscript), and 150 or 300 µg/mL chitosan.
601
602 For qPCR analyses, seedlings were snap-frozen in liquid nitrogen 12 hr post-infection.
603 For HPLC-DAD analyses, seedlings were snap-frozen 24 to 28 hr post-infection. For
604 ChIP analyses, seedlings were snap-frozen 9 hr post-infection.
605
606 4-to-5-week-old adult leaves were treated with 0.0125% Silwet or 0.0125% Silwet and
607 20 µM dexamethasone for 20 sec and incubated on 0.8% (w/v) tissue-culture agar
608 plates on a light cart at 21˚C for 6-8 hr. Leaves were then surface-inoculated with *Pto*
609 DC3000 (OD_{600} = 0.002 or 10^6 colony-forming units (cfu)/cm^2 leaf area) in the presence

610 of 0.01% (v/v) Silwet L-77 (Phytotechnology Laboratories) for 15 sec and incubated on
611 0.8% (w/v) tissue-culture agar plates at 21°C under a 16-hr light cycle (70-80 µE m-2 s-
612 1 light intensity) for 3 days. Leaves were then surface-sterilized in 70% ethanol for 10
613 sec, rinsed in sterile water, surface-dried on paper towels, and the bacteria were
614 extracted into water, using an 8-mm stainless steel bead and a ball mill (20 Hz for 3
615 min). Serial dilutions of the extracted bacteria were plated on LB agar plates for colony-
616 forming units (CFU) counting.
617
618 **RNA isolation and quantitative PCR (qPCR).** Total RNA was extracted from 9-day-old
619 seedlings using TRIzol reagent (Invitrogen, San Diego, CA) according to the
620 manufacturer's instructions. 2.5 µg of total RNA was reverse-transcribed with 3.75 µM
621 random hexamers (Qiagen, Hilden, Germany) and 20 U of ProtoScript II (New England
622 Biolabs, Boston, MA) according to the manufacturer's instructions. The resulting
623 cDNA:RNA hybrids were treated with 10 U of DNase I (Roche, Basel, Switzerland) for
624 30 min at 37°C and purified on PCR clean-up column (Macherey-Nagel, Düren,
625 Germany). qPCR was performed with Kapa SYBR Fast qPCR master mix (Kapa
626 Biosystems, Wilmington, MA) and CFX384 real-time PCR machine (Bio-Rad, Hercules,
627 CA). The thermal cycling program was as follows: 95°C for 3 min; 45 cycles of 95°C for
628 15 sec and 53°C for 30 sec; a cycle of 95°C for 1 min, 53°C for 1 min, and 70°C for 10
629 sec; and 50 cycles of 0.5°C increments for 10 sec. Biological and technical replicates
630 were performed on the same 384-well PCR plate. Average of the three Ct values per
631 biological replicate was converted to difference in Ct value relative to that of control

632 sample. The Pfaffl method (Pfaffl, 2001) and calculated primer efficiencies were used to
633 determine the relative fold increase of the target gene transcript over the *EIF4A1*
634 (*AT3G13920* or *AL3G26100*) housekeeping gene transcript for each biological replicate.
635 Expression values were then calculated relative to WT un-treated samples. Primer
636 sequences and efficiencies are listed in Supplementary Table 6**.**
637
638 **RT-PCR Analysis.** Total RNA was extracted from snap-frozen tobacco leaves or
639 Arabidopsis protoplasts as described by Couto et al (2015) and Oñate-Sánchez and
640 Vicente-Carbajosa (2008). Total RNA (2 *µ*g) was treated with 10 units of DNase I
641 (Roche) for 30 min at 37˚C and then for 15 min at 70˚C, reverse-transcribed with 3.75
642 *µ*M random hexamers (Qiagen) and 20 units of M-MuLV (New England Biolabs) for 1 h
643 at 42˚C and then for 15 min at 70˚C. Tobacco cDNA was diluted 7.5-fold. 4 *µ*L of cDNA
644 was used in 20 *µ*L PCR reactions, and resulting PCR products were separated on 2%
645 agarose gels. PCR was performed on C1000 thermal cycler (Bio-Rad) with the following
646 thermal cycling program: 95˚C for 3 min; 40 cycles of 95˚C for 10 sec, 53˚C for 15 sec,
647 and 72˚C for 7 sec (*AtWRKY33*, *CYP82C2*, *AlCYP82C2*), 15 sec (*NbACTIN1*) or 21 sec
648 (*CYP82C2*-long). Primer sequences are listed in Supplementary Table 6.
649
650 **Camalexin and 4OH-ICN extraction and LC-DAD-MS.** 10-day-old seedlings were
651 snap-frozen, lyophilized, weighed and homogenized using a 5-mm stainless steel bead
652 and ball mill (20 Hz, 4 min). For phytoalexin analysis, homogenate was extracted with
653 300 µL 80% (v/v) aqueous methanol containing 0.08% (v/v) formate and 2.5 µL internal

654 standard (200 µM 4-methoxyindole/4M-I [Sigma-Aldrich] in 100% methanol) per mg
655 sample dry weight. Extracts were sonicated for 5 min and centrifuged at 16,000xg for 2
656 min. The supernatant was filtered using a 0.45-µm polypropylene filter plate (GE
657 Healthcare, Chicago, IL). Samples were separated by reversed-phase chromatography
658 on an Ultimate 3000 HPLC (Dionex, Sunnyvale, CA) system, using a 3.5-µm, 3x150-
659 mm Zorbax SB-Aq column (Agilent, Santa Clara, CA); volume injected was 10 µL. The
660 gradient is shown in Supplementary Table 7. A coupled DAD-3000RS diode array
661 detector (Dionex) collected UV absorption spectra in the range of 190-560 nm, a FLD-
662 311 fluorescence detector (Dionex) collected fluorescence data at 275 nm excitation
663 and 350 nm emission, and a MSQPlus mass spectrometer (Dionex) collected ESI mass
664 spectra in positive and negative ion modes in the range of 100-1000 m/z. Total ICN,
665 4OH-ICN and camalexin amounts were quantified using standard curves of standards
666 prepared in *cyp79B2 cyp79B3* seedling extract and integrated areas in the UV
667 chromatographs at 260-nm for 4M-I (retention time [RT] = 7.7 min); 340-nm for ICN (RT
668 = 11.5 min); 280-nm for ICN degradation product ICA-ME (RT = 9.5 min); and co-eluting
669 4OH-ICN degradation products 4OH-ICA and 4OH-ICA-ME (RT = 10.1 min); and 320
670 nm for camalexin (RT = 12.1 min). For Figure 1b, total camalexin amounts were
671 quantified using integrated areas in the FLD chromatograph. For some experiments, 2.5
672 uL 200 µM indole butrytic acid (IBA; RT = 10.1 min) was added per mg sample dry
673 weight instead of 4M-I. Relative amounts of ICN, 4OH-ICN, and amounts were
674 quantified by dividing the peak areas at m/z 169 [M-H]- (ICN), 174 [M-H]- (ICA-ME), 176

675 [M-H]- (4OH-ICA), 190 [M-H]- (4OH-ICME), and 201 [M+H]+ (camalexin), by that of IBA
676 (m/z 202 [M-H]-).
677
678 **Glucosinolate extraction and LC-DAD-FLD-MS.** Glucosinolates were analyzed as
679 desulfoglucosinolates as previously described by Kliebenstein *et al*. (2001) with some
680 modifications. Briefly, a 96-well 0.45 µm PVDF filter plate (EMD Millipore, Billerica, MA)
681 was charged with 45 mg DEAE Sephadex A25 (GE Heathcare) and 300 µL of water per
682 well and equilibrated at room temp for 2 h. Prior to sample homogenization, the plate
683 was centrifuged at 400xg for 1 min to remove the water. The homogenate was extracted
684 with 500 µL 70% (v/v) aqueous methanol at 67.5˚C for 10 min and centrifuged at
685 16,000xg for 2 min. Added to the supernatant was 3 µL of IS (1.25 mM sinigrin (Sigma-
686 Aldrich) in 80% (v/v) ethanol) per mg sample dry weight. Extract was applied to and
687 incubated on the ion exchanger for 10 min. The sephadex resin was washed three
688 times with 70% (v/v) methanol, three times with distilled deionized water (ddH_2O), and
689 two times with 20 mM sodium acetate (pH 5). 20µL of 25 mg/mL aryl sulfatase (Type H1
690 from *Helix pomatia*, Sigma-Aldrich) was applied to and incubated on the sephadex resin
691 at RT overnight (Hogge *et al*., 1988). The plate was centrifuged at 400xg for 1 min, and
692 desulfoglucosinolates were eluted from the sephadex resin by two 100-µL washes with
693 60% (v/v) methanol and two 100-µL washes with ddH_2O. Eluate volume was reduced to
694 250-350 µL using an evaporator. Samples were separated using the gradient shown in
695 Supplementary Table 7**.** A coupled DAD-3000RS diode array detector, FLD-311
696 fluorescence detector (Dionex), and MSQPlus mass spectrometer collected UV

697 absorption spectra at 229-nm, fluorescence spectra at 275/350-nm (ex/em), and ESI
698 mass spectra in positive/negative ion modes at 100-1000 m/z, respectively.
699 Glucosinolates were quantified using integrated areas of desulfoglucosinolates in the
700 UV chromatographs at 229-nm and published response factors (Clarke, 2010).
701
702 **Chromatin immunoprecipitation and (q)PCR.** ChIP was performed as previously
703 described by Chezem et al. (2017) with some modifications. Approximately two-
704 hundred-and-ten 9-day-old seedlings were inoculated with *Pto* DC3000 *avrRpm1* to
705 OD_{600} of 0.013 and co-treated with mock solution of DMSO (M) or 20 μM
706 dexamethasone (D) for 9 hr. Following nuclear extraction, samples were sonicated in a
707 Covaris S2 sonicator (Covaris, Woburn, MA) using 10% duty, 7% intensity, 200 cycles
708 per burst for a total time of 11 min. Chromatin immunoprecipitation was performed using
709 Anti-FLAG M2 Affinity Gel (Sigma-Aldrich). Beads were pre-treated with 0.1% (w/v) non-
710 fat milk in 1X PBS and 0.5 mg/mL sheared salmon sperm DNA (Invitrogen). Following
711 de-crosslinking, DNA samples were phenol-chloroform-extracted, diluted to the same
712 OD_{260} concentration, and 1.5 μL was used in 15 μL PCR reaction. PCR analysis was
713 performed on nuclear extracts prior to antibody incubation (input) and after ChIP. PCR
714 conditions were as follows: 95°C for 3 min; 40 cycles of 95°C for 15 sec, 53°C for 15
715 sec, and 72°C for 1 min; 72°C for 5 min. Densitometric determination of signal intensity
716 in each ChIP and input sample was calculated using ImageJ. Fold enrichment was
717 determined by calculating the ratio of PCR product intensities in ChIP D/M to Input D/M.
718 In cases where amplicons were absent, an arbitrary value of 10 was assigned. For

719 *EPL2*, qPCR analysis was additionally performed to confirm absence of amplicons in
720 ChIP samples. RLU counts at the 25th cycle were used for quantification. Primer
721 sequences are listed in Supplementary Table 6**.**
722
723 **Comparative genomics**. All phylogenetic species trees were adapted from Koch and
724 Kiefer (2005) and Couvreur et al. (2009). To generate novel phylogenetic maximum
725 likelihood (ML) trees, sequences were aligned using MUSCLE in MEGA7 (Kumar *et al*.,
726 2016) and JTT model (for CYP82C and LINE alignments) or Tamura-Nei model (for the
727 *EPCOT3* alignment). Sequences for all genes with the description "non-LTR
728 retrotransposon family (LINE)" (N=263) were batch-downloaded from TAIR
729 (https://arabidopsis.org). Of these, sequences containing intact reverse transcriptase
730 domains ("PGPDG", "LIPK", "FRPISL", or "FADD" sequences; N=126) were used for
731 subsequent phylogenetic analysis. Gaps were removed from the CYP82C alignment,
732 leaving a total of 480 codons. *EPCOT3* alignments were visualized in JalView
733 (http://www.jalview.org/; Waterhouse *et al*., 2009). Information on genomes used for
734 synteny analysis is shown in Supplementary Table 8.
735
736 Selection estimates based on nonsynonymous-to-synonymous substitution ratios were
737 calculated from the CYP82C ML tree (Supplementary Text 1). A Newick tree file was
738 generated from this ML tree (Supplementary Figure 4b; Supplementary Table 2) and for
739 Branch site models, branches were pre-defined. CodeML analysis in PAML (Yang,
740 2007) was then conducted with the following modified parameters: ncatG = 8;

741 CodonFreq = 3. The M0 test was performed with model = 0 and NSsites = 0. The M1a
742 null test was performed with model = 0 and NSsites = 1. A more stringent null test (fixed
743 omega) was performed for each Branch site model to be tested (model = 2 and NSsites
744 = 2), where omega was fixed to 1. Branch site models were then tested with unfixed
745 omega. Likelihood ratio tests were performed by comparing critical values and degrees
746 of freedom between each unfixed Branch site test and either the M1a test or the
747 corresponding fixed-omega test. Pre-defined branches with *P* values less than 0.05 for
748 both tests were regarded as under positive selection (Supplementary Figure 2).
749
750 The protein structure of CYP82C2 was generated using Intensive modeling mode in
751 Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index; Kelley *et al.*, 2015)
752 and visualized in MacPyMOL (Schrödinger, LLC). Amino acid conservation was scored
753 using the Bayesian Best model in Consurf (http://consurf.tau.ac.il/2016/; Ashkenazy *et*
754 *al.*, 2016).
755
756 **Bioinformatics.** Coexpression data was obtained from ATTED-ii (http://atted.jp/;
757 Obayashi et al., 2018). Mutual ranks less than 200 are indicative of strong co-
758 expression (Obayashi et al., 2018). Epigenetics data was obtained from Roudier et al.
759 (2011) and confirmed using data from Liu et al. (2018). Percent identity matrices were
760 constructed from Clustal Omega Multiple Sequence Alignments
761 (https://www.ebi.ac.uk/Tools/msa/clustalo/). Promoter alignment plots were generated
762 using mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml; Frazer et al., 2004)

763

764 **Data and code availability**

765 The authors declare that all data supporting the findings of this study are available
766 within the manuscript and the Supplementary Information or are available from the
767 corresponding authors upon request. Custom code to visualize homology models in
768 MacPyMOL and conduct statistics in R are also available upon request.

769

770

771 **References**

772 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism
773 in *Arabidopsis thaliana*. *Cell* **166**(2), 481-491 (2016).

774 Aoyama, T. & Chua, N.H. A glucocorticoid-mediated transcriptional induction system
775 in transgenic plants. *Plant J* **11**(3), 605–612 (1997).

776 Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and
777 visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*
778 **44**(W1), W344-W350 (2016).

779 Bak 2001

780 Birkenbihl, R.P., Diezel, C., Somssich, I.E. Arabidopsis WRKY33 is a key transcriptional
781 regulator of hormonal and metabolic responses toward *Botrytis cinerea* infection.
782 *Plant Physiol* **159**(1), 266-285 (2012).

783 Birkenbihl, R.P., Kracher, B., Somssich, I.E. Induced Genome-Wide Binding of Three
784 Arabidopsis WRKY Transcription Factors during Early MAMP-Triggered Immunity.
785 *Plant Cell* **29**(1), 20-38 (2017).
786 Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., Innes, R.W. A disease
787 resistance gene in Arabidopsis with specificity for two different pathogen avirulence
788 genes. *Plant Cell* **6**(7), 927-933 (1994).
789 Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures
790 of enhancer activity during embryonic development. *Nat Genet* **44**(2), 148-156
791 (2012).
792 Bourque, G. Transposable elements in gene regulation and in the evolution of
793 vertebrate genomes. *Curr Opin Genet Dev* **19**(6), 607-612.
794 Böttcher, C. *et al.* The multifunctional enzyme CYP71B15 (PHYTOALEXIN
795 DEFICIENT3) converts cysteine-indole-3-acetonitrile to camalexin in the indole-3-
796 acetonitrile metabolic network of Arabidopsis thaliana. *Plant Cell* **21**(6), 1830-1845
797 (2009).
798 Cabrera, J.C., Messiaen, J., Cambier, P., Van Cutsem, P. Size, acetylation and
799 concentration of chitooligosaccharide elicitors determine the switch from defence
800 involving PAL activation to cell death and water peroxide production in Arabidopsis
801 cell suspensions. *Physiol Plant* **127**(1), 44-56 (2006).
802 Camara, M.D., Ancell, C.A., Pigliucci, M. Induced mutations: a novel tool to study
803 phenotypic integration and evolutionary constraints in *Arabidopsis thaliana*. *Evol*
804 *Ecol Res* 2(8), 1009-1029 (2000).

805 Chae, L., Kim, T., Nilo-Poyanco, R., Rhee, S.Y. Genomic signatures of specialized
806 metabolism in plants. *Science* **344**(6183), 510-513 (2014).
807 Chatfield, S.P. & Raizada, M.M. Ethylene and shoot regeneration: hookless1 modulates
808 de novo shoot organogenesis in *Arabidopsis thaliana*. *Plant Cell Rep* **27**(4), 655-666
809 (2008).
810 Chan, Y.F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent
811 deletion of a *Pitx1* enhancer. *Science* **237**(5963), 302-305 (2010).
812 Chuong, E.B., Elde, N.C., Feschotte, C. Regulatory evolution of innate immunity
813 through co-option of endogenous retroviruses. *Science* **351**(6277), 1083-1087
814 (2016).
815 Clarke, D.B. Glucosinolates, structures and analysis in food. *Anal Methods* **2**(4), 301-
816 416 (2010).
817 Clay, N.K., *et al.* Glucosinolate metabolites required for an Arabidopsis innate immune
818 response. *Science* **323**(5910), 95-101 (2009).
819 Clough, S.J. & Bent, A.F. Floral dip: a simplified method for Agrobacterium-mediated
820 transformation of *Arabidopsis thaliana*. *Plant J* **16**(6), 735-743 (1998).
821 Couto D., Stransfeld L., Arruabarrena A., Zipfel C., Lozano-Durán R. Broad
822 application of a simple and affordable protocol for isolating plant RNA. *BMC Res*
823 *Notes* **8**, 154 (2015).
824 Couvreur, T.L.P. *et al.* Molecular phylogenetics, temporal diversification, and
825 principles of evolution in the mustard family (Brassicaceae) *Mol Biol Evol* **27**(1),
826 55-71 (2009).

827 de Souza, F.S.J., Franchini, L.F., Rubinstein, M. Exaptation of transposable elements
828 into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol*
829 **30**(6), 1239-1251 (2013).
830 Debener T., Lehnackers H., Arnold M., Dangl J.L. Identification and molecular mapping
831 of a single Arabidopsis thaliana locus determining resistance to a phytopathogenic
832 Pseudomonas syringae isolate. *Plant J* **1**(3), 289-302 (1991).
833 Dixon, R.A. & Strack. D. Phytochemistry meets genome analysis and beyond.
834 *Phytochemistry* **62**(6), 815-816 (2003).
835 Denoux, C. *et al.* Activation of defense response pathways by OGs and Flg22 elicitors
836 in Arabidopsis seedlings. *Mol Plant* **1(**3), 423-445 (2008).
837 Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat Rev*
838 *Genet* **9**(5), 397-405 (2008).
839 Felix G., Duran J.D., Volko S., Boller T. Plants have a sensitive perception system for
840 the most conserved domain of bacterial flagellin. *Plant J* **18**(3),265-276 (1999).
841 Force, A.M. *et al.* Preservation of duplicate genes by complementary, degenerative
842 mutations. *Genet* **151**(4), 1531-1545 (1999).
843 Frazer, K.A. *et al.* VISTA: computational tools for comparative genomics. *Nucleic*
844 *Acids Res* **32**(Web Server issue), W273-W279 (2004).
845 Glawischnig, E., Hansen, B.G., Olsen, C.E., Halkier, B.A. Camalexin is synthesized
846 from indole-3-acetaldoxime, a key branching point between primary and secondary
847 metabolism in *Arabidopsis*. *Proc Natl Acad Sci USA* **101**(21), 8245-8250 (2004).

848 Grotewold, E. Plant metabolic diversity: a regulatory perspective. *Trends Plant Sci*
849 **10**(2), 57-62 (2005).
850 Hammerschmidt, R. PHYTOALEXINS: What have we learned after 60 years? *Annu Rev*
851 *Phytopathol* **37**, 285-306 (1999).
852 Hardtke, C.S., Müller, J., Berleth, T. Genetic similarity among *Arabidopsis thaliana*
853 ecotypes estimated by DNA sequence comparison. *Plant Mol Biol* 32(5), 915-922
854 (1996).
855 Hartmann, T. From waste products to ecochemicals: fifty years research of plant
856 secondary metabolism. *Phytochemistry* **68**(22-24), 2831-2846 (2007).
857 Heintzman, N.D. *et al.* (2007) Distinct and predictive chromatin signatures of
858 transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**(3),
859 311-318 (2007).
860 Hénaff, E., Vives, C., Desvoyes, B., Chaurasia, A., Payet, J., Gutierrez, C.,
861 Casacuberta, J.M. Extensive amplification of the E2F transcription factor binding
862 sites by transposons during evolution of Brassica species. *Plant J* **77**(6), 852-862
863 (2014).
864 Hoffman, B.G. et al. Locus co-occupancy, nucleosome positioning, and H3K4me1
865 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and
866 liver. *Genome Res* **20**(8), 1037-1061 (2010).
867 Hogge, L.R., Reed, D.W., Underhill, E.W., Haughn, G.W. HPLC separation of
868 glucosinolates from leaves and seeds of *Arabidopsis thaliana* and their
869 identification using thermospray liquid chromatography-mass spectrometry. *J*

870 *Chromatogr Sci* **26**, 551-556 (1988).

871 Hohmann, N., Wolf, E.M., Lysak, M.A., Koch, M.A. A time-calibrated road map of
872 Brassicaceae species radiation and evolutionary history. *Plant Cell* **27**(10), 2770-
873 2784 (2015).

874 Hu, T.T. et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome
875 size change. *Nat Genet* **43**(5), 476-481 (2011).

876 Hugouvieux, V., Barber, C.E., Daniels, M.M. Entry of *Xanthomonas campestris* pv.
877 *campestris* into hydathodes of *Arabidopsis thaliana* leaves: a system for studying
878 early infection events in bacterial pathogenesis. *Mol Plant-Microbe Interact* **11**(6),
879 537-543 (1998).

880 Hull, A.K., Vij, R., Celenza, J.L. *Arabidopsis* cytochrome P450s that catalyze the first
881 step of tryptophan-dependent indole-3-acetic acid biosynthesis. *Proc Natl Acad Sci*
882 *USA* **97**(5), 2379-2384 (2000).

883 Jefferson, R.A., Kavanagh, T.A., Bevan, M.W. GUS fusions: beta-glucuronidase as a
884 sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901-3907
885 (1987).

886 Jones, J.D.G. & Dangl, J.L. The plant immune system. *Nature* **444**(7117), 323-329
887 (2006).

888 Kagan, I.A. & Hammerschmidt, R. Arabidopsis ecotype variability in camalexin
889 production and reaction to infection by *Alternaria brassicicola*. *J Chem Ecol* **28**(11),
890 2121-2140 (2002).

891 Kelley, L.A. *et al.* The Phyre2 web portal for protein modeling, prediction and

892 analysis. *Nat Protoc* **10**(6), 845-858 (2015).

893 Klein, A.P., Anarat-Cappillino, G., Sattely, E.S. Minimum set of cytochromes P450 for
894 reconstituting the biosynthesis of camalexin, a major Arabidopsis antibiotic. *Angew*
895 *Chem Int Ed Engl* **52**(51), 13625-13628 (2013).

896 Kliebenstein, D.J. *et al.* Genetic control of natural variation in Arabidopsis
897 glucosinolate accumulation. *Plant Physiol* **126**(2), 811-825 (2001).

898 Koch, M.A. & Kiefer, M. Genome evolution among cruciferous plants: a lecture from
899 the comparison of the genetic maps of three diploid species—*Capsella rubella,*
900 *Arabidopsis lyrata* subsp. *petraea,* and *A. thaliana*. *Am J Bot* **92**(4), 761-767
901 (2005).

902 Kover, P.X. & Schaal, B.A. Genetic variation for disease resistance and tolerance
903 among Arabidopsis thaliana accessions. *Proc Natl Acad Sci USA* **99**(17), 11270-
904 11274 (2002).

905 Kruse, T. *et al.* In planta biocatalysis screen of P450s identifies 8-methoxypsoralen as a
906 substrate for the CYP82C subfamily yielding original chemical structures. *Chem Biol*
907 **15**(2), 149-156 (2008).

908 Kumar, S., Stecher, G., Tamura, K. MEGA7: molecular evolutionary genetics
909 analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**(7), 1870-1874 (2016).

910 Levine, M. & Davidson, E.H. Gene regulatory networks for development. *Proc Natl Acad*
911 *Sci USA* **102**(14), 4936-4942 (2005).

912 Li, G. et al. Dual-level regulation of ACC synthase activity by MPK3/MPK6 cascade and
913 its downstream WRKY transcription factor during ethylene induction in Arabidopsis.
914 *PLoS Genet* **8**(6), e1002767 (2012).
915 Liu, S. *et al.* Negative regulation of ABA signaling by WRKY33 is critical for Arabidopsis
916 immunity towards *Botrytis cinerea* 2100. *eLife* **4**, e07295 (2015).
917 Liu, Y. *et al.* PCSD: a plant chromatin state database. *Nucleic Acids Res* **46**(D1),
918 D1157-D1167 (2018).
919 Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H. Reverse transcription of
920 R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for
921 non-LTR retrotransposition. *Cell* **72**(4), 595-605 (1993).
922 Lynch, M. The lower bound to the evolution of mutation rates. *Genome Biol Evol* 3,
923 1107-1118 (2011).
924 Malik, H.S., Burke, W.D., Eickbush, T.H. The age and evolution of non-LTR
925 retrotransposable elements. *Mol Biol Evol* **16**(6), 793-805 (1999).
926 Mansfield, J.W. Antimicrobial compounds and resistance: the role of phytoalexins and
927 phytoanticipins. In Mechanisms of Resistance to Plant Diseases, A. Slusarenko,
928 RSS Fraser and LC. van Loon eds (Kluwer Academic Publishers Dordrecht, The
929 Netherlands), pp. 325-370 (2000).
930 Martin, C., Ellis, N., Rook, F. Do transcription factors play special roles in adaptive
931 variation? *Plant Physiol* **154**(2), 506-511 (2010).
932 McClintock, B. Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol*
933 **21**, 197-216 (1956).

934 McNellis, T.W. *et al.* Glucocorticoid-inducible expression of a bacterial avirulence
935 gene in transgenic Arabidopsis induces hypersensitive cell death. *Plant J* **14**(2),
936 247-257 (1998).
937 Mikkelsen, M.D., Hansen, C.H., Wittstock, U., Halkier, B.A. Cytochrome P450 CYP79B2
938 from *Arabidopsis* catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a
939 precursor of indole glucosinolates and indole-3-acetic acid. *J Biol Chem* **275**(43),
940 33712-33717 (2000).
941 Moore, R.C. & Purugganan, M.D. The evolutionary dynamics of plant duplicate genes.
942 *Curr Opin Plant Biol* **8**(2), 122-128 (2005).
943 Mukherjee, A.K., Lev, S., Gepstein, S., Horwitz, B.B. A compatible interaction of
944 *Alternaria brassicicola* with *Arabidopsis thaliana* ecotype DiG: evidence for a specific
945 transcriptional signature. *BMC Plant Biol* **9**(1), 31 (2009).
946 Müller, A.A. Die Induktion von rezessiven Letalmutationen durch Äthylmethansulfonat
947 beiArabidopsis. *Theor Appl Genet* **36**(5), 201-220 (1966).
948 Murashige, T. & Skoog, F. A revised medium for rapid growth and bio assays with
949 tobacco tissue cultures. *Physiol Plant* **15**(3), 473–497 (1962).
950 Murgia, I., Tarantino, D., Soave, C., Morandini, P. Arabidopsis *CYP82C4* expression is
951 dependent on Fe availability and circadian rhythm, and correlates with genes
952 involved in the early Fe deficiency response. *J Plant Physiol* **168**(9), 894-902 (2011).
953 Nafisi, M. *et al.* Arabidopsis cytochrome P450 monooxygenase 71A13 catalyzes the
954 conversion of indole-3-acetaldoxime in camalexin synthesis. *Plant Cell* **19**(6), 2039-
955 2052 (2007).

956 Navarro, L. *et al.* The transcriptional innate immune response to flg22. Interplay and
957 overlap with Avr gene-dependent defense responses and bacterial pathogenesis.
958 *Plant Physiol* **135**(2), 1113-1128 (2004).
959 Obayashi, T. *et al.* ATTED-II in 2018: A plant coexpression database based on
960 investigation of statistical property of Mutual Rank Index. *Plant Cell Physiol* **59**(1),
961 e3 (2018).
962 Ohno, S. Evolution by Gene Duplication. Springer-Verlag: Heidelberg, Germany (1970).
963 Omranian, N. et al. Differential metabolic and coexpression networks of plant
964 metabolism. *Trends Plant Sci* **20**(5), 266-268 (2015).
965 Oñate-Sánchez L. & Vicente-Carbajosa J. DNA-free RNA isolation protocols for
966 Arabidopsis thaliana, including seeds and siliques. *BMC Res Notes* **1**, 93 (2008).
967 Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-
968 PCR. *Nucleic Acids Res* **29**(9), e45 (2001).
969 Povero, G. *et al.* Transcript profiling of chitosan-treated Arabidopsis seedlings. *J Plant*
970 *Res* **124**(5), 619-629 (2011).
971 Prud'homme, B., Gompel, N., Carroll, S.B. Emerging principles of regulatory evolution.
972 *Proc Natl Acad Sci USA* **104**(Suppl 1), 8605-8612 (2007).
973 Qiu, J.L. *et al.* Arabidopsis MAP kinase 4 regulates gene expression through
974 transcription factor release in the nucleus. *EMBO J* **27(**16), 2214-2221 (2008).
975 Rajniak, J., Barco, B., Clay, N.K., Sattely, E.S. A new cyanogenic metabolite in
976 Arabidopsis required for inducible pathogen defence. *Nature* **525**(7569), 376-379
977 (2015).

978 Rajniak, J. *et al.* Biosynthesis of redox-active metabolites in response to iron deficiency
979 in plants. *Nat Chem Biol* **14**(5), 442-450 (2018).
980 Rédei, G.P. Single locus heterosis. *Zeitschrift für Vererbungslehre* **93**(1), 164-170
981 (1962).
982 Rédei, G.P. & Koncz, C. Classical mutagenesis. In Methods in Arabidopsis Research,
983 pp. 16-82 (1993).
984 Rinerson, C.I. *et al.* The evolution of WRKY transcription factors. *BMC Plant Biol* **15**, 66
985 (2015).
986 Rogers, W.A., *et al.* Recurrent modification of a conserved *cis*-regulatory element
987 underlies fruit fly pigmentation diversity. *PLoS Genet* **9**(8), e1003740 (2013).
988 Roudier, F. *et al.* Integrative epigenomic mapping defines four main chromatin states in
989 Arabidopsis. *EMBO J* **30**(10), 1928-1938 (2011).
990 Rushton, P.J., Somssich, I.E., Ringler, P., Shen, Q.J. WRKY transcription factors.
991 *Trends Plant Sci* **15**(5), 247-258 (2010).
992 Schluttenhofer, C. &, Yuan, L. (2015) Regulation of specialized metabolism by WRKY
993 transcription factors. *Plant Physiol* **167**(2), 295-306 (2015).
994 Sheen J. A transient expression assay using *Arabidopsis* mesophyll protoplasts.
995 http://genetics.mgh.harvard.edu/sheenweb/ (2002).
996 Slattery, M. *et al.* Absence of a simple code: how transcription factors read the genome.
997 *Trends Biochem Sci* **39**(9), 381-399 (2014).
998 Slotkin, R. K, & Martienssen, R. Transposable elements and the epigenetic regulation of
999 the genome. *Nat Rev Genet* **8**(4), 272 (2007).

1000 Spitz, F. & Furlong, E.E. Transcription factors: from enhancer binding to developmental
1001 control. *Nat Rev Genet* **13**(9), 613-626 (2012).
1002 Tao, Y. *et al.* Quantitative nature of Arabidopsis responses during compatible and
1003 incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant*
1004 *Cell* **15**(2), 317-330 (2003).
1005 Thomma, B.P., Nelissen, I., Eggermont, K., Broekaert, W.F. Deficiency in phytoalexin
1006 production causes enhanced susceptibility of *Arabidopsis thaliana* to the fungus
1007 *Alternaria brassicicola. Plant J* **19**(2), 163-171 (1999).
1008 Tohge, T. & Fernie, A.R. Co-expression and co-responses: within and beyond
1009 transcription. *Front Plant Sci* **3**, 248 (2012).
1010 Tsuji, J. et al. (1992) Phytoalexin accumulation in *Arabidopsis thaliana* during the
1011 hypersensitive reaction to *Pseudomonas syringae* pv *syringae*. *Plant Physiol* **98**(4),
1012 1304-1309 (1992).
1013 Wang, Y., Li, X., Hu, H. H3K4me2 reliably defines transcription factor binding regions in
1014 different cells. *Genomics* **103**(2), 222-228 (2014).
1015 Waterhouse, A.M. *et al.* Jalview Version 2—a multiple sequence alignment editor
1016 and analysis workbench. *Bioinformatics* **25**(9), 1189-1191 (2009).
1017 Weng, J.K., Philippe, R.N., Noel, J.P. (2012) The rise of chemodiversity in plants.
1018 *Science* **336**(6089), 1667-1670 (2012).
1019 Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements.
1020 *Nat Rev Genet* **8**(12), 973-982 (2007).

1021 Wink, M. Evolution of secondary metabolites from an ecological and molecular
1022 phylogenetic perspective. *Phytochemistry* **64**(1), 3-19 (2003).
1023 Wittkopp, P.J. & Kalay, G. *Cis*-regulatory elements: molecular mechanisms and
1024 evolutionary processes underlying divergence. *Nat Rev Genet* **13**(1), 59-69 (2012).
1025 Wray, G.A. The evolutionary significance of *cis*-regulatory mutations. *Nat Rev Genet*
1026 **8**(3), 206-216 (2007).
1027 Zheng, Z., Qamar, S.A., Chen, Z., Mengiste, T. (2006) Arabidopsis WRKY33
1028 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant*
1029 *J* **48** (4), 592-605 (2006).
1030 Zhou, J., Wang, J., Zheng, Z., Fan, B., Yu, J. Q., Chen, Z. Characterization of the
1031 promoter and extended C-terminal domain of Arabidopsis WRKY33 and functional
1032 analysis of tomato WRKY33 homologues in plant stress responses. *J Exp Bot*
1033 **66**(15), 4567-4583 (2015).
1034 Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D., Boller, T., & Felix, G.
1035 Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts
1036 Agrobacterium-mediated transformation. *Cell* **125**(4), 749-760 (2006).
1037

1038 **Acknowledgements**

1039 We thank J.L. Celenza for the *cyp79B2 cyp79B3* mutant. We thank E.S. Sattely for
1040 ICN/ICN-ME, 4OH-ICA/4OH-ICA-ME and camalexin standards. This work was
1041 supported by T32-GM007499 (to B.B.) and Elsevier/Phytochemistry Young Investigator
1042 Award (to N.K.C.).

Author Contributions

B.B. and N.K.C. performed pathogen assays and ChIP-PCR experiments. N.K.C. performed RT-PCR experiments. B.B. and Y.K. profiled accessions and species. B.B. performed all other experiments. B.B. and N.K.C. interpreted the results and wrote the paper.

Competing Interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and material requests can be addressed to Brenden Barco.