bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

*Correspondence should be addressed to B. Barco (brenden.barco@yale.edu)

Current addresses:

(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

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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; Weng *et al*., 2012). They are also often organized into transcription factor (TF) regulons

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of co-regulated genes for optimal timing, amplitude, and tissue-specific pathway gene expression and subsequent metabolite accumulation (Grotewold, 2005; Hartmann, 2007; Martin *et al*., 2010; Tohge & Fernie, 2012; Omranian et al., 2015).

Changes in *cis*-regulatory modules such as enhancers and promoters can accelerate the capture of duplicated genes into regulons, thus driving phenotypic diversity (Levine and Davidson, 2005; Prud'homme *et al*., 2007; Wray, 2007; Wittkopp & Kalay, 2012; Rogers *et al*., 2013). Enhancers consist of transcription factor binding sites (TFBSs) and are derived either through mutation or co-option of a TFBS-carrying transposable element (TE) (Spitz & Furlong, 2012; Wittkopp & Kalay, 2012). TE exaptations have been hypothesized to be responsible for the rapid transcriptional rewiring of gene regulatory networks in ancient lineages of vertebrates (Feschotte 2008; Bourque 2009; Lynch *et al*., 2011; de Souza *et al*., 2013; Chuong et al., 2016) and plants (Hénaff *et al*., 2014), but the physiological significance of this rewiring, if any, is still unknown.

Bacteria elicit two primary immune defense modes in plants, pattern- and effector-triggered immunity (PTI and ETI) (Jones & Dangl, 2006). Pathogenic bacteria additionally compromise PTI via specific virulence effector proteins (effector-triggered susceptibility, ETS; Jones & Dangl, 2006). PTI involves the extracellular perception of conserved molecules known as microbe-associated molecular patterns (MAMPs), whereas ETI involves the cytosolic perception of effectors. Although ETI results in the formation of more rapid and robust pathogen-specific response including the hypersensitive response (HR), a form of programmed cell death (Jones & Dangl, 2006),

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

both result in the ability of naïve host cells to generate, through non-self perception and subsequent transcriptional reprogramming, pathogen-inducible specialized metabolites necessary for defense (Hammerschmidt, 1999; Mansfield, 2000; Clay *et al*., 2009).

Three pathogen-inducible tryptophan (Trp)-derived defense metabolites – camalexin, 4-methoxyindol-3-ylmethylgucosinolate (4M-I3M), and 4-hydroxyindole-3-carbonylnitrile (4OH-ICN) – have been shown to expand innate immunity in *Arabidopsis thaliana* (Bednarek *et al.,* 2009; Clay *et al.,* 2009; Thomma *et al.,* 1999; Tsuji *et al.,* 1992; Rajniak *et al.,* 2015). The three biosynthetic pathways share an early step, which is the conversion of Trp to indole-3-actetaldoxime (IAOx) via the genetically redundant P450 monooxygenases CYP79B2 and CYP79B3 (Fig. 1a) (Zhao *et al*., 2002; Glawischnig *et al*., 2004; Rajniak *et al*., 2015). The camalexin and 4OH-ICN pathways additionally share the conversion of IAOx to indole-3-cyanohydrin (ICY) by partially redundant P450s CYP71A12 and CYP71A13 (Fig. 1a) (Nafisi *et al*., 2007; Klein *et al*., 2013; Rajniak *et al*., 2015). CYP71A13 and CYP71B15/PAD3 catalyze further reactions, leading to camalexin production, whereas the flavin-dependent oxidase FOX1/AtBBE3 and P450 CYP82C2 convert ICY to 4OH-ICN (Fig. 1a) (Nafisi *et al*., 2007; Böttcher *et al*., 2009; Rajniak *et al*., 2015). 4M-I3M is widely distributed across the mustard family (Brassicaceae), whereas camalexin is restricted to the Camelineae tribe of Brassicaceae (Bednarek *et al*., 2011). The evolutionary conservation of 4OH-ICN has not yet been investigated.

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

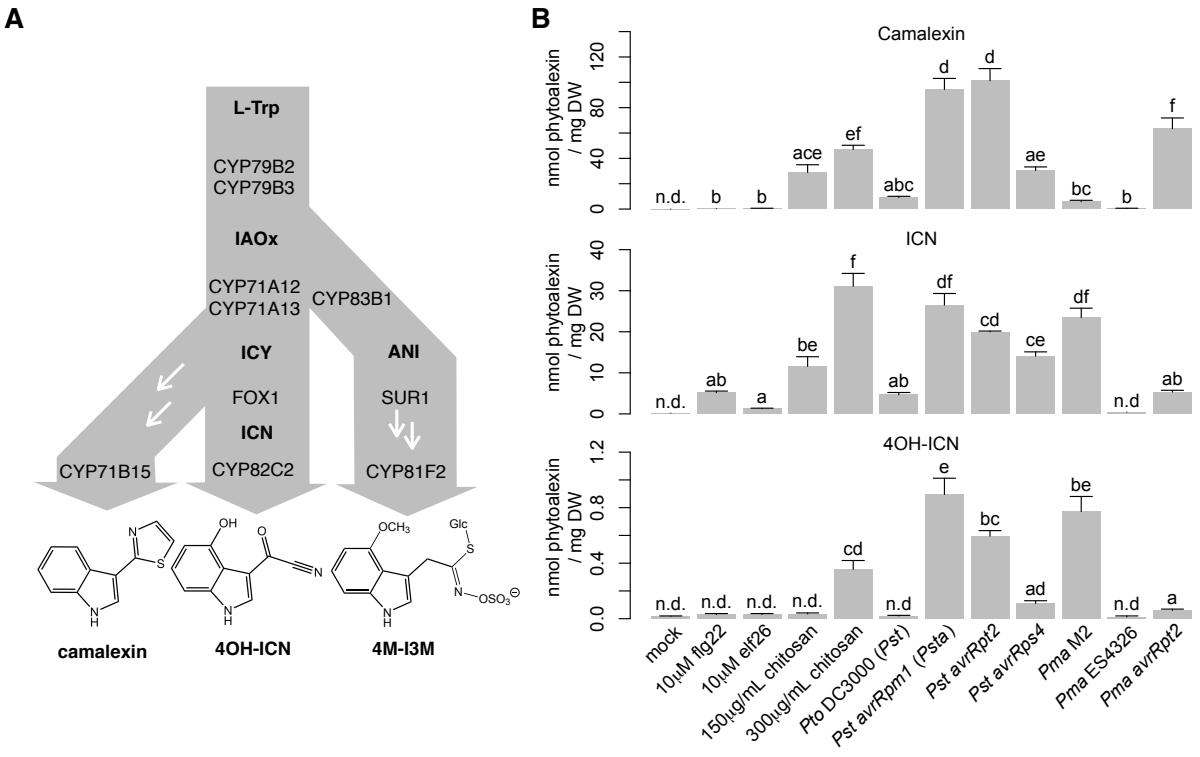


Figure 1. 4OH-ICN is synthesized under ETI-like responses. (**a**). Schematic of tryptophan (L-Trp)-derived specialized metabolism in *A. thaliana*. White arrows denote the presence of additional enzymes. ICY, indole cyanohydrin; ANI, *aci*-nitro indole. (**b**)**.** LC-DAD-FLD-MS analysis of camalexin (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.

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The TF WRKY33 has been shown to regulate the pathogen-inducible biosynthesis of camalexin in *A. thaliana* and its orthologs regulate numerous unrelated specialized metabolites in other flowering plant lineages (Qiu *et al.,* 2005; Liu *et al.,* 2015; Birkenbihl *et al.,* 2017; Schluttenhofer & Yuan, 2015). The group I class of WRKYs to which WRKY33 belongs is an ancient clade of regulators; orthologs in the green alga *Chlamydomonas reinhardtii* may be ancestral to all higher plant WRKYs (Rinerson *et al*., 2015; Schluttenhofer & Yuan, 2015). While all WRKY TFs bind to the W-box core sequence [TTGAC(T/C)], WRKY33 preferentially binds W-boxes that are within 500 nt of the 'WRKY33-specific' motif [(T/G)TTGAAT]) (Rushton *et al.,* 2010; Liu *et al*., 2015).

Here, we show that a recent, lineage-specific TE exaptation resulted in the expansion of a core regulon within the framework of Arabidopsis Trp-derived defense metabolism. Specifically, the LINE retrotransposon *EPCOT3* retroduplicated from a WRKY33-TFBS-carrying progenitor and inserted upstream of the newly duplicated gene *CYP82C2*. Subsequent chromatin remodeling in *A. thaliana* lead *EPCOT3* to become a *bona fide* enhancer with demonstrated biochemical, regulatory, physiological, and fitness-promoting by way of WRKY33-binding, pathogen-responsive *CYP82C2* transcription, 4OH-ICN biosynthesis, and antibacterial defense.

Results

4OH-ICN requires ETI-like responses. To identify the major Trp-derived specialized metabolites synthesized in ETI in *A. thaliana*, we compared host transcriptional and metabolic responses to the PTI-eliciting bacterial MAMPs flg22, elf26, and fungal

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MAMP chitosan, the PTI/ETS-eliciting pathogens *Pseudomonas syringae* pv. *tomato* DC3000 (*Pto* DC3000 or *Pst*), *Pseudomonas syringae* pv. *maculicola* ES4326 (*Pma*) and the ETI-eliciting pathogens *Pst avrRpm1* (*Psta*), *Pst avrRpt2*, *Pst avrRps4*, *Pma* M2, and *Pma avrRpt2* under similar conditions as those of previous studies (Denoux *et al*., 2008; Clay *et al*., 2009). *Psm* M2 is an ETI-eliciting strain from which the *avrRpm1* gene was originally isolated (Debener *et al.*, 1991). Both flg22 and *Psta* induced genes involved in 4OH-ICN, camalexin and 4M-I3M biosynthesis, with 4OH-ICN and camalexin biosynthetic genes having a higher level of induction than those of 4M-I3M in *Psta*-inoculated plants (Supplementary Fig. 1a; Denoux *et al*., 2008). In contrast to the quantitative differences observed in transcriptional responses between PTI and ETI (Tao *et al*., 2003; Navarro *et al*., 2004), the metabolite responses between PTI and ETI differed largely qualitatively. 4OH-I3M and 4M-I3M were present in uninfected plants and accumulated to modest levels at the expense of parent metabolite I3M in flg22- and *Psta*-inoculated plants (Supplementary Fig. 1b) (Clay *et al*., 2009). By comparison, ICN, 4OH-ICN, and camalexin were absent in uninfected plants and at low-to-undetectable levels in plants treated with saturating concentrations of the bacterial MAMPs flg22 and elf26 (10 μM; Felix *et al*., 1999; Zipfel *et al*., 2006). In contrast, ICN, 4OH-ICN and camalexin accumulated to high levels upon inoculation with ETI-inducing pathogens (Fig. 1b; Supplementary Fig. 1c). Furthermore, camalexin, ICN, and 4OH-ICN metabolism was greatly diminished, and indole glucosinolate levels were mostly unchanged in the *rpm1* mutant, which is impaired in ETI recognition of *Psta* (Bisgrove *et al*., 1994) (Supplementary Fig. 1b-c). By contrast, camalexin and ICN were absent in uninfected plants and largely at low-to-undetectable levels in plants treated with

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MAMPs and PTI/ETS-eliciting pathogens, with 4OH-ICN not detected in most cases. One exception was the fungal MAMP chitosan. 150 µg/mL chitosan induced high levels of camalexin and detectable levels of ICN, consistent with previous observations of camalexin biosynthetic genes upregulation (Fig. 1b) (Povero *et al*., 2011). Higher chitosan concentrations (≥200 µg/mL) have been shown to induce HR-like cell death in Arabidopsis (Cabrera et al., 2006), a phenomenon commonly observed for ETI (Jones and Dangl, 2006). To our surprise, 300 µg/mL chitosan additionally induced detectable levels of 4OH-ICN (Fig. 1b). These results suggest that 4OH-I3M, 4M-I3M, camalexin, and ICN are synthesized in response to multiple PTI elicitors, whereas 4OH-ICN biosynthesis is specific to ETI-like responses.

WRKY33* is required to activate 4OH-ICN in response to *Psta. 4OH-ICN biosynthetic genes are highly co-expressed with each other (Rajniak *et al*., 2015) and with camalexin biosynthetic genes (Supplementary Fig. 1d), which are in the WRKY33 regulon (Qiu *et al*., 2008; Birkenbihl *et al*., 2012). To determine whether 4OH-ICN biosynthetic genes are also in the WRKY33 regulon, we compared camalexin, ICN and 4OH-ICN levels between wild-type and a *wrky33* loss-of-function mutant that encodes two differently truncated proteins (Fig. 2a; Zheng *et al*., 2006). Consistent with a previous report (Qiu *et al*., 2008), *wrky33* was impaired in camalexin biosynthesis in response to *Psta* and *Pst avrRps4* (Fig. 2b; Supplementary Fig. 2a). The *wrky33* mutant was similarly impaired in 4OH-ICN biosynthesis (Fig. 2b; Supplementary Fig. 2a). These

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

158 results indicate that WRKY33 is required for camalexin and 4OH-ICN biosynthesis in

159 response to multiple ETI elicitors.

160

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

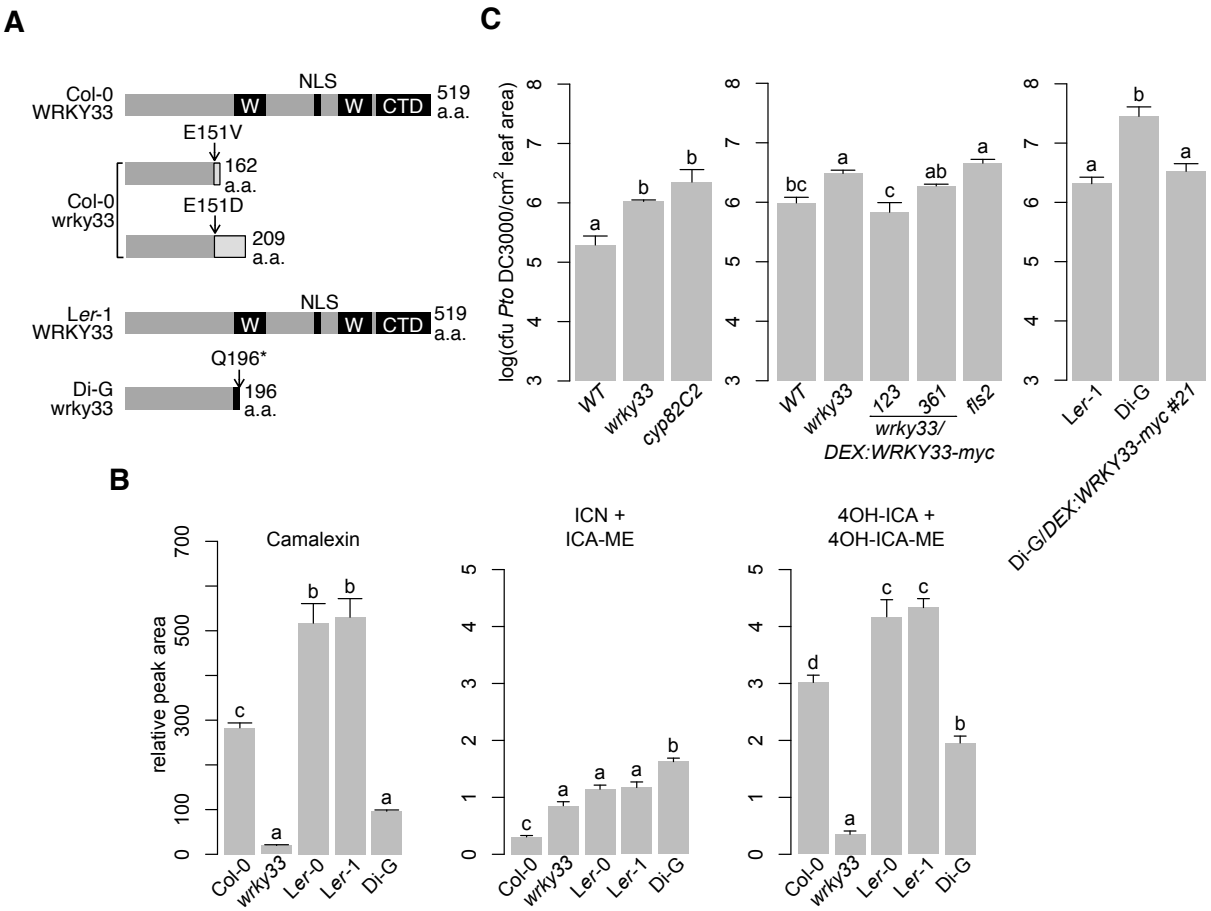


Figure 2. Intraspecific variation in *WRKY33* affects 4OH-ICN and immunity. (**a**) Schematic of WRKY33 proteins in Col-0, Col-0 *wrky33*, L*er*-1 and Di-G. Black boxes denote WRKY domains (W), nuclear localization signal (NLS), or C-terminal domain (CTD). (**b**) LC-DAD-MS analysis of camalexin, ICN, 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

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

172 similar results.

173

174

175

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 3a-b). In addition, differences observed in the metabolite response between Landsberg accessions and Di-G most closely resembled those between Col-0 and *wrky33* mutant (Fig. 2b; Supplementary Fig. 3a). These results led us to hypothesize that genetic variation in a regulatory gene, as opposed to an immune signaling gene, is responsible for the metabolite phenotypes observed in Di-G. To test this hypothesis, genetic variation between Di-G and three sequenced Landsberg accessions (La-0, L*er*-0, and L*er*-1) were used to identify 354 genes that were differentially mutated to high effect in Di-G (Supplementary Fig. 3c). Twenty-eight of these mutated Di-G genes were annotated by Gene Ontology to have roles in defense, including *WRKY33* (Supplementary Table 1). We confirmed by Sanger sequencing that Di-G *WRKY33* harbors a nonsense mutation early in the N-terminal DNA-binding motif (Fig. 2a), likely abolishing protein function. Our findings indicate that camalexin and 4OH-ICN are sensitive to intraspecific variation in *WRKY33*.

Camalexin and 4OH-ICN promote plant fitness by contributing non-redundantly to pathogen defense against the fitness-reducing *Pst* (Kover & Scaal, 2002; Rajniak *et al*., 2015). To confirm that disease resistance to *Pst* is also sensitive to intraspecific variation in *WRKY33*, we measured bacterial growth in adult leaves of *wrky33* and Di-G and their respective (near-)isogenic accessions Col-0 and L*er*-1. *wrky33* and Di-G were more susceptible to *Pst* than their (near)isogenic relatives and comparable to the 4OH-ICN biosynthetic mutant *cyp82C2* (Fig. 2c; Rajniak *et al*., 2015).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

WRKY33 activates 4OH-ICN biosynthesis. To confirm that the 4OH-ICN biosynthetic pathway is in the WRKY33 regulon, we first compared *WRKY33*, *CYP71A13, CYP71B15, FOX1* and *CYP82C2* transcript levels among WT, *wrky33*, *wrky33*/*DEX:WRKY33-flag*, and *wrky33*/*DEX:WRKY33-myc*. Consistent with previous reports (Qiu *et al*., 2008), *CYP71A13*, *CYP71B15*, and *FOX1* expression was down-regulated in *wrky33* plants in response to *Psta* and upregulated in both *wrky33*/*DEX:WRKY33-flag* and *wrky33*/*DEX:WRKY33-myc* (Fig. 3a) (Supplementary Fig. 4f, 5a). Interestingly, *CYP82C2* expression and 4OH-ICN production were restored in *wrky33*/*DEX:WRKY33-flag* but not *wrky33*/*DEX:WRKY33-myc* or Di-G/*DEX:WRKY33-myc* plants (Fig. 2d, 3a) (Supplementary Fig. 4d-f), likely due to the interference of the larger myc tag with the WRKY33 C-terminus, a region previously linked with transactivation activity (Zhou *et al*., 2015). These transcriptional and metabolic findings

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

243 indicate that WRKY33 mediates camalexin and 4OH-ICN biosynthesis in response to

244 pathogen effectors.

245

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

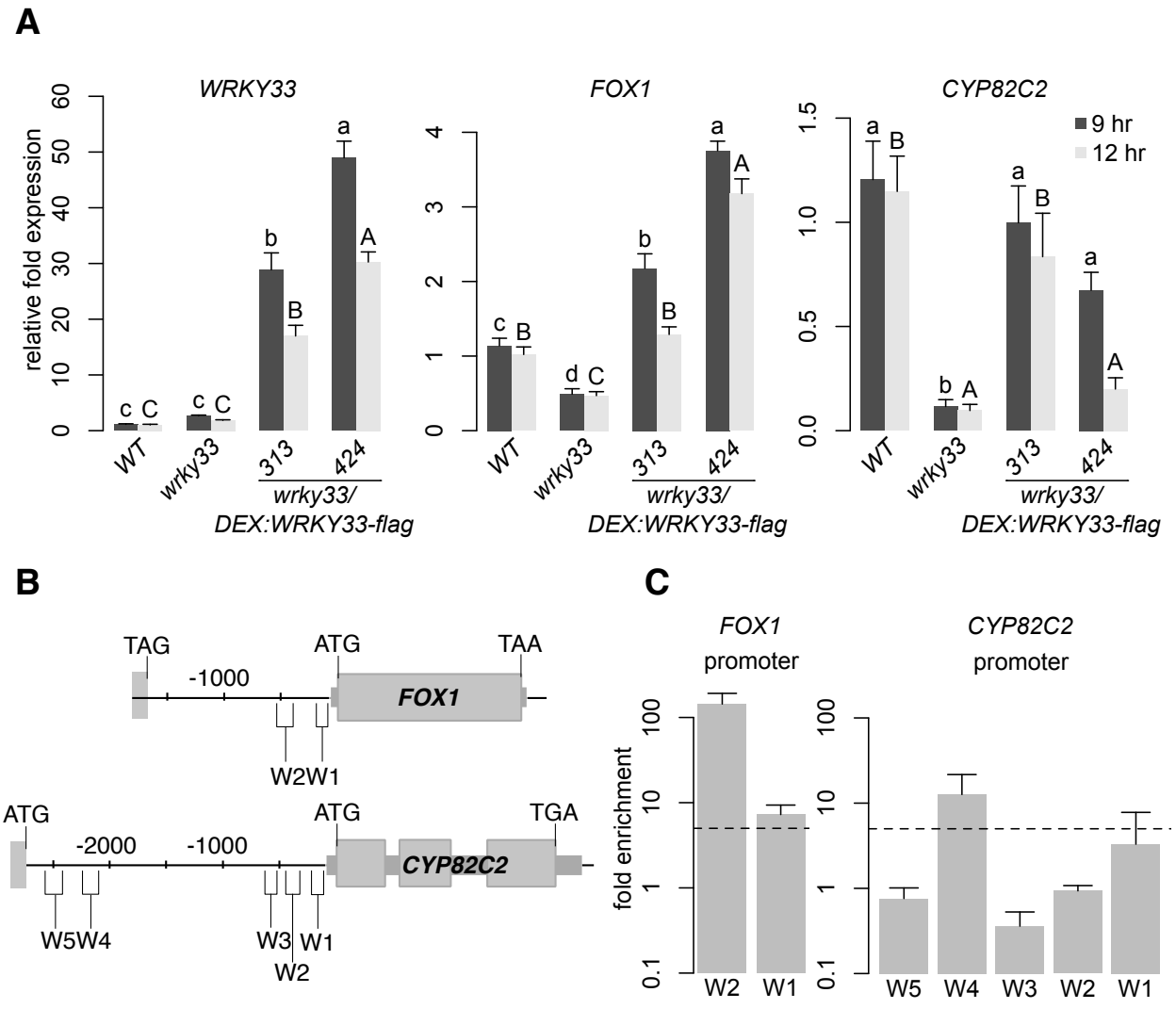


Figure 3. WRKY33 directly activates 4OH-ICN biosynthetic genes. (**a**) qPCR analysis of 4OH-ICN regulatory and biosynthetic genes in seedlings inoculated with 20 μM dex and *Psta* for 9 and 12 hr. Different letters denote statistically significant differences ($P < 0.05$, one-factor ANOVA coupled to 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*

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

255 plants co-treated with 20 μM dex (D) or mock solution (M) and *Psta* for 9 hr. Dashed
256 line represents the 5-fold cutoff between weak and strong TF-DNA interactions. Data
257 represent mean ± SE of four replicates.
258
259 We then tested for WRKY33-binding to W-box-containing regions upstream of
260 camalexin and 4OH-ICN biosynthetic genes in dex-treated and *Psta*-infected
261 *wrky33/DEX:WRKY33-flag* seedlings by chromatin immunoprecipitation (ChIP)-PCR.
262 WRKY33 has been shown to bind to a W-box region upstream of *CYP71A12*
263 (Birkenbihl *et al*., 2017), a region that also contains three WRKY33-specific motifs and
264 is consistent with WRKY33's reported binding site preference (Liu *et al*., 2015). We
265 additionally observed that *Psta*-induced WRKY33 bound strongly (greater than 5-fold
266 enrichment) to a single W-box region upstream of *FOX1* and *CYP82C2* (W2 and W4,
267 respectively; Fig. 3b-c; Supplementary Fig. 5b). Both regions also contain one to three
268 WRKY33-specific motifs. Together with our expression analysis, our findings indicate
269 that WRKY33 uses preferred WRKY33-binding sites to directly activate 4OH-ICN
270 biosynthetic genes in response to pathogen effectors.
271
272 Interestingly, *Psta*-induced WRKY33 did not bind to the W5 region upstream of
273 *CYP82C2* (Fig. 3c), a W-box region that does not contain any WRKY33-specific motifs
274 and is just upstream of neighboring gene of unknown function *At4g31960* (Fig. 3b).
275 WRKY33 reportedly binds to W5 in response to flg22 and *B. cinerea* (Liu *et al*., 2015;
276 Birkenbihl *et al*., 2017). By contrast, *Psta*-induced WRKY33 bound strongly to W1
277 region upstream of *CYP71B15* (Supplementary Fig. 5c-d), a W-box region that also

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

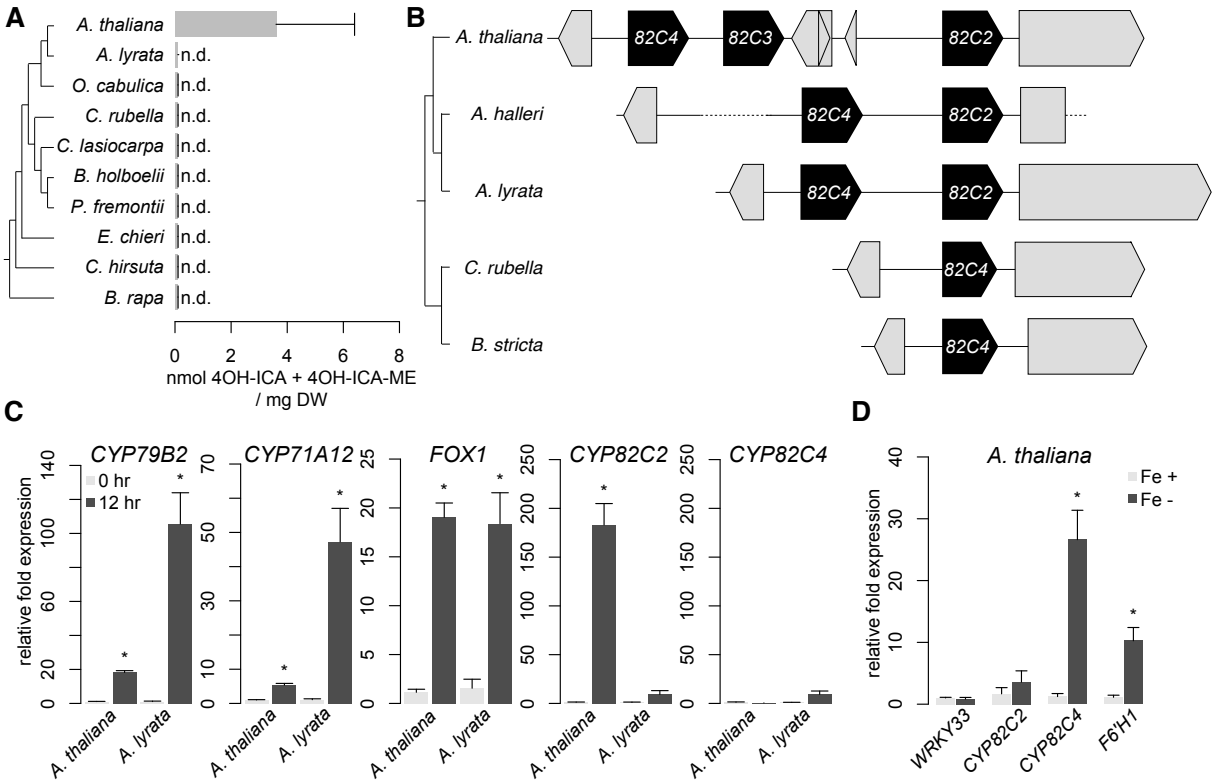


Figure 4. Regulatory neofunctionalization of *CYP82C2*. (**a**) (Left) Phylogenetic species tree. (Right) HPLC-DAD analysis of 4OH-ICN in seedlings inoculated with *Psta* for 30 hr. Data represent mean ± SE of three independent experiments (n = 4 biological replicates), each with *A. thaliana* as a positive 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 untreated samples ($P < 0.05$, two-tailed t-test).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In *A. thaliana*, *CYP82C2* resides in a near-tandem cluster with paralogs *CYP82C3* and *CYP82C4* (Fig. 4b). We performed phylogenetic and syntenic analyses to identify putative *CYP82C2* orthologs in ICN-synthesizing species. All identified homologs are syntenic to *CYP82C2* or *CYP82C4*, and encode proteins with >88% identity to one another (Fig. 4b; Supplementary Fig. 6b-c). *CYP82C3* is present only in *A. thaliana*, and although more similar to *CYP82C2* than *CYP82C4* in sequence, it is not functionally redundant with *CYP82C2* (Fig. 4b; Supplementary Fig. 6b; Rajniak *et al*., 2015). *CYP82C4* is required for the biosynthesis of sideretin, a widely conserved, phenylalanine-derived metabolite required for iron acquisition (Rajniak *et al*., 2018). *CYP82C4* has syntenic orthologs in the mustard family, correlating with the distribution of sideretin biosynthesis (Fig. 4b; Supplementary Fig. 6b**;** Rajniak *et al*., 2018). By contrast, *CYP82C2* has syntenic orthologs only within the *Arabidopsis* genus (Fig. 4b; Supplementary Fig. 6b). These results suggest that *CYP82C2* duplicated from *CYP82C4* prior to the formation of the *Arabidopsis* genus and then acquired a new expression pattern and/or catalytic function prior to *A. thaliana* speciation approx. 2 million years later (Hu *et al*., 2011; Hohmann *et al*., 2015).

CYP82C2 and CYP82C4 were previously characterized to 5-hydroxylate with equal efficiency the specialized metabolite 8-methoxypsoralen, a molecule structurally reminiscent of ICN and sideretin (Kruse *et al*., 2008). The apparent similarities in substrate specificity and catalytic function suggest that *CYP82C2* may have diverged from *CYP82C4* in expression but not protein function. To test this, we first compared the expression of *CYP82C2* and *CYP82C4* in *A. lyrata* and *A. thaliana* in response to

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

331 *Psta*. 4OH-ICN biosynthetic genes *CYP79B2, CYP71A12* and *FOX1* were upregulated
332 in both species, consistent with the common presence of ICN (Fig. 4a and c). By
333 contrast, *CYP82C2* levels were respectively upregulated and unchanged in *A. thaliana*
334 and *A. lyrata*, correlating with the distribution of 4OH-ICN in these species (Fig. 4a and
335 c). *CYP82C4* expression was unchanged in both species (Fig. 4c). These results
336 indicate that 4OH-ICN biosynthesis is linked with pathogen-induced expression of
337 *CYP82C2.*
338
339 We then compared the aligned upstream sequences of *CYP82C2* and *CYP82C4* in *A.*
340 *lyrata* and *A. thaliana* and observed good sequence conservation among orthologs but
341 poor conservation among paralogs (Supplementary Fig. 6d), indicating that sequences
342 upstream of *CYP82C4* and *CYP82C2* were independently derived. We performed
343 expression analysis in *A. thaliana* to confirm that *CYP82C2* and *CYP82C4* have
344 different expression patterns. *CYP82C2* expression is upregulated in response to *Psta*
345 and unchanged under iron deficiency (Fig. 4c-d; Supplementary Fig. 1a**;** Rajniak *et al*.,
346 2015). Conversely, *CYP82C4* is upregulated under iron deficiency and unchanged in
347 response to *Psta* (Fig. 4c-d; Murgia *et al*., 2011; Rajniak *et al*., 2018). Finally, *CYP82C4*
348 was unchanged in *Psta*-challenged *wrky33* and *wrky33*/*DEX:WRKY33-flag*
349 (Supplementary Fig. 6e). Our findings suggest that *CYP82C2* diverged from *CYP82C4*
350 by acquiring WRKY33 regulation for its pathogen-induced expression.
351
352 We next assessed dN/dS ratios along branches of the CYP82C phylogenetic tree
353 (Supplementary Fig. 6b) and found good support for purifying selection acting on

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

354 CYP82C enzymes (ω=0.21), and no support for positive selection acting on
355 CYP82C2/3 enzymes (Supplementary Table 2). Lastly, we identified non-conserved
356 amino acid residues among CYP82C homologs and mapped this information onto a
357 homology model of CYP82C2. The protein inner core, which encompasses the active
358 site and substrate channel, is highly conserved among CYP82C homologs
359 (Supplementary Fig. 6f), and is consistent with CYP82C2 and CYP82C4's reportedly
360 redundant catalytic functions (Kruse *et al*., 2008). Altogether, our findings suggest that
361 *CYP82C2* underwent regulatory neofunctionalization (Moore & Purugganan, 2005),
362 diverging from *CYP82C4* in expression but not protein function.
363
364 **TE *EPCOT3* is a *CYP82C2* enhancer**. WRKY33 regulation of *CYP82C2* is mediated by
365 a WRKY33-TFBS in the W4 region (Figs. 3 and 5a; Supplementary Fig. 5c). Preferential
366 WRKY33-binding at this region should also be influenced by chromatin features
367 associated with *cis*-regulatory elements like enhancers and basal promoters (Slattery
368 *et al*., 2014). To investigate how *CYP82C2* acquired WRKY33-binding for its pathogen-
369 induced expression, we compared the aligned upstream sequences of *CYP82C*
370 homologs in ICN-synthesizing species. We observed three large upstream sequences
371 specific to *A. thaliana CYP82C2*, hereafter named ***E****ighty-two-C2* ***P****romoter* ***C****ontained*
372 ***O****nly in A.* ***T****haliana1-3* (*EPCOT1–3*; Fig. 5a). *EPCOT3* in particular is a 240nt region that
373 completely encompasses W4 (Fig. 5a), indicating that the WRKY33's regulation of
374 *CYP82C2* in response to *Psta* may be species-specific. Further bioinformatics analysis
375 revealed that *EPCOT3* has the epigenetic signature of an active enhancer (Roudier *et*
376 *al*., 2011; Liu *et al*., 2018). Relative to neighboring sequences, *EPCOT3* is enriched with

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

activating histone mark H3K4me2 and lacks the repressive histone mark H3K27me3 (Fig. 5b) (Heintzman *et al*., 2007; Hoffman *et al*., 2010; Roudier *et al*., 2011; Bonn *et al*., 2012; Wang *et al*., 2014). Our findings suggest that *EPCOT3* functions as an enhancer that mediates WRKY33-binding and activation of *CYP82C2* in response to pathogen effectors.

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

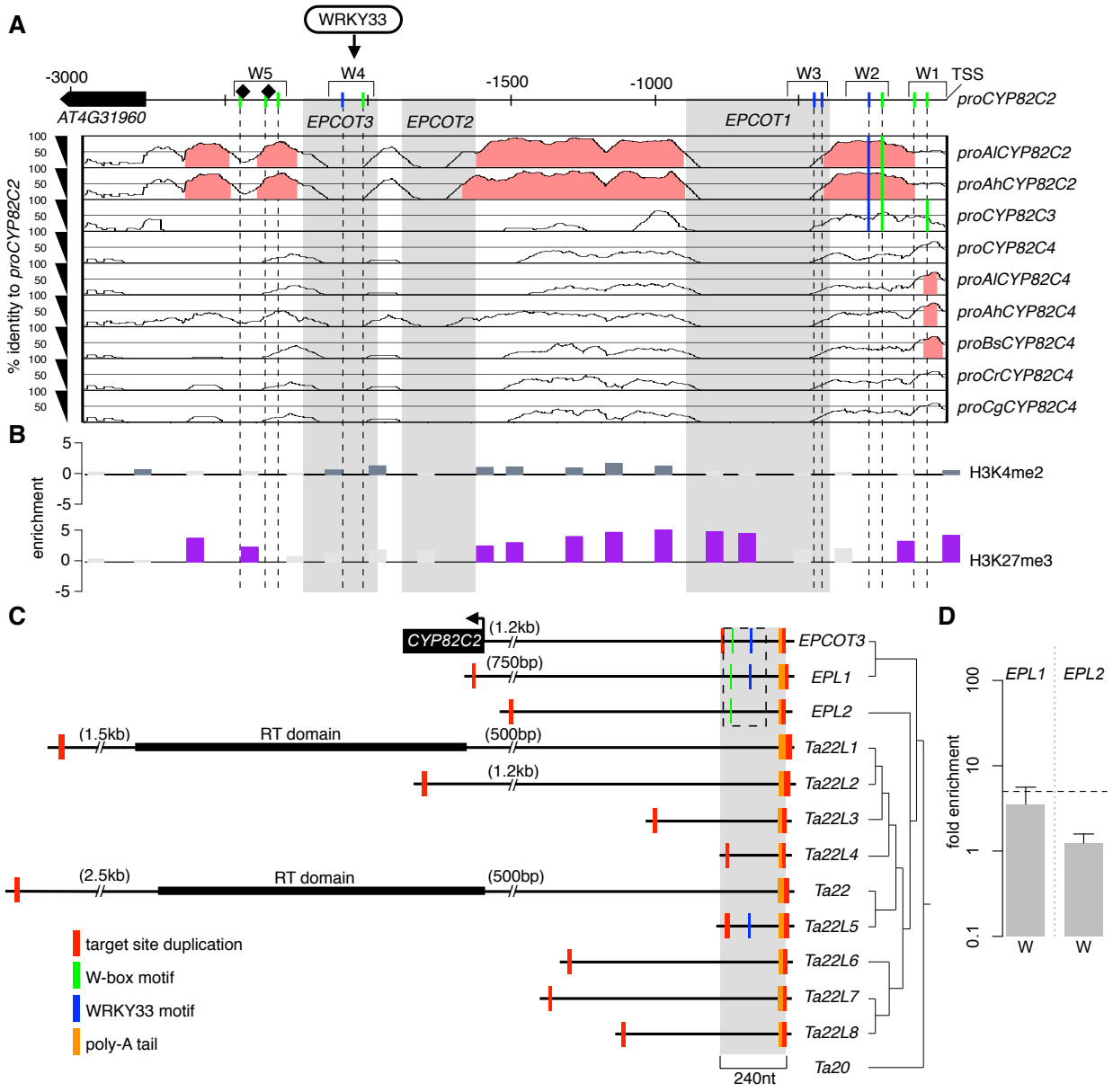


Figure 5. TE *EPCOT3* is a *CYP82C2* enhancer. (**a**) mVISTA plot of *CYP82C2* upstream sequence, indicating nt positions of unique (*EPCOT1–3;* gray boxes) and conserved regions (≥70% sequence identity; pink) among homologous sequences. Also indicated are positions of W-boxes (green) and WRKY33-specific motifs (blue) that are present (solid lines) or absent (dashed lines) in each homologous sequence, previously known WRKY33-TFBSs (diamonds) and ChIP-tested regions (W1-5). TSS,

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transciptional start site; *Al, Arabidopsis lyrata; Ah, Arabidopsis halleri*; *Cr, Capsella rubella*; *Bs, Boechera stricta; Cg, Capsella grandiflora*. (**b**) Epigenetic map of *CYP82C2* upstream sequence, indicating nt positions of significant amounts of H3K4me2 (blue-gray bars), and H3K27me3 (purple bars). (**c**) (Left) Schematic of *EPCOT3* and related LINE retrotransposons in *A. thaliana* drawn to scale, indicating nt positions of *CYP82C2* and reverse transcriptase (RT) domain. A more detailed tree is available as Supplementary Text 1. (Right) Phylogenetic maximum likelihood tree. Dashed box represent region containing W-boxes (green lines) and/or WRKY33-binding motifs (blue lines) within *EPCOT3, EPL1* and *EPL2*. (**d**) ChIP-PCR analysis of W-box-containing regions (W) within *EPL1* and *EPL2* in *wrky33*/*DEX:WRKY33-flag* plants co-treated with 20 μM dex (D) or mock solution (M) and *Psta* for 9 hr. Data represent mean ± SE of four replicates. Dashed line represents the 5-fold cutoff between weak and strong TF-DNA interactions.

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

EPCOT3 contains a 3' poly-A tail and is flanked by variable-length target site duplications (Fig. 5c; Supplementary Fig. 7a), which are hallmarks of eukaryotic LINE retrotransposons (Malik *et al*., 1999). LINE retrotransposition (reverse transcription and integration) results in frequent 5'-truncation of retrocopies (Luan *et al*., 1993). We identified eleven variably truncated retrocopies similar to *EPCOT3* throughout the genome, including *Ta22*, one of the first LINEs characterized in *A. thaliana* (Fig. 5c; Supplementary Fig. 7a-b, Supplementary Table 3; Wright *et al*., 1996). *EPCOT3*-related LINEs were sorted into two groups roughly correspondent to their phylogenetic placement: *EPCOT3-LIKE* (*EPL*) for those with high identity (>65%) to *EPCOT3* and *Ta22* or *Ta22-LIKE* (*Ta22L*) for the remainder (Supplementary Fig. 7a; Supplementary Table 3). Only *Ta22* and *Ta22L1* are full-length LINEs (Fig. 5c), presumably encoding the proteins necessary for their own transposition and for the transposition of nonautonomous family members like *EPCOT3*. We also identified two syntenic species-specific *Ta22L*s, but no *EPLs*, in *A. lyrata* (Supplementary Table 3). Given the 80% overall sequence identity between *A. thaliana* and *A. lyrata* (Hu *et al*., 2011), this data indicates that *EPCOT3* and *EPLs* arose from retrotransposition following the speciation of *A. thaliana.*

Of all the retrocopies, *EPL1* is most similar to *EPCOT3* (85.4% identity), sharing the W-box and WRKY33-specific motif, whereas *EPL2* is less similar (67%) and lacks the WRKY33-specific motif (Fig. 5c; Supplementary Table 3, Supplementary Fig. 7a). *EPL1* and *EPL2* are much less truncated than *EPCOT3* (Fig. 5c), and lack epigenetic signatures typical of *cis*-regulatory sequences (Supplementary Fig. 7c) (Roudier *et al*.,

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2011; Liu *et al*., 2018). To investigate whether the sequence information and chromatin features associated with *EPL*s are sufficient for WRKY33-binding, we tested for WRKY33-binding to *EPL* sequences homologous to the W4 region of *EPCOT3* in dex-treated, *Psta*-infected *wrky33*/*DEX:WRKY33-flag* plants by ChIP-(q)PCR. Compared to *EPCOT3* (Fig. 3c), WRKY33 respectively bound weakly or not at all to *EPL1* and *EPL2* (Fig. 5d; Supplementary Fig. 7d). Our findings suggest the following history: (1) *EPL1* likely retroduplicated from *EPL2* or its progenitor, which already contained a W-box; (2) *EPL1* then acquired a WRKY33-specific motif by mutation; (3) *EPCOT3* retroduplicated from *EPL1* and then acquired epigenetic signatures of an enhancer, thereby allowing selection to act on standing variation rather than *de novo* mutation for *CYP82C2* recruitment into the 4OH-ICN biosynthetic pathway.

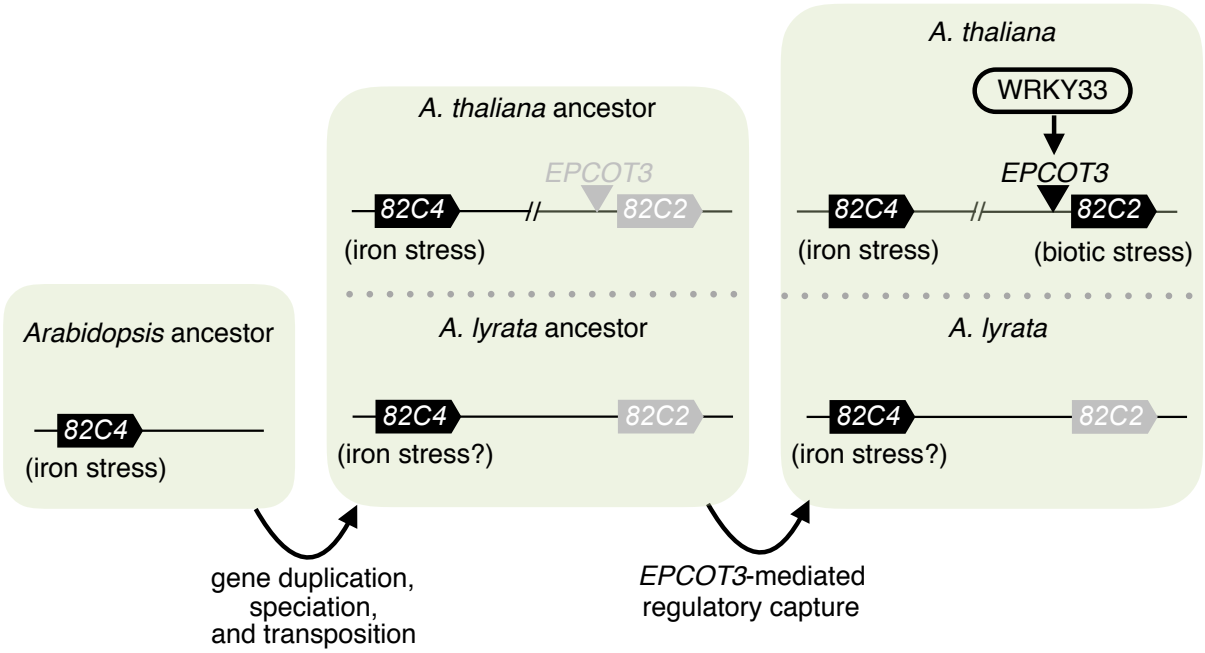
Discussion

TEs were originally conceived to act as “controlling elements” of several loci in the genome (McClintock, 1956), and exaptation of TEs into *cis*-regulatory modules has been hypothesized to be responsible for the rapid transcriptional rewiring in more ancient lineages of vertebrates (Feschotte 2008; Bourque 2009; de Souza et al., 2013). However, few (if any) evolutionarily recent TE exaptation events in vertebrates and higher plants have been demonstrated to have biochemical, regulatory, physiological and fitness-promoting functions (de Souza et al., 2013). With well over a dozen genomes available including the genetic model *A. thaliana*, the mustard family presents an excellent system for examining such events. In this study, we show that *EPCOT3* is a TE-derived enhancer that mediates WRKY33-binding, pathogen-responsive

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcription of *CYP82C2,* synthesis of the species-specific metabolite 4OH-ICN, and pathogen defense (Fig. 6). These results provide the first instance of a recent TE exaptation responsible for the rewiring of a new gene into an ancient regulon, ultimately leading to a positive effect on fitness.

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



455

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

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

457 Although the *EPL1/EPCOT3* progenitor retrotransposed a preferred WRKY33-TFBS in
458 the form of *EPCOT3* upstream of *CYP82C2*, a further series of epigenetic modifications
459 were needed to facilitate optimal access of *EPCOT3* by WRKY33 (Fig. 6). *EPL1* exists
460 in a silenced heterochromatin state (Supplementary Fig. 7c), typical for TEs (Slotkin &
461 Martienssen, 2007), and is bound weakly by WRKY33 (Fig. 5d), whereas *EPCOT3* is in
462 an open chromatin state (Fig. 5b; Roudier *et al*., 2011; Liu *et al*., 2018) and bound
463 strongly by WRKY33 (Fig. 3c). The more severe 5'-truncation of *EPCOT3* could
464 account for its release from TE silencing mechanisms, and the initially weak WRKY33-
465 binding could provide a 'seed' for chromatin remodelers to drive the exaptation of
466 newly retrotransposed *EPCOT3* into a *bona fide* enhancer. Further epigenomic
467 sampling within *Arabidopsis* is needed to better clarify the epigenetic transformations
468 underlying the *EPCOT3* exaptation event.
469
470 Compared to closely-related Landsberg accessions (Supplementary Fig. 3; Hardtke *et*
471 *al*., 1996), Di-G synthesizes less camalexin and 4OH-ICN (Fig. 2b; Kagan &
472 Hammerschmidt, 2002), is more susceptible to a range of bacterial and fungal
473 pathogens (Fig. 2c) (Hugouvieux *et al*., 1998; Kagan & Hammerschmidt, 2002;
474 Mukherjee *et al*., 2009), and is more sensitive to the phytohormone ethylene (Chatfield
475 *et al*., 2008). WRKY33 has been implicated in camalexin biosynthesis (Qiu *et al*., 2008),
476 antifungal defense (Zheng *et al*., 2006), and ethylene biosynthesis (Li *et al*., 2012). We
477 identified WRKY33 as causal for some if not all of these phenotypes in Di-G. This is the
478 first report of WRKY33's involvement in antibacterial defense and is consistent with the

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

479 contribution of camalexin and 4OH-ICN towards antibacterial defense (Rajniak *et al*.,
480 2015).
481
482 WRKY33 is an ancient transcription factor responsible for many fitness-promoting
483 traits in plants, thus it is unexpected that an *A. thaliana* accession would have a
484 naturally occurring *wrky33* mutation (C536T transversion). Di-G is the sole member of
485 1,135 sequenced accessions to have a high-effect single nucleotide polymorphism
486 (SNP) in *WRKY33* (1001 Genomes Consortium, 2016). Di-G and L*er*-0 have long been
487 models for studies in mutagenesis (Rédei, 1962, Müller, 1966), and thus a possibility
488 exists that Di-G may have originated from an ethyl methanesulfonate (EMS)
489 mutagenesis screen of L*er*-0. Historical EMS mutagenesis experiments generated
490 upwards of tens of thousands of mutations per cell (Müller 1966; Rédei & Koncz, 1993;
491 Camara *et al*., 2000), well within the range of ~25,000 SNPs that are not concordant
492 between Di-G and L*er*-0 (Supplementary Fig. 2f). However, features of EMS mutations
493 (i.e. transversion mutations) or X-ray mutations (i.e. indels) are not enriched in the Di-G
494 pseudogenome relative to related pseudogenomes (Supplementary Table 4). These
495 findings suggest that the wrky33 Di-G mutation is naturally derived.
496
497 **Methods**
498 **Plant materials and growth conditions.** For qPCR and HPLC-DAD analyses, surface-
499 sterilized *Arabidopsis thaliana* seeds were sown in 12-well microtiter plates sealed with
500 Micropore tape (3M, St. Paul, MN), each well containing ~15 seeds and 1 mL filter-
501 sterilized 1X Murashige and Skoog (MS; Murashige & Skoog, 1962) media (pH 5.7–5.8)

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(4.43 g/L MS basal medium with vitamins [Phytotechnology Laboratories, Shawnee Missions, KS], 0.05% MES hydrate, 0.5% sucrose). Iron deficient media was made as previously described by Rajniak et al (2018). For *Polyctenium fremontii*, surface-sterilized seeds were sown on MS agar plates. On day 9, seedlings were transferred to 6-well microtiter plates, each well containing ~15 seeds and 3 mL MS media. For all other species, surface-sterilized seeds were sown in 6-well plates, each well containing ~15 seeds and 3 mL MS media. On day 9, media were refreshed prior to bacterial elicitation. Microtiter plates were placed on grid-like shelves over water trays on a Floralight cart (Toronto, Canada), and plants were grown at 21°C with 60% humidity under a 16-hr light cycle (70-80 µE m-2 s-1 light intensity). For chromatin immunoprecipitation analyses, approximately 200 seeds were sown in a 100mm x 15mm petri plate containing 20mL of 1X MS media. Media were exchanged for fresh media on day 9. Microtiter plates were placed on grid-like shelves over water trays on a Floralight cart (Toronto, Canada), and plants were grown at 21°C with 60% humidity under a 16-hr light cycle (70-80 µE m-2 s-1 light intensity). For bacterial infection assays, seedlings were transferred to and grown on soil [3:1 mix of Farfard Growing Mix 2 (Sun Gro Horticulture, Vancouver, Canada) to vermiculite (Scotts, Marysville, OH)] at 22˚C daytime/18˚C nighttime with 60% humidity under a 12-hr light cycle [50 (dawn/dusk) and 100 (midday) µE m-2 s-1 light intensity].

Seed stock information is shown in Supplementary Table 5.

Vector construction and transformation. To generate the *DEX:WRKY33-flag* construct, *WRKY33* was PCR-amplified from genomic DNA using the primers

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

WRKY33gXhoF (5'-AACTCGAGAAGAACAAGAACCATCAC-3'), and W33flgSpeIR (5'-CGACTAGTCTACTTGTCGTCATCGTCTTTGTAGTCGGGCATAAACGAATCGAAA-3') and subcloned into the *Xho*I/*Spe*I sites of pTA7002 vector (Aoyama and Chua, 1997; McNellis et al., 1998). To generate the *DEX:WRKY33-myc* construct, *WRKY33* was PCR-amplified using the primers WRKY33gXhoF and WRKY33gStuR (5'-AAGGCCTGGCATAAACGAATCGAAAAATG-3') and subcloned into the *Xho*I/*Stu*I sites of a version of pTA7002 modified to contain 6 tandem copies of the c-Myc epitope downstream of the StuI site (Chezem *et al.*, 2017). The constructs were introduced into *Arabidopsis thaliana wrky33* plants via *Agrobacterium*-mediated floral dip method (Clough and Bent, 1998), and transformants were selected on agar media containing 15 µg/mL hygromycin B (Invitrogen, Carlsbad, CA).

Bacterial infection and MAMP elicitation. A single colony of *Pseudomonas syringae* pv. *maculicola* (*Pma*) M2 (containing *avrRpm1*, but not *avrRps4* or *avrRpt2*), *Pma* ES4326 (containing no aforementioned effectors), *Pma* ES4326 *avrRpt2*, *Pseudomonas syringae* pv. *tomato* DC3000 (*Pto* DC3000 or *Pst,* containing no aforementioned effectors), *Pst avrRpm1*, *Pst avrRps4*, and *Pst avrRpt2* from a freshly streaked 3-day-old agar plate were used to inoculate 2 mL of LB containing appropriate antibiotics. Strains were grown to log phase, washed in sterile water twice, resuspended in water to OD_{600} of 0.2, and incubated at room temperature with no agitation for 3-6 and prior to infection. Chitosan (90% deacetylated chitin; Spectrum Chemical Mfg Corp, New Brunswich, NJ) was prepared in 0.1 N acetic acid and neutralized with 0.1 N NaOH to pH 5.8 to a stock concentration of 1.2 mg/mL.

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

548

549 Hydroponically grown 9-day-old seedlings were inoculated with bacterial strains to
550 OD_{600} of 0.013 or treated with 10 μM flg22 (QRLSTGSRINSAKDDAAGLQIA; Genscript,
551 Nanjing, China), 10 μM elf26 (ac-SKEKFERTKPHVNVGTIGHVDHGKTT; Genscript), and
552 150 or 300 μg/mL chitosan.

553

554 For qPCR analyses, seedlings were snap-frozen in liquid nitrogen 12 hr post-infection.
555 For HPLC-DAD analyses, seedlings were snap-frozen 24 to 28 hr post-infection. For
556 ChIP analyses, seedlings were snap-frozen 9 hr post-infection.

557

558 4-to-5-week-old adult leaves were treated with 0.0125% Silwet or 0.0125% Silwet and
559 20 μM dexamethasone for 20 sec and incubated on 0.8% (w/v) tissue-culture agar
560 plates on a light cart at 21˚C for 6-8 hr. Leaves were then surface-inoculated with *Pto*
561 DC3000 (OD_{600} = 0.002 or 10^6 colony-forming units (cfu)/cm^2 leaf area) in the presence
562 of 0.01% (v/v) Silwet L-77 (Phytotechnology Laboratories) for 15 sec and incubated on
563 0.8% (w/v) tissue-culture agar plates at 21°C under a 16-hr light cycle (70-80 μE m-2 s-
564 1 light intensity) for 3 days. Leaves were then surface-sterilized in 70% ethanol for 10
565 sec, rinsed in sterile water, surface-dried on paper towels, and the bacteria were
566 extracted into water, using an 8-mm stainless steel bead and a ball mill (20 Hz for 3
567 min). Serial dilutions of the extracted bacteria were plated on LB agar plates for
568 colony-forming units (CFU) counting.

569

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

570 **RNA isolation and quantitative PCR (qPCR).** Total RNA was extracted from 9-day-
571 old seedlings using TRIzol reagent (Invitrogen, San Diego, CA) according to the
572 manufacturer's instructions. 2.5 µg of total RNA was reverse-transcribed with 3.75 µM
573 random hexamers (Qiagen, Hilden, Germany) and 20 U of ProtoScript II (New England
574 Biolabs, Boston, MA) according to the manufacturer's instructions. The resulting
575 cDNA:RNA hybrids were treated with 10 U of DNase I (Roche, Basel, Switzerland) for
576 30 min at 37°C and purified on PCR clean-up column (Macherey-Nagel, Düren,
577 Germany). qPCR was performed with Kapa SYBR Fast qPCR master mix (Kapa
578 Biosystems, Wilmington, MA) and CFX384 real-time PCR machine (Bio-Rad, Hercules,
579 CA). The thermal cycling program was as follows: 95°C for 3 min; 45 cycles of 95°C for
580 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
581 sec; and 50 cycles of 0.5°C increments for 10 sec. Biological and technical replicates
582 were performed on the same 384-well PCR plate. Average of the three Ct values per
583 biological replicate was converted to difference in Ct value relative to that of control
584 sample. The Pfaffl method (Pfaffl, 2001) and calculated primer efficiencies were used to
585 determine the relative fold increase of the target gene transcript over the *EIF4A1*
586 (*AT3G13920* or *AL3G26100*) housekeeping gene transcript for each biological
587 replicate. Expression values were then calculated relative to WT un-treated samples.
588 Primer sequences and efficiencies are listed in Supplementary Table 6.
589
590 **Camalexin and 4OH-ICN extraction and LC-DAD-MS.** 10-day-old seedlings were
591 snap-frozen, lyophilized, weighed and homogenized using a 5-mm stainless steel bead
592 and ball mill (20 Hz, 4 min). For phytoalexin analysis, homogenate was extracted with

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

593 300 µL 80% (v/v) aqueous methanol containing 0.08% (v/v) formate and 2.5 µL internal
594 standard (200 µM 4-methoxyindole/4M-I [Sigma-Aldrich] in 100% methanol) per mg
595 sample dry weight. Extracts were sonicated for 5 min and centrifuged at 16,000xg for 2
596 min. The supernatant was filtered using a 0.45-µm polypropylene filter plate (GE
597 Healthcare, Chicago, IL). Samples were separated by reversed-phase chromatography
598 on an Ultimate 3000 HPLC (Dionex, Sunnyvale, CA) system, using a 3.5-µm, 3x150-
599 mm Zorbax SB-Aq column (Agilent, Santa Clara, CA); volume injected was 10 µL. The
600 gradient is shown in Supplementary Table 7. A coupled DAD-3000RS diode array
601 detector (Dionex) collected UV absorption spectra in the range of 190-560 nm, a FLD-
602 311 fluorescence detector (Dionex) collected fluorescence data at 275 nm excitation
603 and 350 nm emission, and a MSQPlus mass spectrometer (Dionex) collected ESI mass
604 spectra in positive and negative ion modes in the range of 100-1000 m/z. Total ICN,
605 4OH-ICN and camalexin amounts were quantified using standard curves of standards
606 prepared in *cyp79B2 cyp79B3* seedling extract and integrated areas in the UV
607 chromatographs at 260-nm for 4M-I (retention time [RT] = 7.7 min); 340-nm for ICN (RT
608 = 11.5 min); 280-nm for ICN degradation product ICA-ME (RT = 9.5 min); and co-
609 eluting 4OH-ICN degradation products 4OH-ICA and 4OH-ICA-ME (RT = 10.1 min);
610 and 320 nm for camalexin (RT = 12.1 min). For Figure 1b, total camalexin amounts
611 were quantified using integrated areas in the FLD chromatograph. For some
612 experiments, 2.5 uL 200 µM indole butrytic acid (IBA; RT = 10.1 min) was added per
613 mg sample dry weight instead of 4M-I. Relative amounts of ICN, 4OH-ICN, and
614 amounts were quantified by dividing the peak areas at m/z 169 [M-H]- (ICN), 174 [M-

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

615 H]- (ICA-ME), 176 [M-H]- (4OH-ICA), 190 [M-H]- (4OH-ICME), and 201 [M+H]+
616 (camalexin), by that of IBA (m/z 202 [M-H]-).
617
618 **Glucosinolate extraction and LC-DAD-FLD-MS.** Glucosinolates were analyzed as
619 desulfoglucosinolates as previously described by Kliebenstein *et al*. (2001) with some
620 modifications. Briefly, a 96-well 0.45 µm PVDF filter plate (EMD Millipore, Billerica, MA)
621 was charged with 45 mg DEAE Sephadex A25 (GE Heathcare) and 300 µL of water per
622 well and equilibrated at room temp for 2 h. Prior to sample homogenization, the plate
623 was centrifuged at 400xg for 1 min to remove the water. The homogenate was
624 extracted with 500 µL 70% (v/v) aqueous methanol at 67.5˚C for 10 min and
625 centrifuged at 16,000xg for 2 min. Added to the supernatant was 3 µL of IS (1.25 mM
626 sinigrin (Sigma-Aldrich) in 80% (v/v) ethanol) per mg sample dry weight. Extract was
627 applied to and incubated on the ion exchanger for 10 min. The sephadex resin was
628 washed three times with 70% (v/v) methanol, three times with distilled deionized water
629 (ddH_2O), and two times with 20 mM sodium acetate (pH 5). 20µL of 25 mg/mL aryl
630 sulfatase (Type H1 from *Helix pomatia*, Sigma-Aldrich) was applied to and incubated
631 on the sephadex resin at RT overnight (Hogge *et al*., 1988). The plate was centrifuged
632 at 400xg for 1 min, and desulfoglucosinolates were eluted from the sephadex resin by
633 two 100-µL washes with 60% (v/v) methanol and two 100-µL washes with ddH_2O.
634 Eluate volume was reduced to 250-350 µL using an evaporator. Samples were
635 separated using the gradient shown in Supplementary Table 7**.** A coupled DAD-
636 3000RS diode array detector, FLD-311 fluorescence detector (Dionex), and MSQPlus
637 mass spectrometer collected UV absorption spectra at 229-nm, fluorescence spectra

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

638 at 275/350-nm (ex/em), and ESI mass spectra in positive/negative ion modes at 100-
639 1000 m/z, respectively. Glucosinolates were quantified using integrated areas of
640 desulfoglucosinolates in the UV chromatographs at 229-nm and published response
641 factors (Clarke, 2010).
642
643 **Chromatin immunoprecipitation and (q)PCR.** ChIP was performed as previously
644 described by Chezem et al. (2017) with some modifications. Approximately two-
645 hundred-and-ten 9-day-old seedlings were inoculated with *Pto* DC3000 *avrRpm1* to
646 OD_{600} of 0.013 and co-treated with mock solution of DMSO (M) or 20 μM
647 dexamethasone (D) for 9 hr. Following nuclear extraction, samples were sonicated in a
648 Covaris S2 sonicator (Covaris, Woburn, MA) using 10% duty, 7% intensity, 200 cycles
649 per burst for a total time of 11 min. Chromatin immunoprecipitation was performed
650 using Anti-FLAG M2 Affinity Gel (Sigma-Aldrich). Beads were pre-treated with 0.1%
651 (w/v) non-fat milk in 1X PBS and 0.5 mg/mL sheared salmon sperm DNA (Invitrogen).
652 Following de-crosslinking, DNA samples were phenol-chloroform-extracted and diluted
653 to a common concentration prior to PCR. 1.5uL immunoprecipitated ChIP-DNA was
654 used in a 15mL PCR reaction. PCR analysis was performed on nuclear extracts prior to
655 antibody incubation (input) and after ChIP. PCR conditions were as follows: 95°C for 3
656 min; 40 cycles of 95°C for 15 sec, 53°C for 15 sec, and 72°C for 1 min; 72°C for 5 min.
657 Densitometric determination of signal intensity in each ChIP and input sample was
658 calculated using ImageJ. Fold enrichment was determined by calculating the ratio of
659 PCR product intensities in ChIP D/M to Input D/M. In cases where amplicons were
660 absent, an arbitrary value of 10 was assigned. For *EPL2*, qPCR analysis was

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

661 additionally performed to confirm absence of amplicons in ChIP samples. RLU counts
662 at the 25th cycle were used for quantification. Primer sequences are listed in
663 Supplementary Table 6**.**
664
665 **Comparative genomics**. All phylogenetic species trees were adapted from Koch and
666 Kiefer (2005) and Couvreur et al. (2009). To generate novel phylogenetic maximum
667 likelihood (ML) trees, sequences were aligned using MUSCLE in MEGA7 (Kumar *et al*.,
668 2016) and JTT model (for CYP82C and LINE alignments) or Tamura-Nei model (for the
669 *EPCOT3* alignment). Sequences for all genes with the description "non-LTR
670 retrotransposon family (LINE)" (N=263) were batch-downloaded from TAIR
671 (https://arabidopsis.org). Of these, sequences containing intact reverse transcriptase
672 domains ("PGPDG", "LIPK", "FRPISL", or "FADD" sequences; N=126) were used for
673 subsequent phylogenetic analysis. Gaps were removed from the CYP82C alignment,
674 leaving a total of 480 codons. *EPCOT3* alignments were visualized in JalView
675 (http://www.jalview.org/; Waterhouse *et al*., 2009). Information on genomes used for
676 synteny analysis is shown in Supplementary Table 8.
677
678 Selection estimates based on nonsynonymous-to-synonymous substitution ratios were
679 calculated from the CYP82C ML tree (Supplementary Text 1). A Newick tree file was
680 generated from this ML tree (Supplementary Figure 4b; Supplementary Table 2) and for
681 Branch site models, branches were pre-defined. CodeML analysis in PAML (Yang,
682 2007) was then conducted with the following modified parameters: ncatG = 8;
683 CodonFreq = 3. The M0 test was performed with model = 0 and NSsites = 0. The M1a

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

null test was performed with model = 0 and NSsites = 1. A more stringent null test (fixed omega) was performed for each Branch site model to be tested (model = 2 and NSsites = 2), where omega was fixed to 1. Branch site models were then tested with unfixed omega. Likelihood ratio tests were performed by comparing critical values and degrees of freedom between each unfixed Branch site test and either the M1a test or the corresponding fixed-omega test. Pre-defined branches with *P* values less than 0.05 for both tests were regarded as under positive selection (Supplementary Figure 2).

The protein structure of CYP82C2 was generated using Intensive modeling mode in Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index; Kelley *et al*., 2015) and visualized in MacPyMOL (Schrödinger, LLC). Amino acid conservation was scored using the Bayesian Best model in Consurf (http://consurf.tau.ac.il/2016/; Ashkenazy *et al*., 2016).

Bioinformatics. Coexpression data was obtained from ATTED-ii (http://atted.jp/; Obayashi et al., 2018). Mutual ranks less than 200 are indicative of strong co-expression (Obayashi et al., 2018). Epigenetics data was obtained from Roudier et al. (2011) and confirmed using data from Liu et al. (2018). Percent identity matrices were constructed from Clustal Omega Multiple Sequence Alignments (https://www.ebi.ac.uk/Tools/msa/clustalo/). Promoter alignment plots were generated using mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml; Frazer et al., 2004)

Data availability

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The authors declare that all data supporting the findings of this study are available within the manuscript and the Supplementary Information or are available from the corresponding authors upon request.

References

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

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

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

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

Birkenbihl, R.P., Kracher, B., Somssich, I.E. Induced Genome-Wide Binding of Three Arabidopsis WRKY Transcription Factors during Early MAMP-Triggered Immunity. *Plant Cell* **29**(1), 20-38 (2017).

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., Innes, R.W. A disease resistance gene in Arabidopsis with specificity for two different pathogen avirulence genes. *Plant Cell* **6**(7), 927-933 (1994).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

728 Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures
729 of enhancer activity during embryonic development. *Nat Genet* **44**(2), 148-156
730 (2012).
731 Bourque, G. Transposable elements in gene regulation and in the evolution of
732 vertebrate genomes. *Curr Opin Genet Dev* **19**(6), 607-612.
733 Böttcher, C. *et al.* The multifunctional enzyme CYP71B15 (PHYTOALEXIN
734 DEFICIENT3) converts cysteine-indole-3-acetonitrile to camalexin in the indole-3-
735 acetonitrile metabolic network of Arabidopsis thaliana. *Plant Cell* **21**(6), 1830-1845
736 (2009).
737 Cabrera, J.C., Messiaen, J., Cambier, P., Van Cutsem, P. Size, acetylation and
738 concentration of chitooligosaccharide elicitors determine the switch from defence
739 involving PAL activation to cell death and water peroxide production in Arabidopsis
740 cell suspensions. *Physiol Plant* **127**(1), 44-56 (2006).
741 Camara, M.D., Ancell, C.A., Pigliucci, M. Induced mutations: a novel tool to study
742 phenotypic integration and evolutionary constraints in *Arabidopsis thaliana*. *Evol*
743 *Ecol Res* 2(8), 1009-1029 (2000).
744 Chae, L., Kim, T., Nilo-Poyanco, R., Rhee, S.Y. Genomic signatures of specialized
745 metabolism in plants. *Science* **344**(6183), 510-513 (2014).
746 Chatfield, S.P. & Raizada, M.M. Ethylene and shoot regeneration: hookless1 modulates
747 de novo shoot organogenesis in *Arabidopsis thaliana*. *Plant Cell Rep* **27**(4), 655-666
748 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

749 Chan, Y.F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent
750 deletion of a *Pitx1* enhancer. *Science* **237**(5963), 302-305 (2010).
751 Chuong, E.B., Elde, N.C., Feschotte, C. Regulatory evolution of innate immunity
752 through co-option of endogenous retroviruses. *Science* **351**(6277), 1083-1087
753 (2016).
754 Clarke, D.B. Glucosinolates, structures and analysis in food. *Anal Methods* **2**(4), 301-
755 416 (2010).
756 Clay, N.K., *et al.* Glucosinolate metabolites required for an Arabidopsis innate immune
757 response. *Science* **323**(5910), 95-101 (2009).
758 Clough, S.J. & Bent, A.F. Floral dip: a simplified method for Agrobacterium-mediated
759 transformation of *Arabidopsis thaliana*. *Plant J* **16**(6), 735-743 (1998).
760 Couvreur, T.L.P. *et al.* Molecular phylogenetics, temporal diversification, and
761 principles of evolution in the mustard family (Brassicaceae) *Mol Biol Evol* **27**(1),
762 55-71 (2009).
763 de Souza, F.S.J., Franchini, L.F., Rubinstein, M. Exaptation of transposable elements
764 into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol*
765 **30**(6), 1239-1251 (2013).
766 Debener T., Lehnackers H., Arnold M., Dangl J.L. Identification and molecular mapping
767 of a single Arabidopsis thaliana locus determining resistance to a phytopathogenic
768 Pseudomonas syringae isolate. *Plant J* **1**(3), 289-302 (1991).
769 Dixon, R.A. & Strack. D. Phytochemistry meets genome analysis and beyond.
770 *Phytochemistry* **62**(6), 815-816 (2003).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

771 Denoux, C. *et al.* Activation of defense response pathways by OGs and Flg22 elicitors
772 in Arabidopsis seedlings. *Mol Plant* **1(**3), 423-445 (2008).
773 Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat Rev*
774 *Genet* **9**(5), 397-405 (2008).
775 Felix G., Duran J.D., Volko S., Boller T. Plants have a sensitive perception system for
776 the most conserved domain of bacterial flagellin. *Plant J* **18**(3),265-276 (1999).
777 Force, A.M. *et al.* Preservation of duplicate genes by complementary, degenerative
778 mutations. *Genet* **151**(4), 1531-1545 (1999).
779 Frazer, K.A. *et al.* VISTA: computational tools for comparative genomics. *Nucleic*
780 *Acids Res* **32**(Web Server issue), W273-W279 (2004).
781 Glawischnig, E., Hansen, B.G., Olsen, C.E., Halkier, B.A. Camalexin is synthesized
782 from indole-3-acetaldoxime, a key branching point between primary and secondary
783 metabolism in Arabidopsis. *Proc Natl Acad Sci USA* **101**(21), 8245-8250 (2004).
784 Grotewold, E. Plant metabolic diversity: a regulatory perspective. *Trends Plant Sci*
785 **10**(2), 57-62 (2005).
786 Hammerschmidt, R. PHYTOALEXINS: What have we learned after 60 years? *Annu Rev*
787 *Phytopathol* **37**, 285-306 (1999).
788 Hardtke, C.S., Müller, J., Berleth, T. Genetic similarity among *Arabidopsis thaliana*
789 ecotypes estimated by DNA sequence comparison. *Plant Mol Biol* 32(5), 915-922
790 (1996).
791 Hartmann, T. From waste products to ecochemicals: fifty years research of plant
792 secondary metabolism. *Phytochemistry* **68**(22-24), 2831-2846 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

793 Heintzman, N.D. *et al.* (2007) Distinct and predictive chromatin signatures of
794 transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**(3),
795 311-318 (2007).
796 Hénaff, E., Vives, C., Desvoyes, B., Chaurasia, A., Payet, J., Gutierrez, C.,
797 Casacuberta, J.M. Extensive amplification of the E2F transcription factor binding
798 sites by transposons during evolution of Brassica species. *Plant J* **77**(6), 852-862
799 (2014).
800 Hoffman, B.G. et al. Locus co-occupancy, nucleosome positioning, and H3K4me1
801 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and
802 liver. *Genome Res* **20**(8), 1037-1061 (2010).
803 Hogge, L.R., Reed, D.W., Underhill, E.W., Haughn, G.W. HPLC separation of
804 glucosinolates from leaves and seeds of *Arabidopsis thaliana* and their
805 identification using thermospray liquid chromatography-mass spectrometry. *J*
806 *Chromatogr Sci* **26**, 551-556 (1988).
807 Hohmann, N., Wolf, E.M., Lysak, M.A., Koch, M.A. A time-calibrated road map of
808 Brassicaceae species radiation and evolutionary history. *Plant Cell* **27**(10), 2770-
809 2784 (2015).
810 Hu, T.T. et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome
811 size change. *Nat Genet* **43**(5), 476-481 (2011).
812 Hugouvieux, V., Barber, C.E., Daniels, M.M. Entry of *Xanthomonas campestris* pv.
813 *campestris* into hydathodes of *Arabidopsis thaliana* leaves: a system for studying

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

early infection events in bacterial pathogenesis. *Mol Plant-Microbe Interact* **11**(6), 537-543 (1998).

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

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

Kelley, L.A. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**(6), 845-858 (2015).

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

835 Kruse, T. *et al.* In planta biocatalysis screen of P450s identifies 8-methoxypsoralen as a
836 substrate for the CYP82C subfamily yielding original chemical structures. *Chem Biol*
837 **15**(2), 149-156 (2008).
838 Kumar, S., Stecher, G., Tamura, K. MEGA7: molecular evolutionary genetics
839 analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**(7), 1870-1874 (2016).
840 Levine, M. & Davidson, E.H. Gene regulatory networks for development. *Proc Natl Acad*
841 *Sci USA* **102**(14), 4936-4942 (2005).
842 Li, G. et al. Dual-level regulation of ACC synthase activity by MPK3/MPK6 cascade and
843 its downstream WRKY transcription factor during ethylene induction in Arabidopsis.
844 *PLoS Genet* **8**(6), e1002767 (2012).
845 Liu, S. *et al.* Negative regulation of ABA signaling by WRKY33 is critical for Arabidopsis
846 immunity towards *Botrytis cinerea* 2100. *eLife* **4**, e07295 (2015).
847 Liu, Y. *et al.* PCSD: a plant chromatin state database. *Nucleic Acids Res* **46**(D1),
848 D1157-D1167 (2018).
849 Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H. Reverse transcription of
850 R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for
851 non-LTR retrotransposition. *Cell* **72**(4), 595-605 (1993).
852 Lynch, M. The lower bound to the evolution of mutation rates. *Genome Biol Evol* 3,
853 1107-1118 (2011).
854 Malik, H.S., Burke, W.D., Eickbush, T.H. The age and evolution of non-LTR
855 retrotransposable elements. *Mol Biol Evol* **16**(6), 793-805 (1999).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

856 Mansfield, J.W. Antimicrobial compounds and resistance: the role of phytoalexins and
857 phytoanticipins. In Mechanisms of Resistance to Plant Diseases, A. Slusarenko,
858 RSS Fraser and LC. van Loon eds (Kluwer Academic Publishers Dordrecht, The
859 Netherlands), pp. 325-370 (2000).
860 Martin, C., Ellis, N., Rook, F. Do transcription factors play special roles in adaptive
861 variation? *Plant Physiol* **154**(2), 506-511 (2010).
862 McClintock, B. Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol*
863 **21**, 197-216 (1956).
864 McNellis, T.W. *et al.* Glucocorticoid-inducible expression of a bacterial avirulence
865 gene in transgenic Arabidopsis induces hypersensitive cell death. *Plant J* **14**(2),
866 247-257 (1998).
867 Moore, R.C. & Purugganan, M.D. The evolutionary dynamics of plant duplicate genes.
868 *Curr Opin Plant Biol* **8**(2), 122-128 (2005).
869 Mukherjee, A.K., Lev, S., Gepstein, S., Horwitz, B.B. A compatible interaction of
870 *Alternaria brassicicola* with *Arabidopsis thaliana* ecotype DiG: evidence for a specific
871 transcriptional signature. *BMC Plant Biol* **9**(1), 31 (2009).
872 Müller, A.A. Die Induktion von rezessiven Letalmutationen durch Äthylmethansulfonat
873 beiArabidopsis. *Theor Appl Genet* **36**(5), 201-220 (1966).
874 Murashige, T. & Skoog, F. A revised medium for rapid growth and bio assays with
875 tobacco tissue cultures. *Physiol Plant* **15**(3), 473–497 (1962).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

876 Murgia, I., Tarantino, D., Soave, C., Morandini, P. Arabidopsis *CYP82C4* expression is
877 dependent on Fe availability and circadian rhythm, and correlates with genes
878 involved in the early Fe deficiency response. *J Plant Physiol* **168**(9), 894-902 (2011).
879 Nafisi, M. *et al.* Arabidopsis cytochrome P450 monooxygenase 71A13 catalyzes the
880 conversion of indole-3-acetaldoxime in camalexin synthesis. *Plant Cell* **19**(6), 2039-
881 2052 (2007).
882 Navarro, L. *et al.* The transcriptional innate immune response to flg22. Interplay and
883 overlap with Avr gene-dependent defense responses and bacterial pathogenesis.
884 *Plant Physiol* **135**(2), 1113-1128 (2004).
885 Obayashi, T. *et al.* ATTED-II in 2018: A plant coexpression database based on
886 investigation of statistical property of Mutual Rank Index. *Plant Cell Physiol* **59**(1),
887 e3 (2018).
888 Ohno, S. Evolution by Gene Duplication. Springer-Verlag: Heidelberg, Germany (1970).
889 Omranian, N. et al. Differential metabolic and coexpression networks of plant
890 metabolism. *Trends Plant Sci* **20**(5), 266-268 (2015).
891 Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-
892 PCR. *Nucleic Acids Res* **29**(9), e45 (2001).
893 Povero, G. *et al.* Transcript profiling of chitosan-treated Arabidopsis seedlings. *J Plant*
894 *Res* **124**(5), 619-629 (2011).
895 Prud'homme, B., Gompel, N., Carroll, S.B. Emerging principles of regulatory evolution.
896 *Proc Natl Acad Sci USA* **104**(Suppl 1), 8605-8612 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Qiu, J.L. *et al.* Arabidopsis MAP kinase 4 regulates gene expression through transcription factor release in the nucleus. *EMBO J* **27(**16), 2214-2221 (2008).

Rajniak, J., Barco, B., Clay, N.K., Sattely, E.S. A new cyanogenic metabolite in Arabidopsis required for inducible pathogen defence. *Nature* **525**(7569), 376-379 (2015).

Rajniak, J. *et al.* Biosynthesis of redox-active metabolites in response to iron deficiency in plants. *Nat Chem Biol* **14**(5), 442-450 (2018).

Rédei, G.P. Single locus heterosis. *Zeitschrift für Vererbungslehre* **93**(1), 164-170 (1962).

Rédei, G.P. & Koncz, C. Classical mutagenesis. In Methods in Arabidopsis Research, pp. 16-82 (1993).

Rinerson, C.I. *et al.* The evolution of WRKY transcription factors. *BMC Plant Biol* **15**, 66 (2015).

Rogers, W.A., *et al.* Recurrent modification of a conserved *cis*-regulatory element underlies fruit fly pigmentation diversity. *PLoS Genet* **9**(8), e1003740 (2013).

Roudier, F. *et al.* Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J* **30**(10), 1928-1938 (2011).

Rushton, P.J., Somssich, I.E., Ringler, P., Shen, Q.J. WRKY transcription factors. *Trends Plant Sci* **15**(5), 247-258 (2010).

Schluttenhofer, C. &, Yuan, L. (2015) Regulation of specialized metabolism by WRKY transcription factors. *Plant Physiol* **167**(2), 295-306 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Slattery, M. *et al.* Absence of a simple code: how transcription factors read the genome. *Trends Biochem Sci* **39**(9), 381-399 (2014).

Slotkin, R. K, & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**(4), 272 (2007).

Spitz, F. & Furlong, E.E. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**(9), 613-626 (2012).

Tao, Y. *et al.* Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell* **15**(2), 317-330 (2003).

Thomma, B.P., Nelissen, I., Eggermont, K., Broekaert, W.F. Deficiency in phytoalexin production causes enhanced susceptibility of *Arabidopsis thaliana* to the fungus *Alternaria brassicicola. Plant J* **19**(2), 163-171 (1999).

Tsuji, J. et al. (1992) Phytoalexin accumulation in *Arabidopsis thaliana* during the hypersensitive reaction to *Pseudomonas syringae* pv *syringae*. *Plant Physiol* **98**(4), 1304-1309 (1992).

Tohge, T. & Fernie, A.R. Co-expression and co-responses: within and beyond transcription. *Front Plant Sci* **3**, 248 (2012).

Wang, Y., Li, X., Hu, H. H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics* **103**(2), 222-228 (2014).

Waterhouse, A.M. *et al.* Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**(9), 1189-1191 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

939 Weng, J.K., Philippe, R.N., Noel, J.P. (2012) The rise of chemodiversity in plants.
940 *Science* **336**(6089), 1667-1670 (2012).
941 Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements.
942 *Nat Rev Genet* **8**(12), 973-982 (2007).
943 Wink, M. Evolution of secondary metabolites from an ecological and molecular
944 phylogenetic perspective. *Phytochemistry* **64**(1), 3-19 (2003).
945 Wittkopp, P.J. & Kalay, G. *Cis*-regulatory elements: molecular mechanisms and
946 evolutionary processes underlying divergence. *Nat Rev Genet* **13**(1), 59-69 (2012).
947 Wray, G.A. The evolutionary significance of *cis*-regulatory mutations. *Nat Rev Genet*
948 **8**(3), 206-216 (2007).
949 Zhao, Y. *et al.* (2002) Trp-dependent auxin biosynthesis in Arabidopsis: involvement of
950 cytochrome P450s CYP79B2 and CYP79B3. *Genes Dev* **16**(23), 3100-3112 (2002).
951 Zheng, Z., Qamar, S.A., Chen, Z., Mengiste, T. (2006) Arabidopsis WRKY33
952 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant*
953 *J* **48** (4), 592-605 (2006).
954 Zhou, J., Wang, J., Zheng, Z., Fan, B., Yu, J. Q., Chen, Z. Characterization of the
955 promoter and extended C-terminal domain of Arabidopsis WRKY33 and functional
956 analysis of tomato WRKY33 homologues in plant stress responses. *J Exp Bot*
957 **66**(15), 4567-4583 (2015).
958 Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D., Boller, T., & Felix, G.
959 Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts
960 Agrobacterium-mediated transformation. *Cell* **125**(4), 749-760 (2006).

bioRxiv preprint doi: https://doi.org/10.1101/368340; this version posted October 20, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

961

962 **Acknowledgements**

963 We thank E.S. Sattely for ICN/ICN-ME, 4OH-ICA/4OH-ICA-ME and camalexin
964 standards. This work was supported by T32-GM007499 (to B.B.) and
965 Elsevier/Phytochemistry Young Investigator Award (to N.K.C.).

966

967 **Author Contributions**

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

971

972 **Competing Interests**

973 The authors declare no competing interests.

974

975 **Materials & Correspondence**

976 Correspondence and material requests can be addressed to Brenden Barco.