HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2016 March 17.

Published in final edited form as:

Nature. 2015 September 17; 525(7569): 376–379. doi:10.1038/nature14907.

A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defense

Jakub Rajniak[1], **Brenden Barco**[2], **Nicole K. Clay**[2,*], and **Elizabeth S. Sattely**[1,*]

[1]Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA

[2]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT, 06511, USA

SUMMARY

Thousands of putative biosynthetic genes in *Arabidopsis thaliana* have no known function, suggesting that there are numerous molecules contributing to plant fitness that have not yet been discovered[1,2]. Prime among these uncharacterized genes are cytochromes P450 upregulated in response to pathogens[3,4]. Starting with a single pathogen-induced P450[5], CYP82C2, we used a combination of untargeted metabolomics and co-expression analysis to uncover the complete biosynthetic pathway to a previously unknown *Arabidopsis* metabolite, 4-hydroxyindole-3-carbonyl nitrile (4-OH-ICN), which harbors cyanogenic functionality that is unprecedented in plants and exceedingly rare in nature[6,7]. The aryl cyanohydrin intermediate in the 4-OH-ICN pathway reveals a latent capacity for cyanogenic glucoside biosynthesis[8,9] in *Arabidopsis*. By expressing 4-OH-ICN biosynthetic enzymes in *Saccharomyces cerevisiae* and *Nicotiana benthamiana*, we reconstitute the complete pathway *in vitro* and *in vivo* and validate the functions of its enzymes. 4-OH-ICN pathway mutants show increased susceptibility to the bacterial pathogen *Pseudomonas syringae*, consistent with a role in inducible pathogen defense. *Arabidopsis* has been the preeminent model system[10,11] for studying the role of small molecules in plant innate immunity[12]; our results uncover a new branch of indole metabolism distinct from the canonical camalexin pathway, and support a role for this pathway in the *Arabidopsis* defense response.[13] These results establish a more complete framework for understanding how the model plant *Arabidopsis* uses small molecules in pathogen defense.

To identify P450s potentially involved in the biosynthesis of novel defense-associated small molecules, we obtained raw datasets for all transcriptomics experiments dealing with biotic stress in *A. thaliana* from the NASCArrays database. We examined *CYP* genes present in the probeset and selected a candidate, *CYP82C2*, that is highly expressed under a variety of

---

*Correspondence: sattely@stanford.edu, nicole.clay@yale.edu.

AUTHOR CONTRIBUTIONS:

J.R., B.B., N.K.C., and E.S.S designed experiments. J.R. and B.B. performed experiments. J.R., B.B., N.K.C., and E.S.S. analyzed data and wrote the paper.

COMPETING INTERESTS:

The authors declare no competing interests.

pathogen treatment conditions, but whose native function in *Arabidopsis* is unknown (Fig. 1A).

To identify small molecules whose levels change in a CYP82C2-dependent manner, we performed comparative metabolomics[14] with a homozygous T-DNA insertion line of *CYP82C2*. We used the bacterial pathogen *Pseudomonas syringae* pv. *tomato* DC3000 harboring the *avrRpm1* avirulence gene (*Psta*) as an elicitor since *CYP82C2* expression is strongly upregulated 24 h after inoculation with this strain (Fig. 1A). We analyzed tissue methanolic extracts of 11-day-old seedlings grown hydroponically in the presence of *Psta* by liquid chromatography-mass spectrometry (LC-MS), and computationally compared mutant and wild-type (WT) Col-0 metabolomes. From this analysis, we identified 11 compound mass signals that reproducibly and significantly differ between WT and *cyp82C2* (Fig. 1B); these mass ions are induced after pathogen elicitation and are not bacterially derived (Extended Data Fig. 1A).

We next sought to obtain clues about the structure of these compounds from their tandem mass spectra (MS/MS). MS/MS analysis revealed that the 11 compounds could be divided into two classes (A and B in Fig. 1B), assigned as indole-3-carboxaldehyde (IAL) derivatives with (B) and without (A) hydroxylated indole systems. Moreover, the fact that the *cyp82C2* mutant lacked all the hydroxylated derivatives but accumulated excess amounts of their non-hydroxylated counterparts suggested that CYP82C2 acts as an indolic hydroxylase. However, except for compound **A1** (Fig. 2B), which was confirmed to be indole-3-carboxylic acid methyl ester, the structures of these compounds remained elusive.

To facilitate structural analysis, we investigated whether any of these compounds were exuded into the medium in the *cyp82C2* mutant seedling experiments (Fig. 1D). Filtered spent medium was loaded onto a C18 silica gel cartridge, and non-polar metabolites were eluted with acetonitrile and analyzed by LC-MS. Surprisingly, the profile of spent medium extracted in this manner was notably different from that of tissue methanolic extracts: while small amounts **A2-A7** were present, no **A1** could be detected; instead, a new UV-active compound with m/z = 171.0553 $[M+H]^+$ dominated the LC-MS trace (Fig. 1D). NMR analysis of this compound followed by comparison with a synthetic standard established its identity as the novel metabolite indole-3-carbonyl nitrile (ICN) (Fig. 1C and Extended Data Fig. 2).

Chemically, the most striking feature of ICN is the presence of a highly reactive α-ketonitrile moiety that, to our knowledge, has not been found in any plant natural product; however, benzoyl cyanide has been previously identified in the secretions of millipedes[6,7]. The α-ketonitrile is susceptible to nucleophilic attack, resulting in the displacement of cyanide ion: in alkaline aqueous solution, ICN degrades to indole-3-carboxylic acid (ICA)[15]; in methanol, ICA methyl ester (**A1**) is formed instead, explaining the presence of **A1** and the absence of ICN in methanolic extracts (Fig. 1C). Modifying the tissue extraction procedure by using an acidified 1:1 acetonitrile/water mixture enabled direct detection of ICN by LC-MS; additionally, when deuterated methanol was used, only the deuterated form of **A1** was observed (Extended Data Fig. 1, B to E). Based on its molecular formula and the synthesis of an authentic standard, **A6** was shown to be a serine-ICN addition product (see Fig. 2B).

However, in the presence of cysteine and structurally related compounds, ICN can undergo a spontaneous cycloaddition, resulting in the formation of a thiazoline ring and the net loss of ammonia. This latter observation allowed us to determine the structures of and synthesize standards for compounds **A2-A5**, which are the cycloaddition products of ICN and cysteine (**A4**) or Cys-Gly dipeptide (**A2**) and their thiazole analogs (**A5** and **A3**, respectively, see Fig. 2B, Extended Data Fig. 3 and SI Table 1).

The absence of the hydroxylated analogs **B1-B6** in the *cyp82C2* insertion line pointed to ICN as the likely substrate for this enzyme. Incubation of ICN with yeast-expressed CYP82C2 yielded only a trace amount of hydroxylated ICN, but a significant amount of 4-hydroxy-ICA (4-OH-ICA) (structure shown in Fig. 3), as confirmed by NMR spectroscopy and comparison with a synthetic standard (Extended Data Fig. 4, A to D). Since CYP82C2 shows no activity on ICA, we deduced that CYP82C2 converts ICN to 4-hydroxy-ICN (4-OH-ICN), competing with hydrolysis of ICN to ICA (Extended Data Fig. 4, E and F). Further experiments with chemically synthesized 4-OH-ICN showed that its half-life is approximately 3 minutes in aqueous solution at pH = 7.5 (SI Table 2), rendering direct isolation of the 4-OH-ICN product infeasible. Chemical synthesis of 4-OH-ICN further enabled the synthesis of the 4-hydroxy derivatives of **A1-A6**, confirming that these correspond to compounds **B1-B6** seen in WT tissue extracts (Extended Data Fig. 3 and SI Table 1). Therefore, all the metabolites identified in our initial metabolomics experiment with *cyp82C2* are ultimately derived from ICN, whether as artifacts of the extraction (**A1** and **B1**), or as *in vivo* addition products (**A2-A7**, **B2-B6**).

We next investigated the biosynthesis of ICN, using the *CYP82C2* gene as bait for coexpression analysis. For our pathogen dataset, the *CYP79B2* gene, whose encoded enzyme converts tryptophan (Trp) into indole-3-acetaldoxime (IAOx)[16], has the second highest correlation (Pearson's *r*) with *CYP82C2* among all genes profiled (SI Table 3). We performed a metabolomic analysis of the *cyp79B2 cyp79B3* double knockout line[17], which is deficient in IAOx production. No ICN-derived metabolites are produced in this mutant (Fig. 2A), indicating that ICN is derived from IAOx.

In searching for the enzyme(s) responsible for further conversion of IAOx to ICN, we postulated a biosynthetic route paralleling that of the cyanogenic glycoside dhurrin[8]: a CYP79-catalyzed formation of an aldoxime, followed by a CYP71-catalyzed formation of a cyanohydrin intermediate. In the dhurrin pathway, the cyanohydrin is glucosylated to yield the final product, whereas in ICN biosynthesis, a final dehydrogenation is required to produce an α-ketonitrile (Fig. 2C).

Correlation analysis implicated *CYP71A12*, a P450 linked to camalexin biosynthesis[18], as the most likely candidate gene for the cyanohydrin formation step (SI Table 3). Profiling of the *cyp71A12* T-DNA insertion line, as well as T-DNA insertion lines of its two closest *Arabidopsis* homologs, *CYP71A13* and *CYP71A18*, demonstrated that the *CYP71A12* gene is in fact likely responsible: all ICN derivatives with the exception of **A6** are at ~10% of WT levels in the *cyp71A12* mutant, but unaffected in the *cyp71A13* and *cyp71A18* mutants (Fig. 2A). Levels of camalexin and other indolic metabolites were only slightly changed in whole-seedling tissue extracts of the *cyp71A12* mutant (Extended Data Fig. 5C).

Further correlation analysis using *CYP71A12* as bait revealed a cluster of five tandemly arrayed homologous genes, *At1g26380-At1g26420*, that are highly coexpressed with *CYP71A12* (SI Table 3). At1g26380 encodes a flavin-dependent oxidoreductase known as *FOX1*[19]. We profiled the corresponding homozygous T-DNA insertion lines for these genes and found a 3- to 5-fold reduction in levels of ICN metabolites in the *fox1* mutant, with no significant changes observed for the other mutants (Fig. 2A). Additionally, we observed a build-up of IAL, the expected hydrolysis product of the indole-3-cyanohydrin intermediate (Extended Data Fig. 5D). More strikingly, the *fox1* mutant accumulates new mass signals corresponding to indole cyanogenic glycosides, not previously observed in plants (Extended Data Fig. 6, A to E, structures shown in Fig. 4D). Cyanogenic glycoside compounds are widely distributed in the plant kingdom, but have not yet been detected in *Arabidopsis*.[9] Disruption of the ICN pathway at the FOX1-catalyzed step therefore leads to capture of some portion of the cyanohydrin intermediate by non-specific glycosyltransferases, exactly paralleling dhurrin synthesis[8].

We sought to confirm the proposed biochemical transformations (Fig. 2C) by reconstituting the complete pathway *in vitro*. A combination of yeast microsomal CYP71A12 and CYP82C2 and *N. benthamiana*-expressed FOX1 was sufficient to catalyze the conversion of IAOx to ICN, as illustrated in Figure 3; the production of 4-OH-ICN is inferred from the accumulation of 4-OH-ICA. We also reconstituted the biosynthesis of 4-OH-ICN in the heterologous host *Nicotiana benthamiana*, using transient expression of the four pathway genes necessary for production of 4-OH-ICN from Trp via *Agrobacterium*-mediated transient transformation[20]. We observed significant accumulation of **B1** (from methanol extraction of 4-OH-ICN) only when all pathway genes were present; however, we also noted background levels of ICA and IAL when only early pathway genes were expressed (Extended Data Fig. 7). Notably, when we expressed *CYP79B2* and *CYP71A12* but not *FOX1*, we again observed the accumulation of indole cyanogenic glycoside mass signals (Extended Data Fig. 6F).

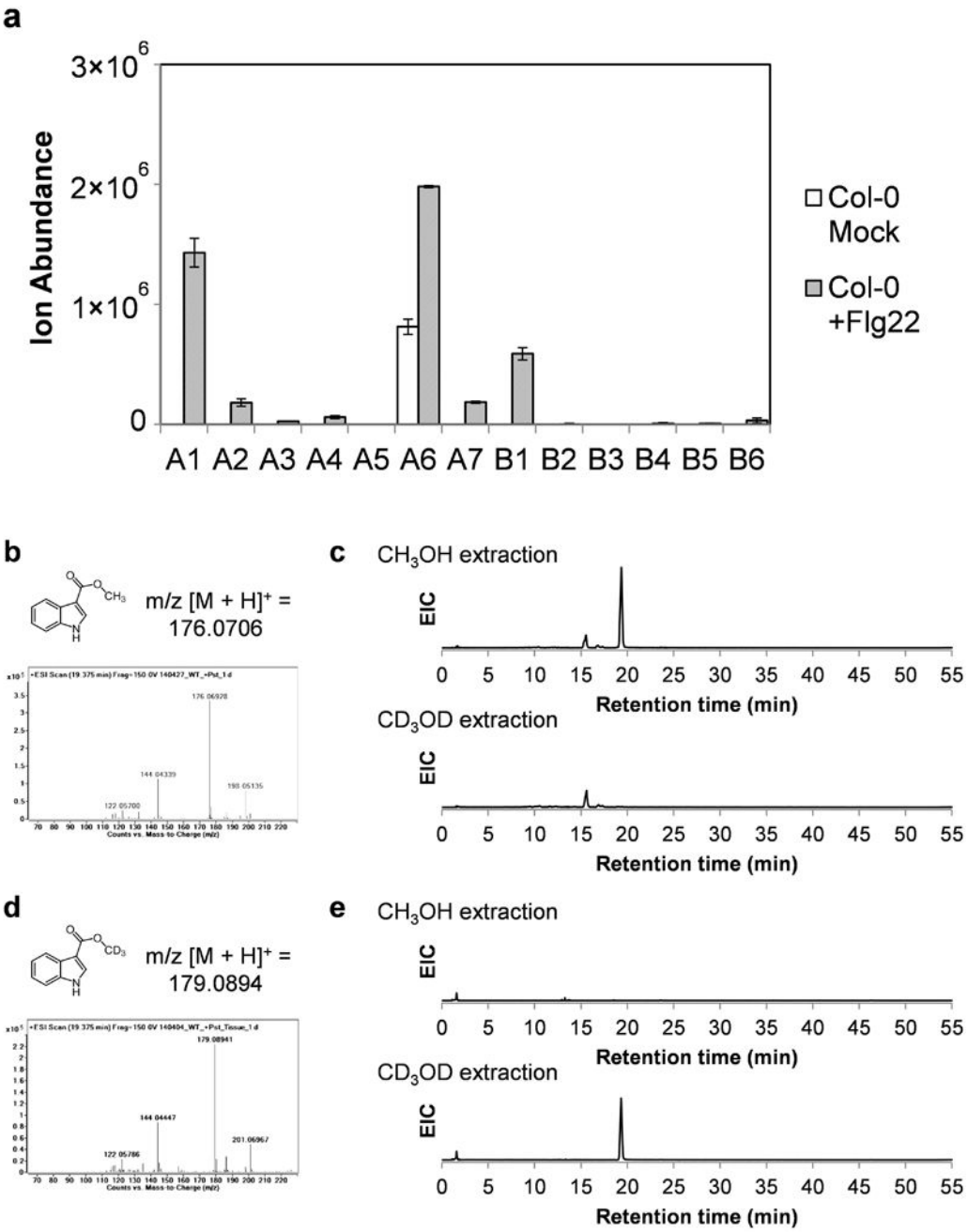
The Trp-derived metabolites camalexin and 4-methoxy indol-3-ylmethylglucosinolate (4-methoxyglucobrassicin) have been shown to play a key role in *Arabidopsis* immunity (Fig. 4D)[10,11,18,21]. To evaluate whether 4-OH-ICN pathway products also contribute to *Arabidopsis* disease resistance, we challenged 4-OH-ICN biosynthetic mutants with a diverse panel of pathogens. Using surface inoculation to mimic the natural infection process, we found that compared to WT, the adult leaves of *cyp71A12* and *cyp82C2* are more susceptible to the virulent bacterial hemibiotroph *Pst* (*Pseudomonas syringae* pv. *tomato* DC3000) and comparable to the immuno-deficient *fls2* mutant, which cannot perceive the bacterial microbe-associated molecular pattern (MAMP) flg22[22,23] (Fig. 4A). Similarly, seedlings of the 4-OH-ICN pathway mutants are more susceptible to *Pst* compared to WT in the presence and absence of flg22 (Fig. 4B), indicating a role for 4-OH-ICN in basal disease resistance against a bacterial pathogen. Notably, the adult leaves and seedlings of the camalexin pathway mutants *cyp71A13* and *pad3* are also more susceptible to *Pst* infection compared to WT (Figure 4A-B), suggesting a previously unrecognized role for camalexin in the antibacterial defense response. To test for a direct role of the ICN pathway metabolites in the plant innate immune response, either as inducible antibacterial or signaling

compounds, we measured their protective effect against subsequent bacterial infection by infecting WT adult leaves with *Pst* after pre-immunizing them with pure compounds and flg22. In comparison to a solvent control, pre-treatment with 4-OH-ICN (but not ICN or camalexin) conferred greater bacterial resistance (Fig. 4C), which supports a direct mechanism of action for 4-OH-ICN in inducible plant defense.

We also observed increased disease symptoms in adult leaves of the *cyp82C2* mutant upon inoculation with spores from the avirulent fungal necrotroph *Alternaria brassicicola* (Extended Data Fig. 8, E and F) and – consistent with a previous report[24] – the virulent necrotroph *Botrytis cinerea* (Extended Data Fig. 8, A and B), but not from the obligate fungal biotroph *Golovinomyces orontii* (Extended Data Fig. 8, C and D). Furthermore, purified ICN and 4-OH-ICN have a growth inhibitory effect on *B. cinerea* and *A. brassicicola* comparable to that of camalexin[25] (Extended Data Fig. 9). However, we cannot rule out the possibility that the role of the 4-OH-ICN pathway in fungal defense is indirect, as adult leaves of the *cyp82C2* mutant appear partially impaired in camalexin production following *Alternaria* treatment (Extended Data Fig. 10).

The camalexin and 4-OH-ICN pathways rely on a pair of paralogous genes, *CYP71A12* and *CYP71A13*, that are members of the CYP71 family linked to innovations in plant metabolism[26] (Fig. 4D). Strikingly, the 4-OH-ICN pathway resembles the widespread cyanogenic glucoside pathway that has been lost in *Brassicaceae*, and appears to be a metabolic re-invention leading to a novel cyanogenic metabolite type derived from Trp.[27,28] It is possible that 4-OH-ICN acts in concert with other Trp-derived metabolites, each contributing to protection against overlapping sets of specific pathogens. Collectively, our data provide additional insight into the *Arabidopsis* defense response and, more generally, how plants use metabolic innovation to expand innate immunity.

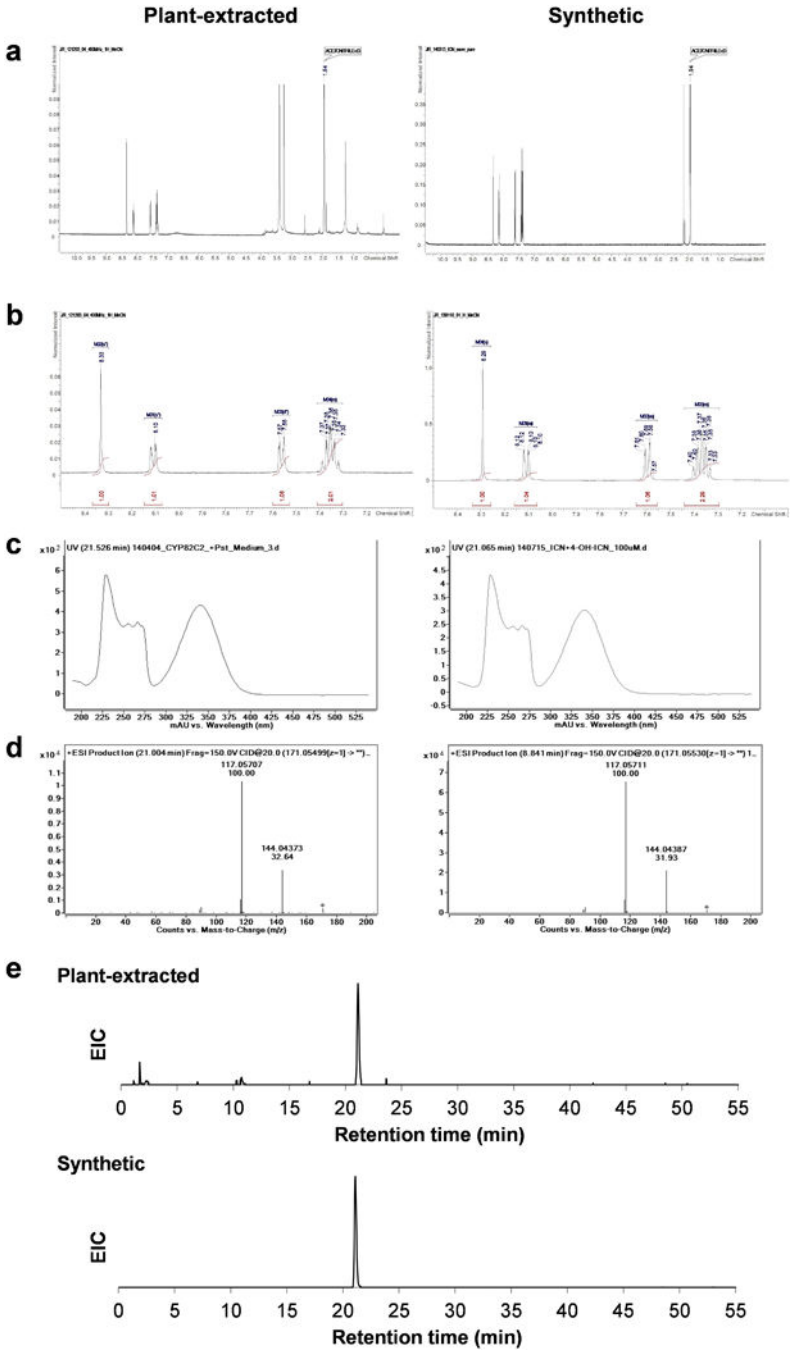
Extended Data



Extended Data Figure 1. Elicitation of compounds identified in metabolomics screen by Flg22 peptide and origin of ICA methyl ester as artifact of the methanol extraction method

(**A**) Levels of compounds in Flg22-elicited *Arabidopsis* Col-0 seedling tissue, quantified as mean $[M+H]^+$ ion (m/z ± 10 ppm) abundances extracted from raw data, with error bars indicating ± standard deviation based on 3 biological replicates. Production of these compounds in axenic plant culture demonstrates that they are plant-derived. (**B**) Structure and mass peaks of ICA methyl ester (compound **A1**) seen in LC-MS analysis, and (**C**) EICs

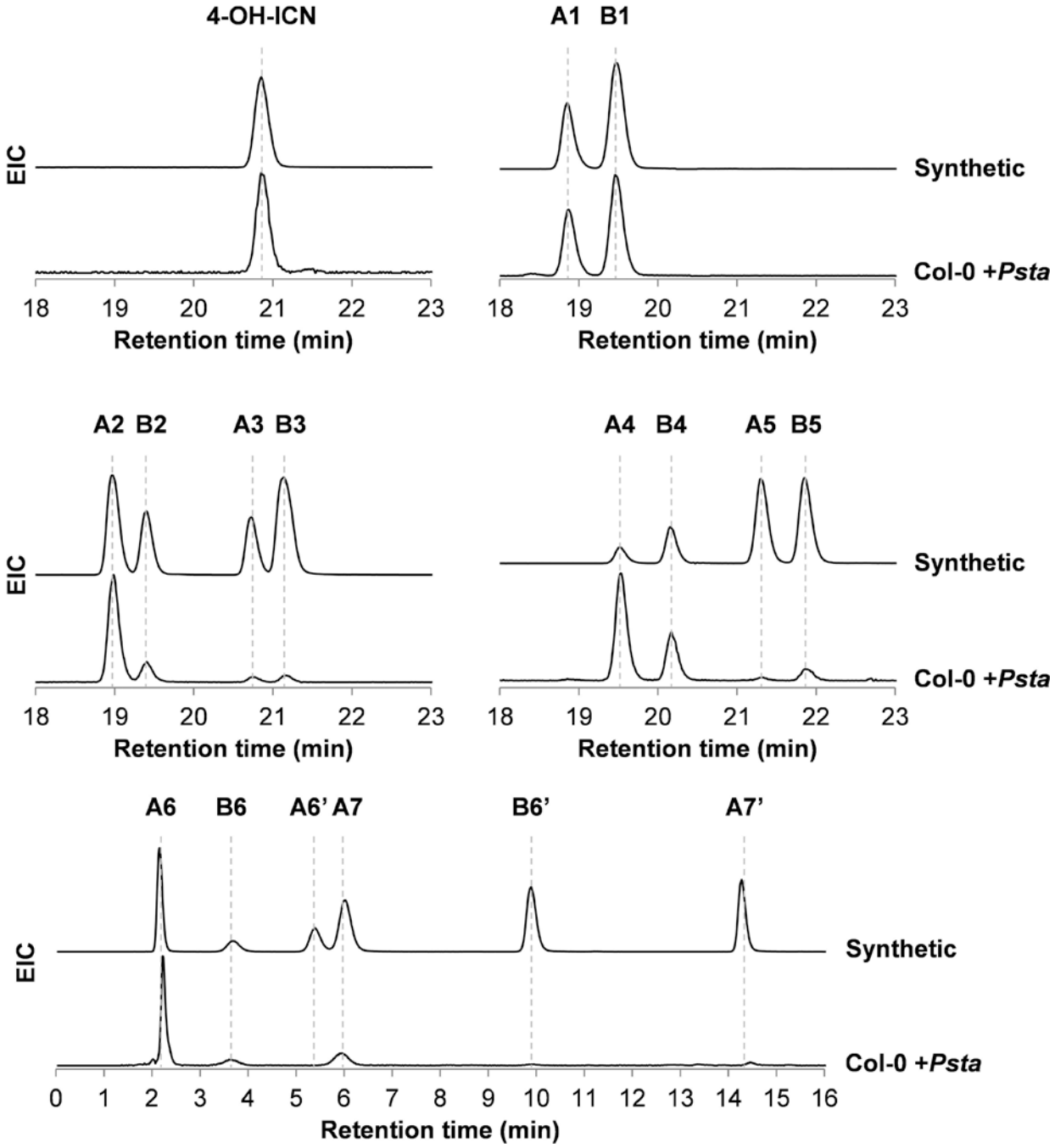
for the expected m/z using a standard extraction with 80:20 CH_3OH/H_2O or with 80:20 CD_3OD/D_2O. (**D**) Structure and mass spectrum peaks seen for the triply deuterated **A1** analog, and (**E**) EICs for the expected m/z using extraction with 80:20 CH_3OH /H_2O, or with 80:20 CD_3OD/D_2O (all EICs are to scale). The presence of the deuterated analog of ICA methyl ester and the complete absence of the non-deuterated compound in plant extracts when CD_3OD is substituted for CH_3OH show that the methyl ester is not a product of *Arabidopsis* metabolism, but arises due to the extraction method as a degradation product of ICN.



Extended Data Figure 2. Comparison of spectra for plant-extracted and synthetic compound establishes identity of ICN as new indolic metabolite produced by *A. thaliana*

(**A**) Full range (δ 10.5 to −0.5) and (**B**) downfield region partial (δ 8.5 to 7.0) ^{1}H NMR spectra in CD_3CN. Upfield contaminants in the full range spectra are presumed to be residual solvent. (**C**) UV/Vis absorbance spectra obtained via a diode array detector during liquid chromatography (LC) analysis. Note that the prominent peak at 230 nm is due to acetonitrile in the LC mobile phase. (**D**) Targeted MS/MS spectra for the parent ICN [M +H]$^{+}$ ion (m/z = 171.0550) at a 20 V collision energy. See SI Table 1 for relative peak

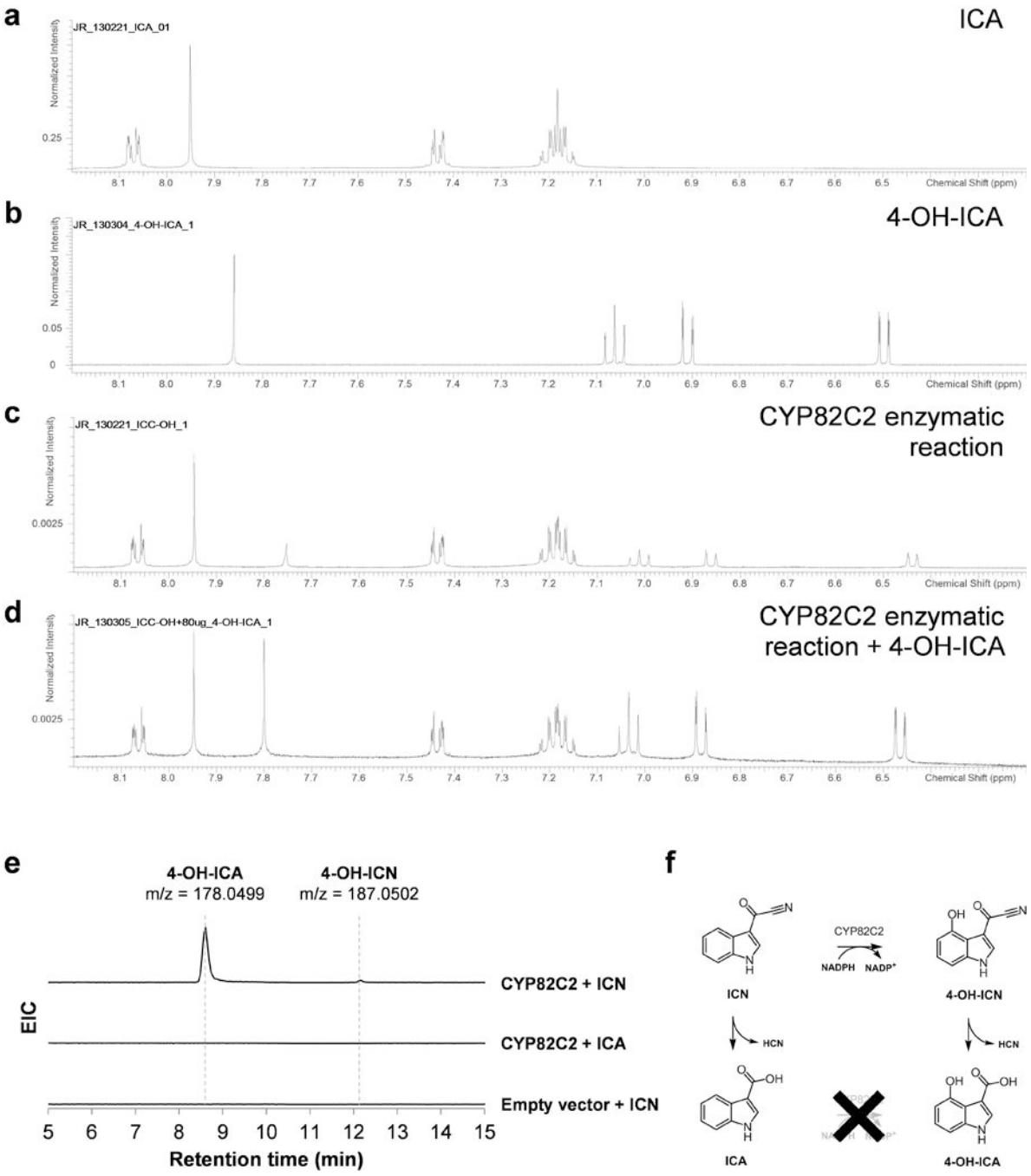
intensities at other collision energies. (**E**) Aligned EICs for the ICN $[M+H]^+$ ion for a Col-0 +*Psta* tissue sample extracted with DMSO and synthetic compound, showing identical retention times.



Extended Data Figure 3. Comparison of plant-extracted ICN derivatives, 4-OH-ICN derivatives and synthetic standards shows identical column elution times for all compounds

Col-0 +*Psta* combined EICs were extracted for the relevant compound $[M+H]^+$ m/z values for a DMSO-extracted medium sample (4-OH-ICN trace), or a MeOH-extracted seedling tissue sample (all other traces), while synthetic EICs were extracted for a mixed standard in

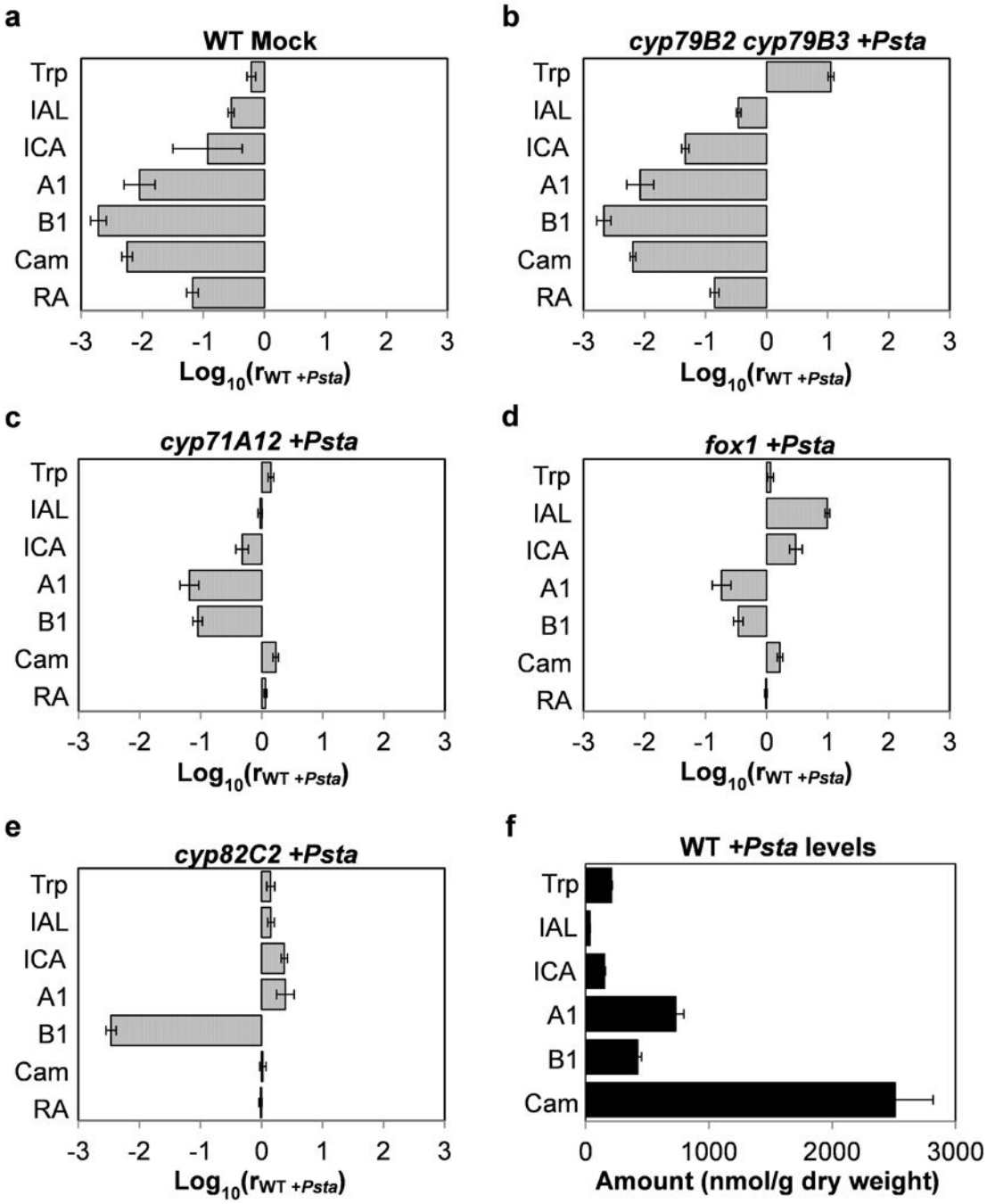
DMSO. Note that chromatograms are not to scale, and the synthetic standard is not equimolar with respect to all compounds due to partial degradation.



Extended Data Figure 4. CYP82C2 is an ICN 4-hydroxylase

^{1}H NMR spectra in CD_3OD of synthetic ICA (**A**) and 4-OH-ICA (**B**). (**C**) Spectrum for large scale enzymatic reaction extract of ICN incubated with CYP82C2. In addition to ICA, resulting from hydrolysis of ICN, peaks for a singly hydroxylated analog of ICA are seen; these are qualitatively consistent with, but shifted slightly upfield (~30–60 Hz) from, the 4-

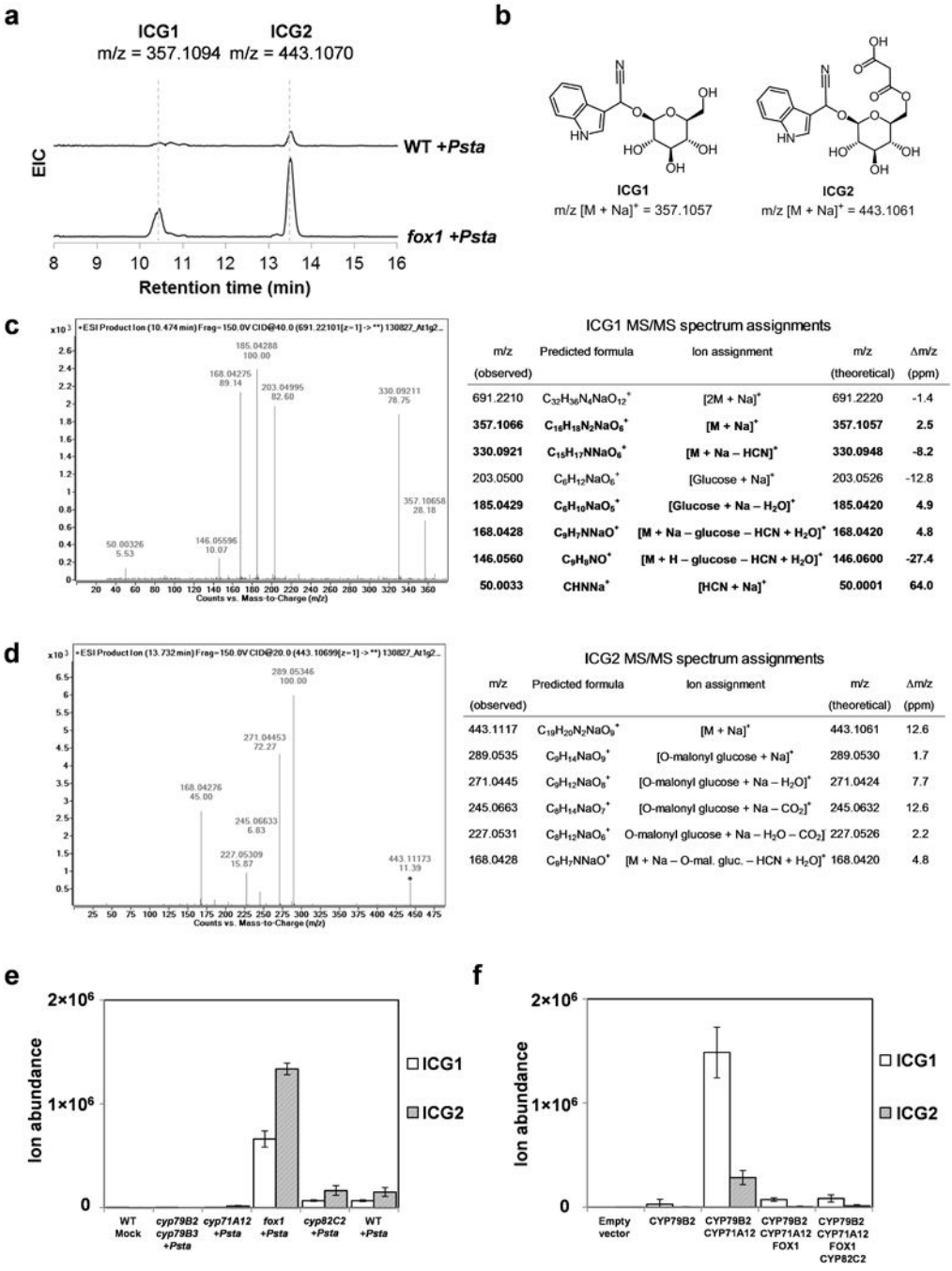
OH-ICA spectrum, possibly due to impurities or a pH effect in the enzymatic reaction sample. (**D**) To confirm the identity conclusively, 80 μg of 4-OH-ICA dissolved in CD_3OD was added to the enzymatic reaction NMR sample prior to acquiring another spectrum: no new peaks are seen, while the prior hydroxylated ICA peaks grow in intensity, establishing the product of the enzymatic reaction as 4-OH-ICA. (**E**) EICs for enzymatic reactions of CYP82C2 on ICN or ICA, or empty vector control incubation with ICN. Only trace amounts of the expected 4-OH-ICN product but significant amounts of 4-OH-ICA are seen for the CYP82C2/ICN reaction. No hydroxylated products are seen for the CYP82C2/ICA or empty vector/ICN reactions, indicating that CYP82C2 catalyzes only the hydroxylation of ICN to 4-OH-ICN, but 4-OH-ICA is seen as the predominant end product due to rapid hydrolysis of 4-OH-ICN (**F**). Chromatograms in this figure were obtained using the 20 min LC-MS gradient (see Methods section 1.9 LC-MS analysis).



Extended Data Figure 5. Levels of numerous *Arabidopsis* indolic metabolites are altered in ICN pathway gene insertion lines compared to WT plants

(**A**-**E**) Relative compound levels for mock treatment condition and indicated pathway insertion line mutants, and (**F**) absolute levels in *Psta*-treated WT (Col-0) seedlings. For panels **A-E**, data bars represent a logarithmically scaled ratio of mean metabolite levels in the indicated line or treatment condition, quantified as $[M+H]^+$ ion abundances by LC-MS analysis with XCMS processing, to levels in *Psta*-treated WT *Arabidopsis* seedlings. In panel **F**, absolute levels for all compounds except RA were quantified by measuring $[M+H]^+$

ion abundances and comparing to standard curves. Error bars indicate ± standard deviation, based on 6 biological replicates. Cam: camalexin; RA: raphanusamic acid; other abbreviations as detailed in the abbreviations list above.



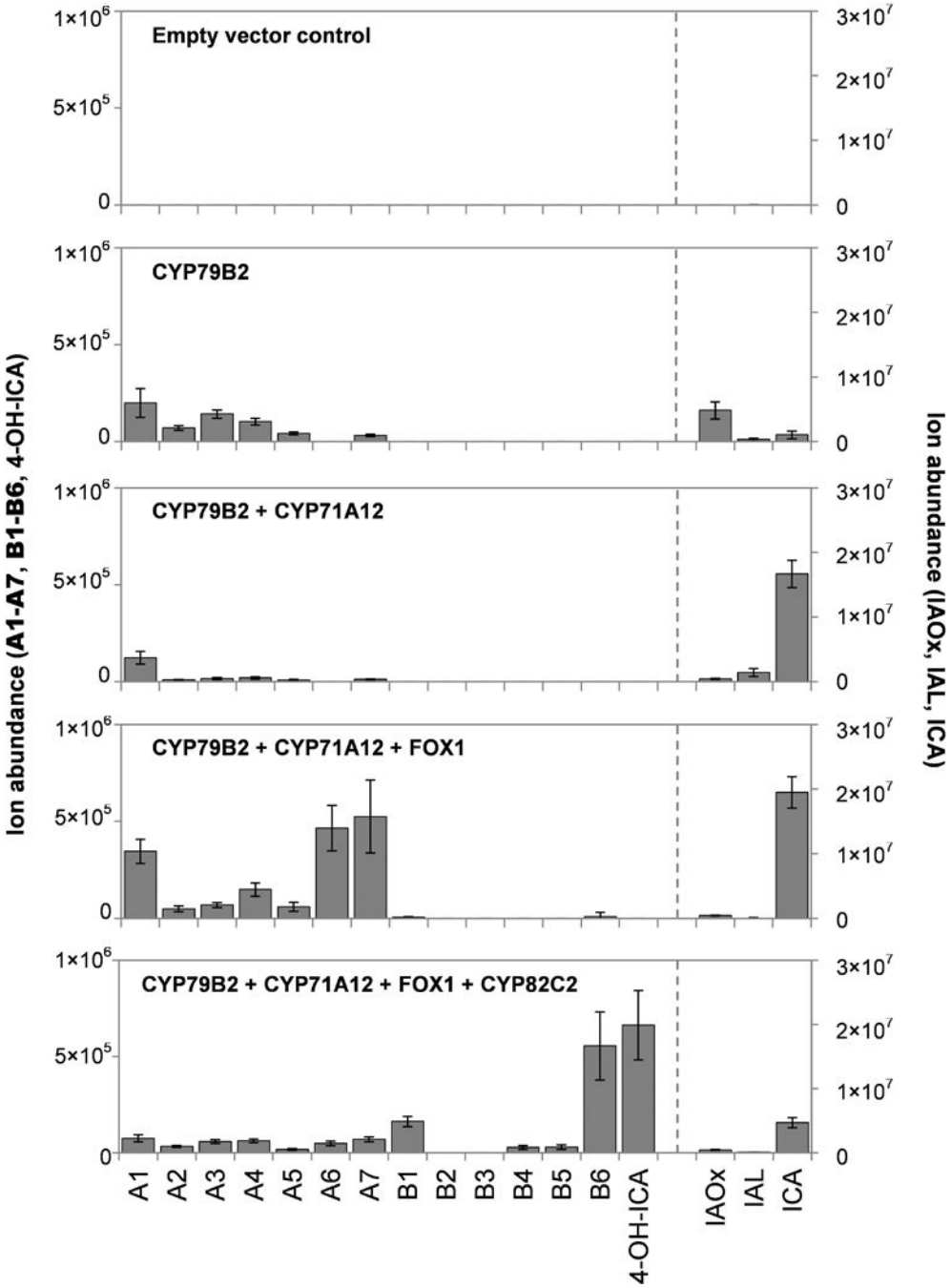
m/z (observed)	Predicted formula	Ion assignment	m/z (theoretical)	Δm/z (ppm)
691.2210	$C_{32}H_{36}N_4NaO_{12}^+$	$[2M + Na]^+$	691.2220	-1.4
357.1066	**$C_{16}H_{18}N_2NaO_6^+$**	**$[M + Na]^+$**	**357.1057**	**2.5**
330.0921	**$C_{15}H_{17}NNaO_6^+$**	**$[M + Na - HCN]^+$**	**330.0948**	**-8.2**
203.0500	$C_6H_{12}NaO_6^+$	$[Glucose + Na]^+$	203.0526	-12.8
185.0429	**$C_6H_{10}NaO_5^+$**	**$[Glucose + Na - H_2O]^+$**	**185.0420**	**4.9**
168.0428	**$C_9H_7NNaO^+$**	**$[M + Na - glucose - HCN + H_2O]^+$**	**168.0420**	**4.8**
146.0560	**$C_9H_8NO^+$**	**$[M + H - glucose - HCN + H_2O]^+$**	**146.0600**	**-27.4**
50.0033	**$CHNNa^+$**	**$[HCN + Na]^+$**	**50.0001**	**64.0**

m/z (observed)	Predicted formula	Ion assignment	m/z (theoretical)	Δm/z (ppm)
443.1117	$C_{19}H_{20}N_2NaO_9^+$	$[M + Na]^+$	443.1061	12.6
289.0535	$C_9H_{14}NaO_9^+$	$[O\text{-malonyl glucose} + Na]^+$	289.0530	1.7
271.0445	$C_9H_{12}NaO_8^+$	$[O\text{-malonyl glucose} + Na - H_2O]^+$	271.0424	7.7
245.0663	$C_8H_{14}NaO_7^+$	$[O\text{-malonyl glucose} + Na - CO_2]^+$	245.0632	12.6
227.0531	$C_8H_{12}NaO_6^+$	$O\text{-malonyl glucose} + Na - H_2O - CO_2]$	227.0526	2.2
168.0428	$C_9H_7NNaO^+$	$[M + Na - O\text{-mal. gluc.} - HCN + H_2O]^+$	168.0420	4.8

Extended Data Figure 6. Putative indole cyanogenic glycosides (ICGs) observed in *Arabidopsis* and in *N. benthamiana* expressing ICN pathway enzymes

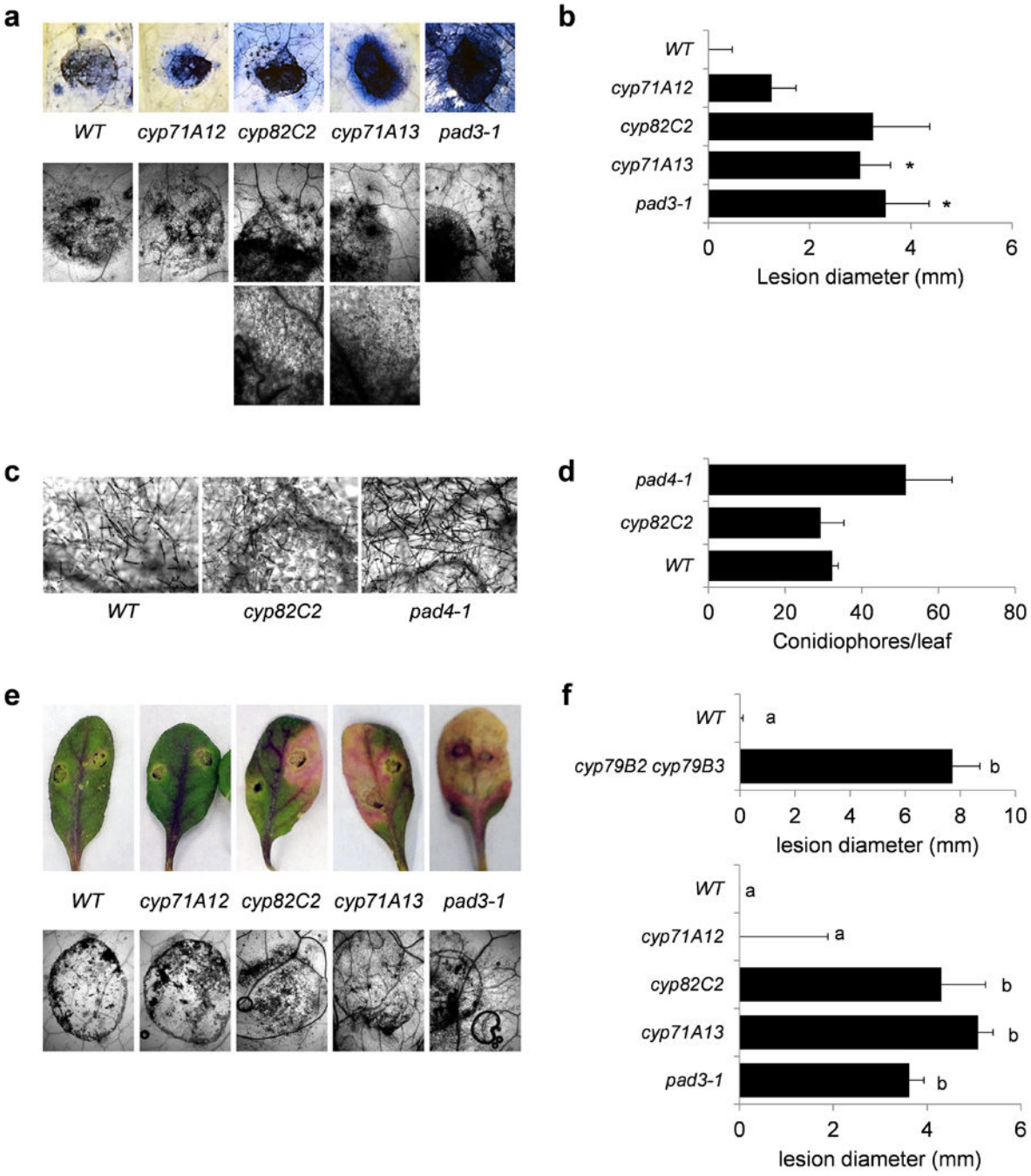
(**A**) EICs for putative ICGs in WT *Arabidopsis* and *fox* mutant elicited with *Psta*. The m/z values shown are median values calculated by XCMS. (**B**) Hypothesized structures and theoretical m/z values for the two ICGs identified. (**C**) MS/MS spectrum for ICG1; m/z

values and relative abundances are shown above each peak. The ion analyzed here (m/z = 691.2210) represents a $[2M+Na]^+$ dimer that is significantly more abundant than the $[M+Na]^+$ ion. Direct analysis of the $[M+Na]^+$ ion (m/z = 357.1057) yielded low abundance spectra that could not be easily analyzed. At lower collision energies, the $[2M+Na]^+$ ion fragments to $[M+Na]^+$, but yields a rich spectrum at 40 V, which is shown. Predicted peak assignments for the ICG1 MS/MS spectrum are shown in the accompanying table. For peaks in bold, exact counterparts could be identified in the dhurrin $[M+Na]^+$ 20 V MS/MS spectrum in the METLIN Metabolite Database. (**D**) MS/MS spectrum obtained for the ICG2 $[M+Na]^+$ ion and predicted peak assignments. While the $[2M+Na]^+$ peak (m/z = 864.2225) is also seen for this compound (not shown), $[M+Na]^+$ is more abundant in this case, and was analyzed directly. (**E**) Levels of ICG1 and ICG2 in ICN pathway mutants and (**F**) in WT plants elicited with *Psta* and *N. benthamiana* expressing ICN pathway enzymes. For panels (**E**) and (**F**), levels are quantified as mean $[M+Na]^+$ ion (m/z ± 10 ppm) abundances extracted from raw data, with error bars indicating ± standard deviation based on 6 biological replicates.



Extended Data Figure 7. ICN pathway metabolites are produced in *N. benthamiana* transiently expressing pathway genes

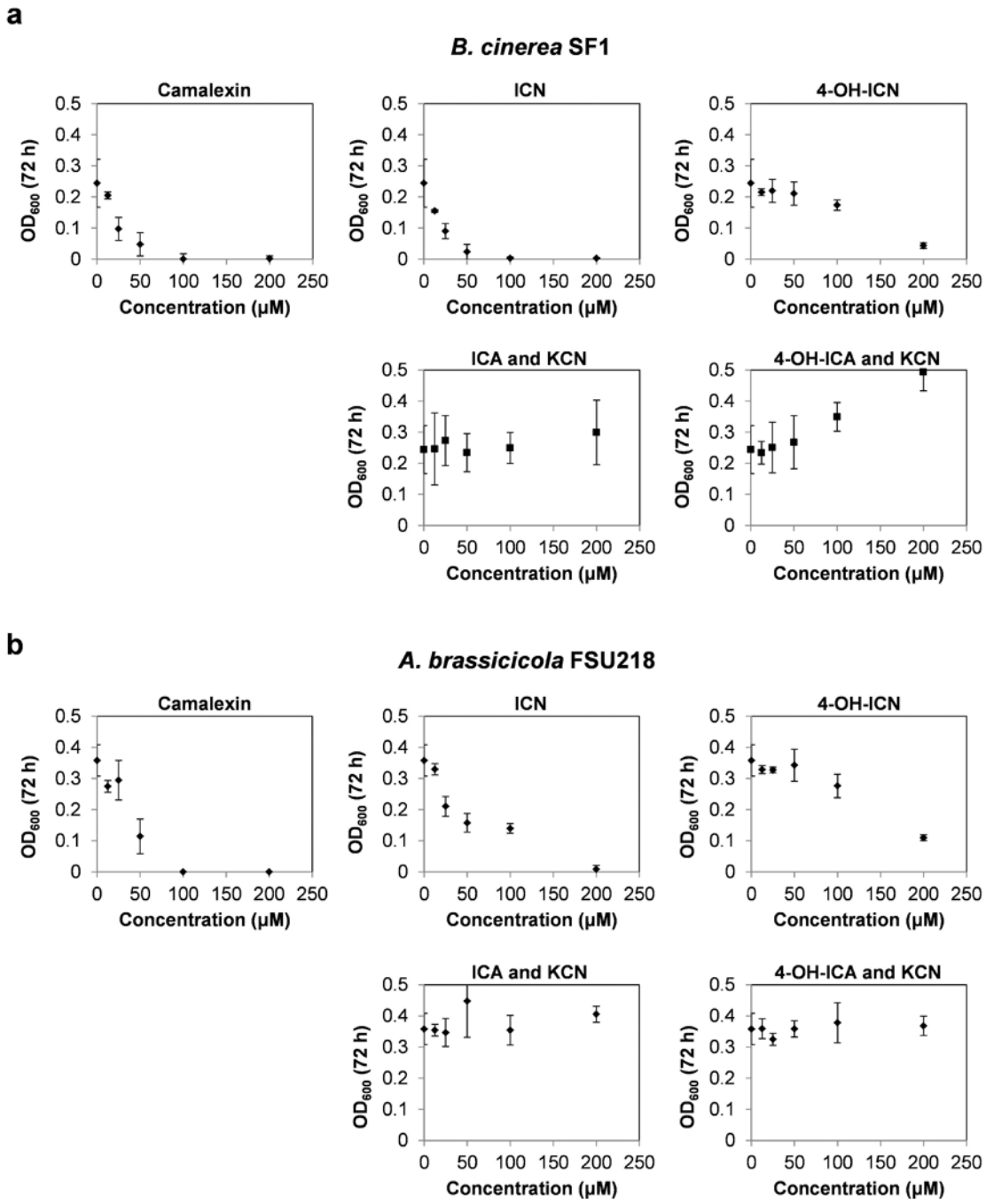
Levels of ICN and 4-OH-ICN derivatives (left axis) and other relevant indolic compounds (right axis), quantified as mean $[M+H]^+$ ion (m/z ± 10 ppm) abundances extracted from raw data, with error bars indicating ± standard deviation based on 6 biological replicates. The set of transiently expressed genes is indicated for each panel. Background levels of ICA and IAL detected when only the early pathway genes CYP71A12 and/or CYP79B2 are expressed indicate potential involvement of endogenous *N. benthamiana* enzymes.



Extended Data Figure 8. ICN pathway metabolites contribute to disease resistance towards *B. cinerea* but not towards *G. orontii*

(**A**) Top: Typical lactophenol trypan blue staining of leaves drop-inoculated with spores from the virulent fungal necrotroph *Botrytis cinerea* to visualize the extent of host cell death (darkly stained areas within and beyond the fungal spore droplet region). Middle: Microscopic analysis of stained leaves to visualize the extent of fungal colonization (stained filamentous fungal hyphae within and beyond the fungal spore droplet region). Images were taken at the same magnification (25×) and are representative of 5 biological replicates.

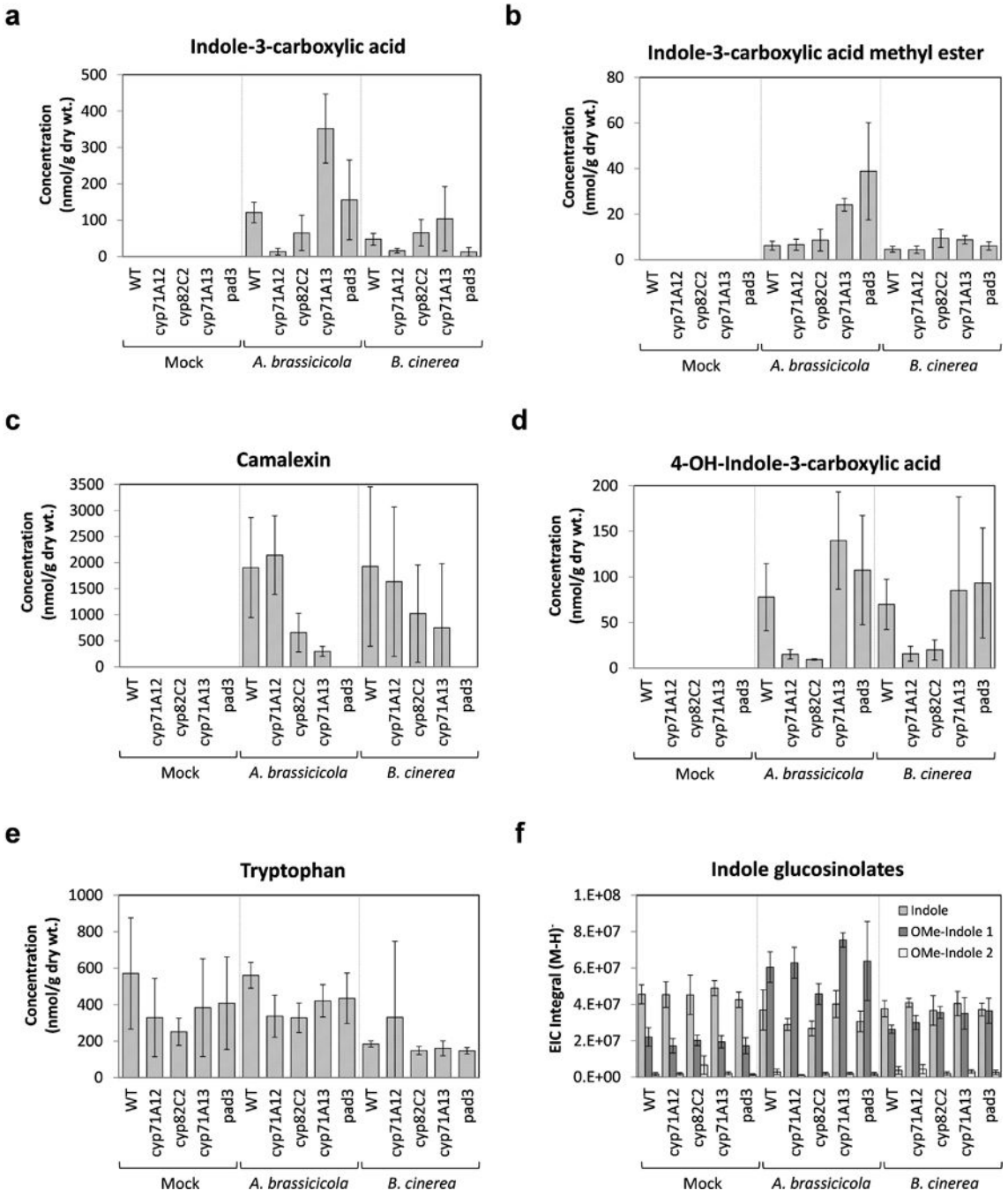
Bottom: Close up images of the fungal hyphae beyond the fungal spore droplet region for *cyp82C2* and *cyp71A13-3* mutants. Images were taken at the same magnification (100×). (**B**) Measurement of the disease lesion diameters in infected leaves. Data represent the median ± standard error for 5 biological replicates. Asterisks denote statistical significance relative to WT ($p < 0.05$, two-tailed *t* test). (**C**) Typical lactophenol trypan blue staining of fungal conidiophores (spore-bearing structures) formed in leaves infected with the adapted powdery mildew *Golovinomyces orontii*. The *pad4-1* mutant is more susceptible to fungal growth by *G. orontii* and thus produces significantly more conidiophores. Images were taken at the same magnification (100×) and are representative of 3 biological replicates. (**D**) Measurement of the number of conidiophores in infected leaves. Data represent the mean ± standard deviation for 3 biological replicates. (**E**) Top: Typical disease symptoms three days after drop inoculation of leaves with spores from the avirulent fungal necrotroph *Alternaria brassicicola*. Bottom: Microscopic analysis of infected leaves after lactophenol trypan blue staining confirming that disease symptoms are consistent with extent of fungal colonization (lightly stained fungal hyphae extending from the fungal spore droplet region) and host cell death (darkly stained areas along and beyond the border of the spore droplet region). Images were taken at the same magnification (25×) and are representative of 10 biological replicates. (**F**) Measurement of the disease lesion diameters in infected leaves. Data represent the median ± standard error of 8 (top graph) or 10 biological replicates (bottom graph). Different letters denote statistically significant differences ($p < 0.05$, two-tailed *t* test).



Extended Data Figure 9. ICN and 4-OH-ICN but not their degradation products inhibit fungal growth *in vitro*

Fungal growth inhibition assays on *Botrytis cinerea* SF1 (**A**) or *Alternaria brassicicola* FSU218 (**B**) with the tested compound (or compound combination) indicated. For compound combinations, the concentration indicated is for each compound; the given combinations approximate the hydrolysis products of ICN or 4-OH-ICN. Growth of fungi in PDB on a microplate was quantified by measuring absorbance at 600 nm (OD_{600}) 72 hours after spore inoculation and subtracting the absorbance at 0 h; see Methods for further details.

Error bars represent ± standard deviation based on 3 biological replicates. Note that the IC_{50} for both camalexin and ICN is approximately 25 μM against *B. cinerea* and 50 μM against *A. brassicicola*. For 4-OH-ICN, the inhibitory effect is not as pronounced, possibly due to rapid degradation of 4-OH-ICN in PDB (see SI Table 2).



Extended Data Figure 10. Levels of indolic compounds in leaves of mature plants after mock treatment or fungal infection

Tissue extracts were analyzed by LC-MS 7 dpi for *Alternaria brassicicola* FSU218 and 5 dpi for *Botrytis cinerea* SF1. (**A**-**E**) Levels of indicated compound, quantified as EIC

integral for the $[M+H]^+$ ion (m/z ± 10 ppm) and converted to absolute amounts by comparison with a standard curve. (**F**) Ion count integrals for indole glucosinolates ($[M-H]^-$ ion, m/z ± 10 ppm). Error bars in all panels represent ± standard deviation based on 6 biological replicates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Fred Ausubel, Mary Beth Mudgett, Chaitan Khosla, Alan Saghatelian, Yves Millet, Christian Danna, Stephanie Galanie, and members of the Sattely and Clay labs for advice on experiments and critical comments on the manuscript. We thank the Salk Institute Genomic Analysis Laboratory for providing the sequence-indexed Arabidopsis T-DNA insertion mutants. We thank George Lomonossoff (John Innes Centre) for providing plasmid pEAQ. This work was supported by R00 GM089985 and DP2 AT008321 (to E.S.S.), T32 GM008412-20 (to J.R.), and T32 GM007499-38 (to B.B.). The early stages of this work were supported by NSF grant MCB-0519898 and NIH grant R37 GM 48707 (awarded to F. Ausubel, Massachusetts General Hospital, Boston, MA).

REFERENCES

1. Chae L, Kim T, Nilo-Poyanco R, Rhee SY. Genomic signatures of specialized metabolism in plants. Science. 2014; 344:510–513. [PubMed: 24786077]
2. D'Auria JC, Gershenzon J. The secondary metabolism of Arabidopsis thaliana: growing like a weed. Current opinion in plant biology. 2005; 8:308–316. [PubMed: 15860428]
3. Bednarek P, Osbourn A. Plant-microbe interactions: chemical diversity in plant defense. Science. 2009; 324:746–748. [PubMed: 19423814]
4. Denoux C, et al. Activation of defense response pathways by OGs and Flg22 elicitors in Arabidopsis seedlings. Molecular plant. 2008; 1:423–445. [PubMed: 19825551]
5. Bak S, et al. Cytochromes p450. The Arabidopsis book / American Society of Plant Biologists. 2011; 9:e0144. [PubMed: 22303269]
6. Jones, Tappey H.; C, WE.; Meinwald, Jerrold; Eisner, Hans E.; Eisner, Thomas. Benzoyl cyanide and mandelonitrile in the cyanogenetic secretion of a centipede. Journal of Chemical Ecology. 1976; 2:421–429.
7. Zagrobelny M, Bak S, Moller BL. Cyanogenesis in plants and arthropods. Phytochemistry. 2008; 69:1457–1468. [PubMed: 18353406]
8. Gleadow RM, Moller BL. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. Annual review of plant biology. 2014; 65:155–185.
9. Tattersall DB, et al. Resistance to an herbivore through engineered cyanogenic glucoside synthesis. Science. 2001; 293:1826–1828. [PubMed: 11474068]
10. Bednarek P, et al. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. Science. 2009; 323:101–106. [PubMed: 19095900]
11. Clay NK, Adio AM, Denoux C, Jander G, Ausubel FM. Glucosinolate metabolites required for an Arabidopsis innate immune response. Science. 2009; 323:95–101. [PubMed: 19095898]
12. Dangl JL, Horvath DM, Staskawicz BJ. Pivoting the plant immune system from dissection to deployment. Science. 2013; 341:746–751. [PubMed: 23950531]
13. Ahuja I, Kissen R, Bones AM. Phytoalexins in defense against pathogens. Trends in plant science. 2012; 17:73–90. [PubMed: 22209038]
14. Vinayavekhin N, Saghatelian A. Untargeted metabolomics. **Chapter 30**, Unit 30 31 31-24. Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]. 2010
15. An alternative route to ICA in Arabidopsis has been reported, see: Bottcher C, et al. The Biosynthetic Pathway of Indole-3-Carbaldehyde and Indole-3-Carboxylic Acid Derivatives in Arabidopsis. Plant physiology. 2014; 165:841–853. [PubMed: 24728709] .

16. Mikkelsen MD, Hansen CH, Wittstock U, Halkier BA. Cytochrome P450 CYP79B2 from Arabidopsis catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. The Journal of biological chemistry. 2000; 275:33712–33717. [PubMed: 10922360]
17. Zhao Y, et al. Trp-dependent auxin biosynthesis in Arabidopsis: involvement of cytochrome P450s CYP79B2 and CYP79B3. Genes & development. 2002; 16:3100–3112. [PubMed: 12464638]
18. Millet YA, et al. Innate immune responses activated in Arabidopsis roots by microbe-associated molecular patterns. The Plant cell. 2010; 22:973–990. [PubMed: 20348432]
19. Boudsocq M, et al. Differential innate immune signalling via Ca(2+) sensor protein kinases. Nature. 2010; 464:418–422. [PubMed: 20164835]
20. Peyret H, Lomonossoff GP. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. Plant molecular biology. 2013; 83:51–58. [PubMed: 23479085]
21. Thomma BP, Nelissen I, Eggermont K, Broekaert WF. Deficiency in phytoalexin production causes enhanced susceptibility of Arabidopsis thaliana to the fungus Alternaria brassicicola. The Plant journal : for cell and molecular biology. 1999; 19:163–171. [PubMed: 10476063]
22. Gomez-Gomez L, Boller T. Flagellin perception: a paradigm for innate immunity. Trends in plant science. 2002; 7:251–256. [PubMed: 12049921]
23. Zipfel C, et al. Bacterial disease resistance in Arabidopsis through flagellin perception. Nature. 2004; 428:764–767. [PubMed: 15085136]
24. Liu F, et al. The Arabidopsis P450 protein CYP82C2 modulates jasmonate-induced root growth inhibition, defense gene expression and indole glucosinolate biosynthesis. Cell research. 2010; 20:539–552. [PubMed: 20354503]
25. Nafisi M, et al. Arabidopsis cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. The Plant cell. 2007; 19:2039–2052. [PubMed: 17573535]
26. Nelson D, Werck-Reichhart D. A P450-centric view of plant evolution. The Plant journal : for cell and molecular biology. 2011; 66:194–211. [PubMed: 21443632]
27. Moller BL. Functional diversifications of cyanogenic glucosides. Current opinion in plant biology. 2010; 13:338–347. [PubMed: 20197238]
28. Rauhut T, Glawischnig E. Evolution of camalexin and structurally related indolic compounds. Phytochemistry. 2009; 70:1638–1644. [PubMed: 19523656]

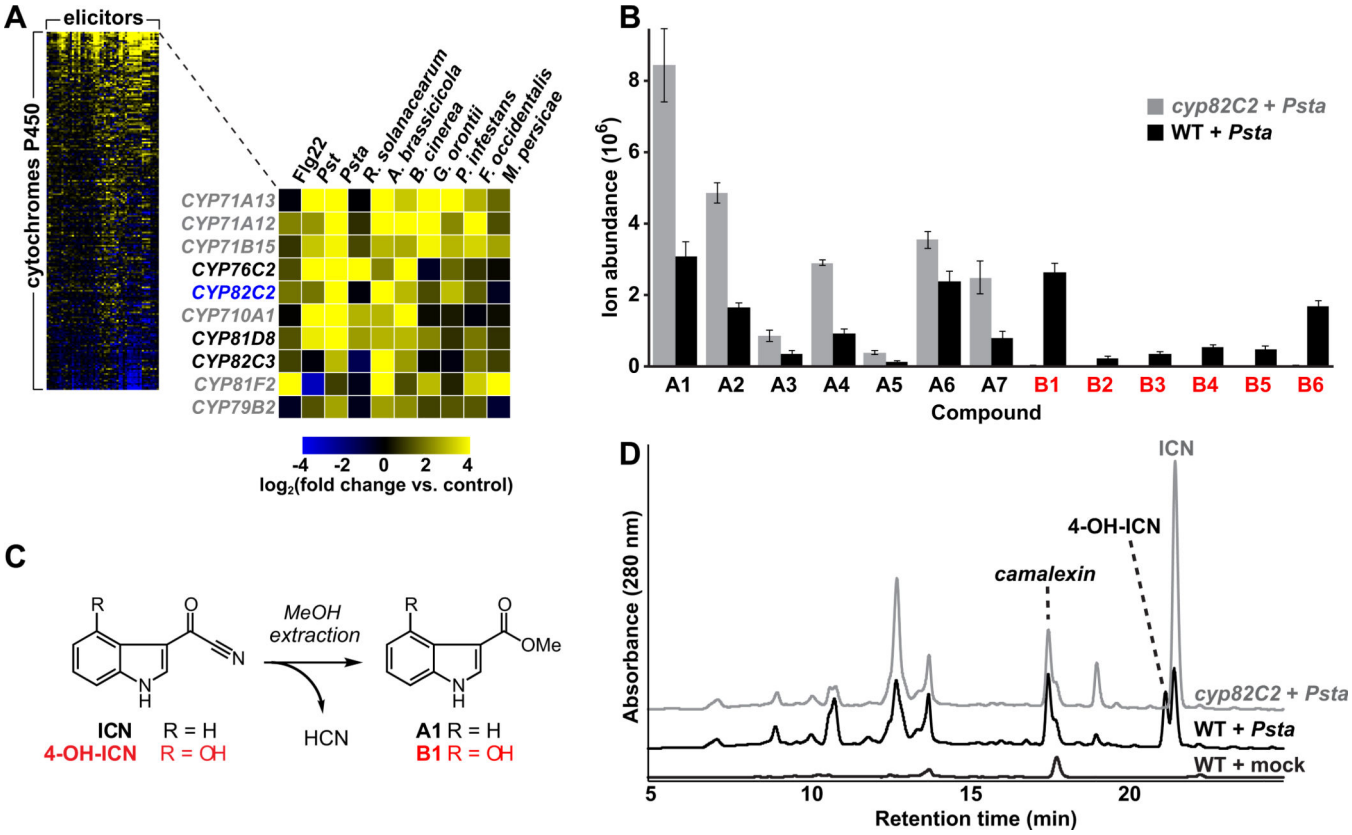


Fig. 1. Transcriptomic and metabolomic analyses implicate CYP82C2 in the biosynthesis of novel pathogen defense-related secondary metabolites

(**A**) Heat map of relative gene expression levels for cytochrome P450 genes in *Arabidopsis* under various pathogen stress conditions. The enlarged map shows the top 10 P450 genes after sorting by mean expression level over all conditions. P450s in gray have previously been biochemically characterized. (**B**) Levels of the most significantly differing metabolites identified in seedling comparative metabolomics experiments with *cyp82C2*. Data represent the mean ± standard deviation of 6 biological replicates. (**C**) ICA methyl ester (**A1**) and 4-OH-ICA methyl ester (**B1**) are methanolic degradation products of ICN and 4-OH-ICN. (**D**) HPLC traces of growth medium for WT and *cyp82C2* seedlings, showing *Psta*-dependent accumulation of ICN and 4-OH-ICN.

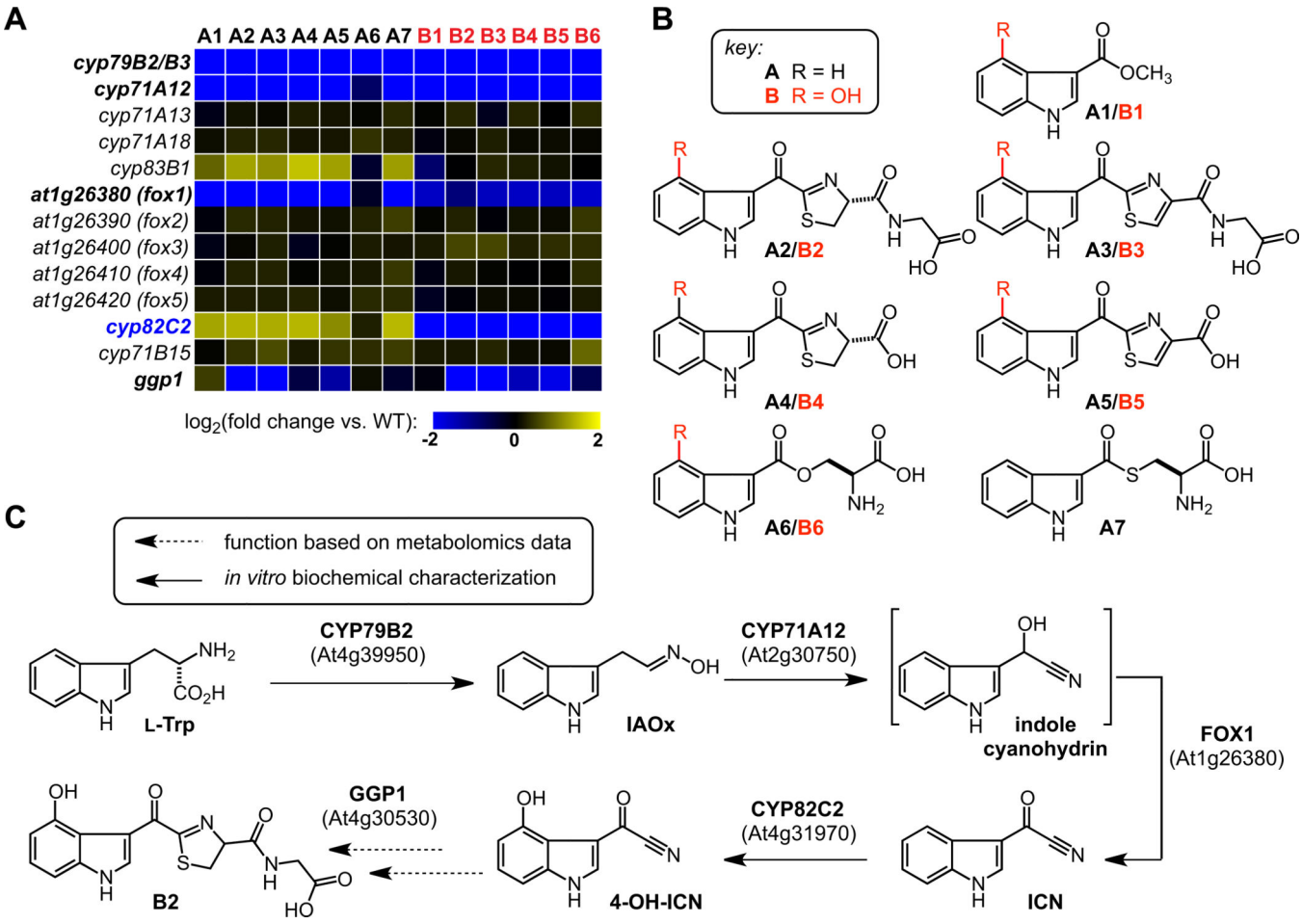


Fig. 2. Targeted metabolic profiling of candidate T-DNA insertion lines helps uncover the entire ICN biosynthetic pathway

(**A**) Heat map of mean ICN-derived metabolite levels relative to WT in *Psta*-elicited T-DNA insertion lines. Mutants in bold have significantly decreased levels of ICN derivatives. Note that **A6** levels are not affected to the same extent as levels of other metabolites in any line except for *cyp79B2/B3*, hinting at an alternative biosynthetic route from IAOx for this metabolite. (**B**) Structures of all ICN derivatives, confirmed by comparison with synthetic standards (see Extended Data Fig. 3 and SI Table 1). (**C**) Proposed biosynthetic pathway from Trp to 4-OH-ICN and downstream metabolites.

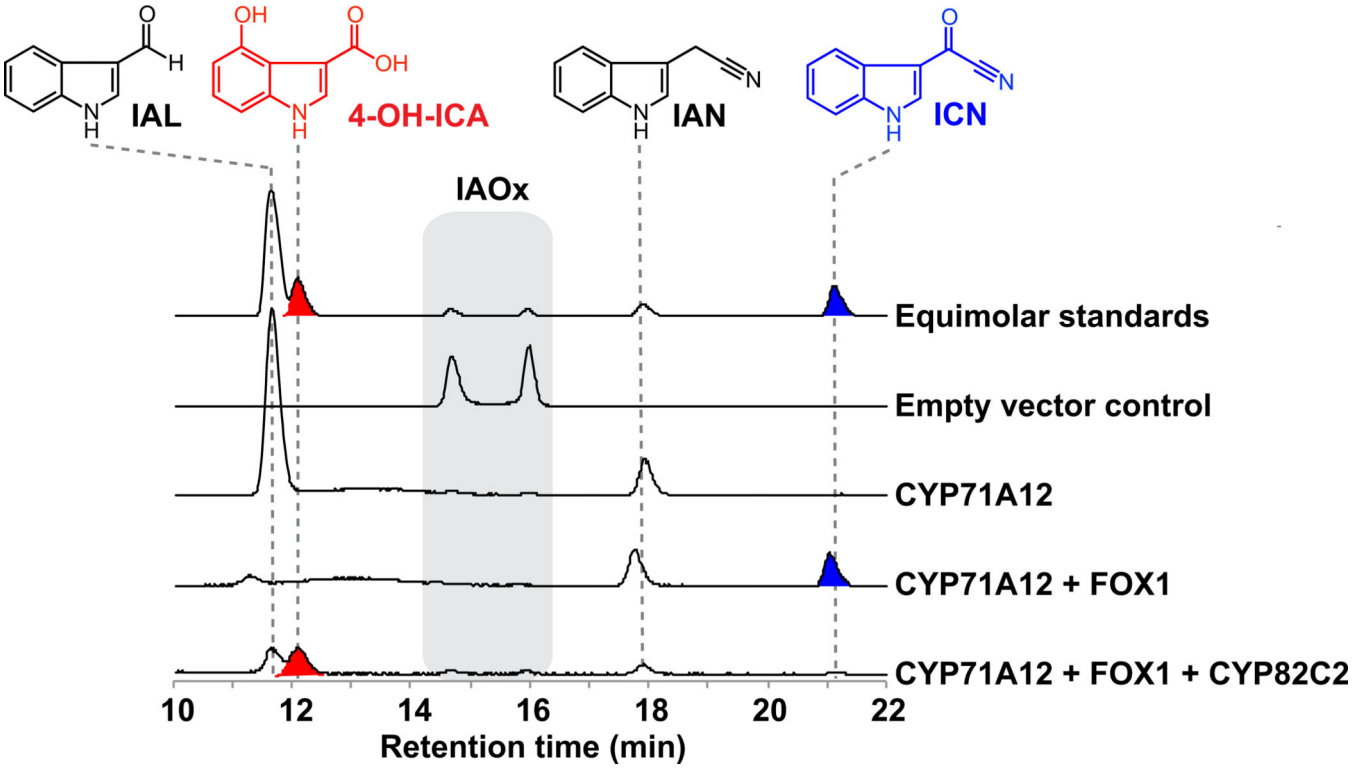


Figure 3. ***In vitro*** **reconstitution of 4-OH-ICN biosynthesis from IAOx**

Combined extracted ion chromatograms for IAOx substrate and reaction products for various subsets of enzymes in the 4-OH-ICN pathway. 4-OH-ICN could not be detected directly and its hydrolysis product 4-OH-ICA is shown instead.

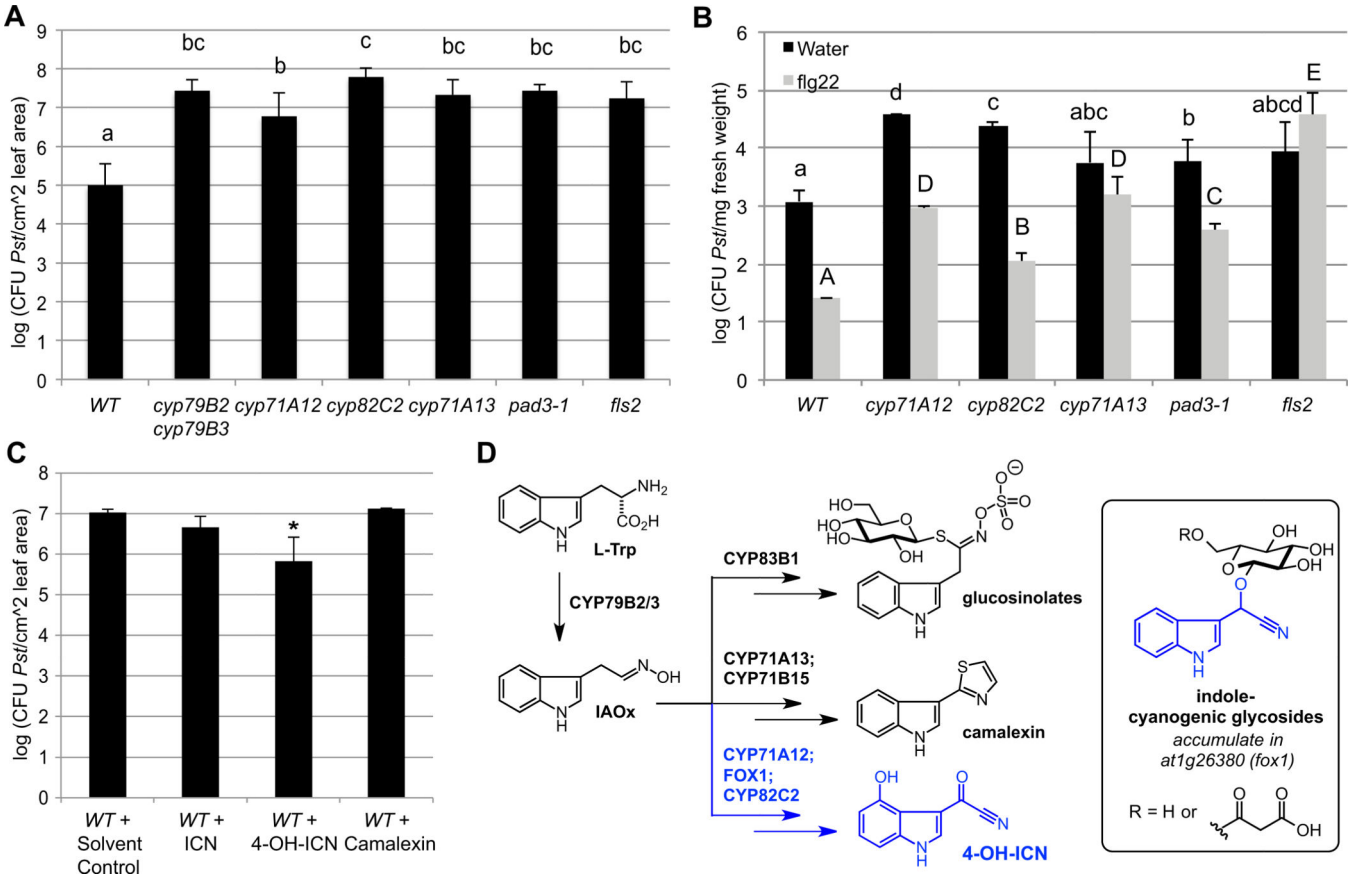


Figure 4. Camalexin and CYP82C2-synthesized 4-OH-ICN contribute nonredundantly to disease resistance against the virulent bacterial pathogen *Pseudomonas syringae*

(**A**) Growth analysis of the virulent *P. syringae* pv. *tomato* DC3000 (*Pst*) in surface-inoculated adult leaves. Data represent the mean ± standard error of 4 biological replicates. Different letters denote statistically significant differences (p <0.05, two-tailed t test). WT, wildtype (Col-0 ecotype). (**B**) Growth analysis of *Pst* in 10-day-old seedlings pretreated with water or 1 μM bacterial MAMP flg22 for 6 h. Data represent the median ± standard error of 4 biological replicates of 10–15 seedlings each. Different letters denote statistically significant differences (p < 0.05, two-tailed t test). (**C**) Growth analysis of *Pst* in wildtype adult leaves pre-immunized with 1 μM flg22 and 100 μM ICN, 4OH-ICN, camalexin or solvent control (DMSO) for 24 h prior to infiltration with *Pst*. Data represent the median ± standard error of 3 biological replicates. Asterisk denotes statistical significance relative to wildtype (p < 0.01, two-tailed t test). Experiment was repeated three times, producing similar results. (**D**) Summary of known major Trp-derived secondary metabolites in *Arabidopsis* and oxidative biosynthetic enzymes that have been used to reconstitute the pathways *in vitro* or *in planta*.