Full title:

Reversing the escape from herbivory: Knockout of cardiac glycoside biosynthesis in wormseed wallflower (*Erysimum cheiranthoides* L., Brassicaceae)

Short running title:

Cardiac glycoside knockouts in *Erysimum*

Authors:

Gordon C. Younkin[1,2] (gcy7@cornell.edu), Martin L. Alani[1†] (mlalani@mit.edu), Anamaría Páez Capador[1] (ap2225@cornell.edu), Hillary D. Fischer[1] (hdf27@cornell.edu), Mahdieh Mirzaei[1] (mm2895@cornell.edu), Amy P. Hastings[3] (aph54@cornell.edu), Anurag A. Agrawal[3] (aa337@cornell.edu), Georg Jander[1‡] (gj32@cornell.edu)

Author affiliations:

1. Boyce Thompson Institute, 533 Tower Rd, Ithaca, NY 14853, USA
2. Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, New York 14853
3. Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

† Present address: Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02142, USA

‡ Corresponding author

ABSTRACT

Like other members of the Brassicaceae, plants in the wallflower genus (*Erysimum*) produce glucosinolates, which are potent defenses against a wide range of herbivores. As a more recently evolved second line of defense, *Erysimum* produces cardiac glycosides, which are allosteric inhibitors of Na^+,K^+-ATPases in animals. Cardiac glycoside biosynthesis has evolved in diverse lineages including foxglove (*Digitalis*, Plantaginaceae) and milkweeds (Apocynaceae), but the full biosynthetic pathway has not been described in any species. We identify and generate CRISPR/Cas9 knockouts of two cytochrome P450 monooxygenases involved in cardiac glycoside biosynthesis in wormseed wallflower (*Erysimum cheiranthoides* L.): *Ec*CYP87A126, which cleaves the side chain from sterol precursors to initiate cardiac glycoside biosynthesis, and *Ec*CYP716A418, which has a role in cardiac glycoside hydroxylation. In the *Ec*CYP87A126 knockout lines, cardiac glycoside production is eliminated, effectively reversing *Erysimum*'s escape from herbivory. For the generalist herbivores green peach aphid (*Myzus persicae* Suzler) and cabbage looper (*Trichoplusia ni* Hübner), cardiac glycosides appear to be largely redundant with glucosinolates, having some effect in choice assays but little to no effect on insect performance. By contrast, the crucifer-feeding specialist cabbage butterfly (*Pieris rapae* L.), which will not oviposit or feed on wildtype *E. cheiranthoides*, is able to complete its life cycle on cardenolide-free *E. cheiranthoides* mutant lines. Thus, our study demonstrates *in vivo* that cardiac glycoside production allows *Erysimum* to escape from a specialist herbivore.

KEYWORDS

Erysimum, cardiac glycoside, evolutionary theory, *Pieris rapae*, herbivory, Brassicaceae

INTRODUCTION

The chemical arms race between plants and their insect herbivores is foundational to our understanding of how ecological interactions generate and maintain biological diversity (Ehrlich & Raven, 1964; Feeny, 1977; Fraenkel, 1959; Gordon, 1961). Under this paradigm, plants that evolve the ability to produce toxic or deterrent metabolites protect themselves from herbivore feeding and enter a "new adaptive zone" in which they may rapidly diversify in the absence of natural enemies (Ehrlich & Raven, 1964). However, as their enemies evolve the ability to tolerate or neutralize these metabolites, they may in turn enter this protected zone (Gordon, 1961), thereby re-applying ecological pressures that force plants to further adapt their defenses.

The Brassicaceae, a botanical family of >4000 species, presents many instances of this chemical arms race between plants and specialized herbivores. Glucosinolates evolved as a defense in this lineage approximately 90 million years ago and facilitated multiple rounds of species radiations that resulted in the high species diversity of the Brassicaceae (Edger et al., 2015). Since its original gain, many insect species have adapted to this defense, evolving the ability to tolerate, detoxify, or sequester glucosinolates (Feeny, 1977; Okamura et al., 2022). As a likely response to selection by these glucosinolate-adapted insects, several lineages within Brassicaceae accumulate novel toxic compounds as a second line of defense: globe candytuft (*Iberis umbellata* L.) makes cucurbitacins (Dong et al., 2021), garlic mustard (*Alliaria petiolata* Bieb.) makes hydroxynitrile glucosides (Frisch & Møller, 2012), scurvy-grass (*Cochlearia* spp. L.) makes tropane alkaloids (Brock et al., 2006), wintercress (*Barbarea vulgaris* W. T. Aiton) makes saponins (Shinoda et al., 2002), and wallflowers (*Erysimum* spp. L.) make cardiac glycosides (Makarevich et al., 1994). These are hypothesized to represent key evolutionary

innovations, allowing these lineages to once more escape their specialized herbivores and rapidly diversify (Dong et al., 2021; Züst et al., 2020).

Cardiac glycosides from *Erysimum* likely are powerful deterrents for crucifer-feeding specialist insects. For example, the glucosinolate-tolerant small and large white cabbage butterflies (*Pieris rapae* L. and *Pieris brassicae* L.) use a wide range of Brassicaceae as host plants but generally avoid *Erysimum*. Experiments involving bioactivity-guided fractionation identified cardiac glycosides as potential agents of this deterrence, and painting cardiac glycosides onto cabbage leaves further established a causal link between the isolated compounds and herbivore behavior (Renwick et al., 1989; Rothschild et al., 1988; Sachdev-Gupta et al., 1993). Similarly, isolated cardiac glycosides were shown to be feeding deterrents to three crucifer-feeding specialist flea beetles (Nielsen, 1978a, 1978b). Characterizing the biosynthetic origins of this second line of defense, along with generating cardiac glycoside-deficient mutant lines, would allow for *in vivo* tests of the role cardenolides play in *Erysimum*'s escape from herbivory. Furthermore, such mutants would allow a comparative assessment of cardiac glycoside biosynthetic genes between *Erysimum* and other cardiac glycoside-producing families.

Accordingly, the aims of this study were two-fold. First, we sought to identify cytochrome P450s involved in cardenolide biosynthesis in wormseed wallflower (*Erysimum cheiranthoides* L.), including the enzyme responsible for catalyzing the first committed step at the branch point between sterol metabolism and cardenolide metabolism, i.e., the cleavage of the sterol side chain to produce pregnenolone (**2**) (Figure 1). Second, we revisited classical ecological experiments with modern tools, using a newly-developed protocol for floral dip stable transformation of *E. cheiranthoides* to generate CRISPR/Cas9-mediated knockouts of cardiac glycoside biosynthetic genes. This effectively reversed its "escape from herbivory" and allowed

us to directly assess the role cardenolides play in defense against both generalist and specialist insects *in vivo*.

MATERIALS AND METHODS

Plants, insects, and growth conditions

All experiments were conducted with the genome-sequenced *E. cheiranthoides* var. Elbtalaue, which has been inbred for at least eight generations (Züst et al., 2020), Arabidopsis Biological Resource Center (https://abrc.osu.edu) accession number CS29250. Plants were grown in Cornell Mix (by weight 56% peat moss, 35% vermiculite, 4% lime, 4% Osmocote slow-release fertilizer [Scotts, Marysville, OH], and 1% Unimix [Scotts]) in Conviron (Winnipeg, CA) growth chambers with a 16:8 photoperiod, 180 μM m^{-2} s^{-1} photosynthetic photon flux density, 60% humidity, and constant 23 °C temperature.

Cabbage looper (*Trichoplusia ni* Hübner) eggs were obtained from Benzon Research (Carlisle, PA) and hatched on artificial diet (Southland Products, Lake Village, AR) in an incubator at 28 °C. Wild-caught *Pieris rapae* butterflies (Ithaca, NY, USA, June 2023) were used to start a lab colony. Adults were fed a 10% sucrose solution and were presented with *Brassica oleracea* var. capitata (Wisconsin Golden Acre cabbage) for oviposition and caterpillar feeding. Green peach aphid (*Myzus persicae* Suzler) were from a lab colony of a previously described, genome-sequenced "USDA" strain (Feng et al., 2023; Ramsey et al., 2007, 2014), which we maintained on *B. oleracea* var. capitata in a growth room with a 16:8 photoperiod and constant 23 °C temperature.

RNA-sequencing analysis

Raw RNA-sequencing reads from 48 *Erysimum* species (Züst et al., 2020) were downloaded from the NCBI Short Read Archive (PRJNA563696)(Strickler et al., 2019). Additional RNA-

sequencing data were collected from *E. cheiranthoides* tissues, including young leaves and roots (PRJNA1015726).

Tissue-specific samples were obtained from six-week-old plants of wildtype *E. cheiranthoides*. Young leaves that had just emerged, measuring approximately 1 cm in length, were harvested for the young leaf samples. The SV Total RNA Isolation Kit with on-column DNAse I treatment (Promega, Madison, WI, USA) was employed to isolate total RNA. The quantity and quality of RNA were evaluated using the RNA Integrity Number (RIN) from a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). For sequencing, 5 µg of purified total RNA, pooled from three replicates, was used for the preparation of strand-specific RNAseq libraries with 14 cycles of final amplification (Zhong et al., 2011). Subsequently, the purified libraries were multiplexed and subjected to sequencing with a paired-end read length of 150 bp using two lanes on an Illumina HiSeq2500 instrument (Illumina, San Diego, CA) at the Cornell University Biotechnology Resource Center (Ithaca, NY). Raw RNA-sequencing reads for species and tissue datasets were pseudoaligned to *E. cheiranthoides* genome v2.0 (PRJNA563696)(Strickler et al., 2019) using kallisto with default parameters, yielding transcript counts (Bray et al., 2016). Output files were normalized and transformed using the transform_counts.R script from the mr2mods pipeline (Wisecaver et al., 2017; https://github.itap.purdue.edu/jwisecav/mr2mods). Fold-change expression between leaf and root tissue was calculated using edgeR (McCarthy et al., 2012; Robinson et al., 2010).

Cloning of candidate genes

Erysimum cheiranthoides RNA was extracted from 2-week-old seedlings and young leaves of 5-week-old plants using the SV Total RNA Isolation System (Promega Corporation, Madison, WI). cDNA was generated using SMARTScribe Reverse Transcriptase (Takara Bio USA, Ann

Arbor, MI). Primers were ordered to include Gateway *attB* recombination sites (Supplemental Table S1), and the coding sequence was amplified from cDNA using Phusion High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA). The gel-purified amplicon was inserted into the pDONR207 vector using Gateway BP Clonase II enzyme mix and then into pEAQ-HT-DEST1 (Sainsbury et al., 2009) using Gateway LR Clonase II enzyme mix (ThermoFisher Scientific, Waltham, MA). The sequences of the inserted genes were verified with Sanger sequencing. All cloning was done using 10-beta Competent *E. coli* (NEB, Ipswich, MA), with transformations done using heat shock at 42 °C. Plasmids were purified using the Wizard Plus SV Minipreps DNA Purification System (Promega Corporation, Madison, WI) and transformed into *Agrobacterium tumefaciens* strain GV3101 using a freeze-thaw method (Weigel & Glazebrook, 2006).

***Transient expression of candidate genes in* Nicotiana benthamiana**

Genes were expressed in leaves of 4-week-old *Nicotiana benthamiana* plants (Bach et al., 2014). A single colony of *A. tumefaciens* strain GV3101 carrying pEAQ-HT-DEST1 carrying a candidate gene was inoculated into a 10 mL culture of LB with 50 µg/mL rifampicin, 20 µg/mL gentamicin, and 50 µg/mL kanamycin and shaken for 24 hours at 28 °C and 230 rpm. The bacteria were pelleted for 10 minutes at 3200 rcf in an Eppendorf Centrifuge 5810 (Hamburg, Germany) and resuspended to OD_{600}=0.5 in a solution containing 10 mM 2-(N-morpholino)ethanesulfonic acid (MES), 10 mM $MgCl_2$, and 400 µM acetosyringone before resting in the dark for 2 hours prior to infiltration into the abaxial surface of leaves using a blunt syringe. Each construct was infiltrated into leaves of at least three separate plants, with pEAQ-HT-DEST1 carrying GFP serving as a negative control. Tissue was collected five days after infiltration for UPLC-MS analysis. A 200 µM solution of pregnenolone (Sigma-Aldrich, St.

Louis, MO) was also infiltrated into separate *N. benthamiana* leaves two days prior to tissue harvest to check for any modifications that may occur *in planta*.

gRNA design and CRISPR/Cas9 constructs

One or two CRISPR guide RNAs (gRNAs) were designed to target the first exon of each candidate gene using the IDT CRISPR-Cas9 guide RNA design tool. Single-stranded DNA oligos were ordered for each gRNA, one containing the forward gRNA sequence and a 5' ATTG, and one containing the reverse complement and a 5' AAAC. gRNA sequences and corresponding oligos are provided in Supplemental Table S1. Complementary oligos were annealed and inserted into either pARV483 in the case of a single gRNA or into pARV370 in the case of multiple gRNAs targeting the same gene, using Type IIS restriction enzyme Aar1 (New England Biolabs). gRNA cassettes including the AtU6-26 promoter, gRNA scaffold, and AtU6-26 terminator were PCR amplified from pARV370 using primers containing PaqCI (New England Biolabs) restriction sites (Supplemental Table S1) and inserted in tandem into pARV380 such that all gRNAs targeting the same gene were contained on a single plasmid. Plasmid maps for pARV483, pARV370, and pARV380 are provided in Supplemental Figure S5.

***Floral dip stable transformation of* Erysimum cheiranthoides**

A floral dip stable transformation protocol for *E. cheiranthoides* was developed based on methods previously published for Arabidopsis (Clough & Bent, 1998) and *Brassica napus* (Wang et al., 2003). *Agrobacterium tumefaciens* strain GV3101 containing a binary plasmid was grown overnight on a shaker at 28 °C and 230 RPM in 5 mL lysogeny broth (LB) at pH 7.5 containing 50 μg/mL rifampicin, 20 μg/mL gentamicin, and 50 μg/mL kanamycin. The 5 mL culture was inoculated into 200 mL fresh LB with the same antibiotics and growth conditions for 24 hours.

To prepare the infiltration solution, *Agrobacterium* cultures were spun down at 3200 rcf in a Eppendorf Centrifuge 5810 for 10 minutes at room temperature and resuspended in a solution containing full strength Murashige and Skoog (MS) salts (Research Products International, Mt. Prospect, IL), 50 g/L sucrose, 0.1 mg/L 6-benzylaminopurine (Sigma-Aldrich), 400 μM acetosyringone (Sigma-Aldrich), and 0.01% Silwet L-77 (PlantMedia.com, Chiang Mai, Thailand) and were allowed to rest for one hour. *Erysimum cheiranthoides* plants just beginning to flower were selected. The inflorescence of each plant was submerged in the bacterial suspension, agitated, and placed under vacuum for 5 minutes. The vacuum was quickly released, and inflorescence was covered with plastic wrap and secured with a twist tie. Plants were kept in dark for 18-24 hours before removing the plastic wrap and returning to standard growth conditions. Seeds were harvested 6 weeks after dipping to be screened for transformants by looking for *DsRed* fluorescence using an SZX12 stereomicroscope equipped with a UV lamp (Olympus, Center Valley, PA).

T7 Endonuclease 1 assay for detecting Cas9-induced mutations

T1 plants were screened for mutations using a T7 endonuclease 1 (T7E1) assay. DNA was extracted from 3-week-old T1 plants by heating a 1 mm leaf disk in 25 μL Extract-N-Amp extraction solution (E7526) at 95 °C for 10 minutes and then adding 25 μL PCR Diluent (E8155, MilliporeSigma, St. Louis, MO). Primers were selected to amplify an approximately 1000 bp region flanking the Cas9 target site (Supplemental Table S1). PCR was carried out using Phire Green Hot Start II Mastermix (ThermoFisher Scientific, Waltham, MA) under manufacturer-recommended conditions. For the T7E1 assay, the Alt-R Genome Editing Detection Kit (Integrated DNA Technologies, Coralville, IA) was used according to manufacturer specifications. In any samples with the presence of non-wildtype amplicons, PCR products were

purified (Wizard SV Gel and PCR Clean-Up System, Promega Corporation, Madison, WI) and sent for Sanger sequencing at the Cornell Biotechnology Resource Center (Cornell University, Ithaca, NY). T2 seeds collected from plants with confirmed target site mutations were screened for the absence of fluorescence and the presence of a homozygous mutation at the target site using Sanger sequencing. T3 seeds collected from these non-transgenic, homozygous mutant plants were used for further analyses.

Metabolite feeding to cyp87a126* E. cheiranthoides *mutants

Predicted cardenolide intermediates were fed to *cyp87a126-2* mutant plants to check for rescue of cardenolide biosynthesis. Two hundred μM of pregnenolone, isoprogesterone (TLC Pharmaceutical Standards, Newmarket, ON), progesterone (Sigma-Aldrich), or 5β-pregnane-3,20-dione (aablocks, San Diego, CA) were suspended in 10 mM MES, 10 mM $MgCl_2$ and injected into the abaxial surface of young leaves of 4-week-old plants. Tissue was collected after 2 days for UPLC-MS analysis.

Metabolite extraction from plant tissue

The same protocol was used for tissue of both *E. cheiranthoides* and *N. benthamiana*. Two 14 mm leaf disks were collected into a 1.7 mL microcentrifuge tube (Laboratory Products Sales Inc., Rochester, NY, USA), either from an infiltrated region of leaf in the case of infiltration experiments, or from the youngest fully expanded leaves of four to five-week-old plants in the case of all other experiments. Tissue was flash frozen in liquid nitrogen and ground with three 3-mm ball bearings (Abbott Ball Company, Hartford, CT) on a 1600 MiniG™ tissue homogenizer (SPEX SamplePrep, Metuchen, NJ). One hundred μL 70% methanol with 15 μM internal standard was added to each sample (ouabain for positive ionization mode, sinigrin for negative ionization mode), which was then vortexed to suspend the plant tissue. Samples were

left to extract for half an hour at room temperature before being centrifuged for ten minutes at 17,000 rcf in a Z207-M microcentrifuge (Hermle, Sayreville, NJ). The supernatant was transferred to a clean 1.7 mL tube and centrifuged again for ten minutes at 17,000 rcf before being transferred to vials for UPLC-MS analysis.

Ultrahigh pressure liquid chromatography coupled to mass spectrometry (UPLC-MS)

Plant and yeast extracts were analyzed on an UltiMate 3000 UHPLC system coupled to a Q-Exactive hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA). The instrument was fitted with a Supelco Titan™ C18 UHPLC Column (80Å, 100 x 2.1 mm, particle size 1.9 µm; Sigma Aldrich). Injections of 2 µL were separated by a short (for quantification) or long (for figures) solvent gradient. Mobile phase A was water + 0.1% (v/v) formic acid and mobile phase B was acetonitrile + 0.1% (v/v) formic acid. All solvents were Optima LC/MS grade (Thermo Fisher Scientific). Short gradient: 0-0.5 minutes, hold at 2% B; 0.5-10 minutes, linear gradient from 2%-97% B; 10-11.5 minutes, hold at 97% B, 11.5-13 minutes, hold at 2% B. Long gradient: 0-5 minutes, hold at 2% B; 5-22 minutes, linear gradient from 2%-97% B; 22-23.5 minutes, hold at 97% B, 23.5-25 minutes, hold at 2% B. The solvent flow rate was 0.5 mL/minute, the column oven was set to 40 °C, and the autosampler temperature was 15 °C for all methods. The mass spectrometer was run in full scan positive ionization mode for detection of cardenolides and in full scan negative ionization mode for detection of glucosinolates.

LCMS data processing

For targeted LC-MS analysis, peak areas for compounds of interest were quantified using a custom processing method in Xcalibur™ Software (Thermo Fisher Scientific). Mass features

used for quantification are provided in Supplemental Table S2 for cardenolides and S3 for glucosinolates.

Na^+,K^+-ATPase Inhibition Assay

The inhibitory effect of plant extracts on porcine (*Sus scrofa* L) Na^+,K^+-ATPase was measured following the protocol described in Petschenka et al. (Petschenka et al., 2023). Wildtype *E. cheiranthoides* was compared with *E. cheiranthoides* mutant lines *cyp87a126-2* and *cyp87a126-1*, with *Arabidopsis thaliana* ecotype Columbia (Col-0) as a glucosinolate-containing, cardenolide-free control. Ten mg of freeze-dried ground plant tissue, three 3-mm ball bearings (Abbott Ball Company), and 1 mL 100% methanol were added to 1.7 mL tubes and were shaken on a 1600 MiniG™ (SPEX SamplePrep, Metuchen, NJ) twice for 1 minute at 1300 rpm. Samples were centrifuged for ten minutes at 17,000 rcf, and 700 µL of supernatant was transferred to fresh tubes before being dried completely in a Savant SpeedVac™ SC110 (Thermo Fisher Scientific). The extracts were resuspended in 14 µL 100% dimethyl sulphoxide (DMSO) by vortexing for 20 seconds, sonicating for 2 x 5 minutes, and centrifuging for 10 seconds at 16,000 rcf. Extracts were diluted 10-fold with deionized water for a final concentration of 10% DMSO. A series of 1x, 4x, 16x, 64x, 256x, 1024x dilutions was prepared in 10% DMSO. Two biological replicates of wildtype *E. cheiranthoides* (each with two technical replicates) and four biological replicates of *cyp87a126-1*, *cyp87a126-2*, and *A. thaliana* (each with one technical replicate) were distributed randomly among two 96-well plates.

Eighty µL of a reaction mix containing 0.0015 units of porcine Na^+,K^+-ATPase was combined with 20 µL of leaf extracts for final well concentrations of 100 mM NaCl, 20 mM KCl, 4 mM $MgCl_2$, 50 mM imidazole, and 2.5 mM ATP at pH 7.4. Each reaction was replicated using an identical reaction mix but lacking KCl as a no-activity background control. Plates were

incubated for 20 minutes at 37 °C before reactions were terminated by the addition of 100 µL sodium dodecyl sulfate (SDS, 10% plus 0.5% Antifoam A). Inorganic phosphate released from enzymatically hydrolyzed ATP was quantified at 700 nm using the method described by Taussky and Shorr (1953). Absorbance values for each reaction well were corrected by their respective background control well. Using the enzymatic activity across sample dilutions, sigmoid dose-response curves were calculated using a logistic function in the nlme package (J Pinheiro & Bates, 2000; Jose Pinheiro & Bates, 2023) in R statistical software (R Core Team, 2020). For each sample, the relative dilution at the inflection point was calculated to estimate the half maximal inhibitory concentration (IC50).

Insect Bioassays

For caterpillar growth and survival assays, individual 2-day-old *T. ni* or *P. rapae* larvae were placed on individual leaves 4-week-old of *E. cheiranthoides* wildtype, *cyp87a126-1*, and *cyp87a126-2* mutant lines (Supplemental Figure S2). For *T. ni*, 12 plants were used for each line, and caterpillars were placed on five leaves of each plant for a total of 60 caterpillars per line. For *P. rapae*, 16 plants were used per line, with caterpillars on a single leaf of the same age per plant. Caterpillars were restricted to a single leaf using 6.5 × 8 cm organza bags (amazon.com, item B073J4RS9C). After eight days, leaf damage was assessed and surviving larvae were moved to a fresh leaf to continue feeding.

Aphid and caterpillar choice assays were conducted in 100x15 mm Petri dishes (Thermo Fisher Scientific, Waltham, MA) sealed with Parafilm. For *T. ni*, 14 mm leaf disks from young leaves were placed in pairs of one wildtype and one mutant leaf disk on wet paper towels along with a single neonate caterpillar, with 20 replicates per mutant line, for a total of 40 replicates (Supplemental Figure S2). After 48 hours, photos were taken of each leaf disk, and leaf area

eaten was quantified using the Leaf Byte app (Getman-Pickering et al., 2020). For *M. persicae*, detached leaves were used instead of leaf disks, and 10 adult aphids were placed in each Petri dish (Supplemental Figure S2). Twelve replicates were done for each mutant line, and after 24 hours, the number of aphids on each leaf was recorded. Any replicates for which either or both leaves or disks were shriveled or limp at the end of the experiment were removed.

Myzus persicae colony growth was measured using synchronized adult aphids that had been reared from first instar nymphs for 7 days on cabbage. Five aphids were transferred from cabbage to bagged 3-week old plants, with 12 replicates for wildtype and each mutant line. After 9 days, the total number of adults and nymphs was recorded for each plant.

Pieris rapae oviposition assays were conducted using lab-reared adult butterflies. One wildtype plant and one mutant plant (either *cyp87a126-1* or *cyp87a126-2*) were placed in a 38 x 38 x 60 cm mesh cage with a mating pair of *P. rapae* butterflies and a 10% sucrose solution (Supplemental Figure S2). Butterflies were monitored daily and the total number of eggs on each plant was recorded once the female butterfly died or after five days. There were eleven replicates for each mutant line, but any instances where no eggs were laid on either plant were removed from the analysis.

Statistical analysis

All statistical analyses were carried out in R statistical software (R Core Team, 2020). The following functions and packages were used: edgeR (McCarthy et al., 2012; Robinson et al., 2010) for differential gene expression analysis, aov and TukeyHSD functions from base R for ANOVA and post-hoc tests. Plots were made using the packages genemodel (Monroe, 2017), MSnbase (Gatto et al., 2020; Gatto & Lilley, 2012), multcompView (Graves et al., 2023). R scripts for all statistical analyses are available on GitHub

(https://github.com/gordonyounkin/EcCYP87A126_scripts), and raw data underlying all figures are available in the Supplementary Information.

Phylogenetic Inference

Homologous sequences from selected species were identified using BLAST against public databases and were aligned using ClustalW (Madeira et al., 2022; Sievers et al., 2011). Gene phylogenies were inferred using IQ-TREE web server (Hoang et al., 2018; Minh et al., 2020; Trifinopoulos et al., 2016) with default parameters, except the number of bootstrap alignments was increased to 10,000.

RESULTS

Identification of candidate genes for cardiac glycoside biosynthesis

To identify cytochrome P450 monooxygenases involved in cardenolide biosynthesis, we examined patterns of gene expression across different *E. cheiranthoides* tissues and between 48 different species in the genus *Erysimum.* Two criteria were used: (1) Grafting experiments showed that cardenolides are synthesized in leaves of *E. cheiranthoides* and transported to the roots (Alani et al., 2021). Therefore, we expected high expression in leaves relative to roots. (2) *Erysimum collinum* produces nearly undetectable levels of cardenolides (Züst et al., 2020). If an enzyme has an exclusive role in cardenolide biosynthesis, it is predicted to be expressed at much lower levels in *E. collinum* relative to all other species of *Erysimum*. Of the 116 cytochrome P450s identified across the two expression datasets, only 3 matched both criteria, with at least 9-fold greater expression in young leaves relative to roots, and expression in *E. collinum* leaves more than 3 standard deviations below the mean of expression levels in other *Erysimum* species (Figure 2A). These 3 cytochrome P450s, *Ec*CYP71B132, *Ec*CYP716A418, and *Ec*CYP87A126,

were selected as candidates for involvement in cardenolide biosynthesis. Full length coding sequences for candidate genes are provided in Supplemental Data S1.

Phylogenetic analysis of candidate genes

All three candidate genes are in clades containing duplication events relative to *A. thaliana* (Figure 2B,D,E). *Ec*CYP87A126 is of particular interest because of recent reports identifying CYP87A members as capable of sterol side chain cleavage in several species including woolly foxglove (*Digitalis lanata* Ehrh.) (Carroll et al., 2023), as well as common foxglove (*D. purpurea* L.), Sodom apple (*Calotropis procera* W. T. Aiton), and *E. cheiranthoides* (Kunert et al., 2023). To better understand the convergence of this activity in diverse lineages, we aligned the amino acid sequences of *Ec*CYP87A126 and *Dl*CYP87A4. We found that the two amino acid substitutions identified by Carroll et al. (Carroll et al., 2023) as necessary for sterol side chain cleaving activity, V355A and A357L in *Dl*CYP87A4, were also present in *Ec*CYP87A126 (Figure 2C).

Cardenolide content is altered in Cas9-generated* cyp87a126 *and* cyp716a418 *mutant lines

We generated independent knockout lines for each of the three candidate cytochrome P450s using Cas9-mediated gene editing (Figure 2F-H, Supplemental Figure S1). None of the knockout lines for any of the three candidates showed an obvious visual phenotype (Figure 3B), and *Ec*CYP71B132 knockout lines displayed no changes in cardenolide content (Figure 3A). Knockout lines for the other two candidates had strong changes in cardenolide accumulation. *Ec*CYP716A418 knockout lines hyperaccumulate glycosides of digitoxigenin (**10**) (Figure 3A,D, one-way ANOVA: $F_{2,10}$=74.01 $p<0.001$; Tukey's HSD: WT-*cyp716a418-1* $p<0.001$, WT-*cyp716a418-2* $p<0.001$), apparently lacking the ability to hydroxylate digitoxigenin at C19 to form cannogenol (**11**), cannogenin (**12**), and strophanthidin (**13**) (Figure 1). *Ec*CYP87A126

knockout lines accumulated barely detectable levels of cardenolides (Figure 3A), suggesting that it is an essential enzyme in cardenolide biosynthesis. Despite a 1000-fold decrease in total cardenolide-related peak area (Figure 3C, one-way ANOVA: $F_{2,12}$=271 $p<0.001$; Tukey's HSD: WT-*cyp87a126-1* $p<0.001$, WT-*cyp87a126-2* $p<0.001$), *cyp87a126* lines display no difference in aliphatic (Figure 3C, one-way ANOVA: $F_{2,10}$=2.32 $p=0.15$) or indole (one-way ANOVA: $F_{2,10}$=1.71 $p=0.23$) glucosinolate abundance (Figure 3C).

Transient expression of* E. cheiranthoides *cytochrome P450s in* N. benthamiana *leaves

To investigate the *in planta* activity of *Ec*CYP716A418 and *Ec*CYP87A126, full length coding sequences were cloned and transiently expressed in *N. benthamiana* leaves, with substrate co-infiltration where necessary. Based on the strong phenotype of the knockout lines, we expected *Ec*CYP716A418 to hydroxylate digitoxigenin at C19. However, no activity was detected upon co-infiltration with digitoxigenin in *N. benthamiana* leaves. For *N. benthamiana* transiently expressing *Ec*CYP87A126, UPLC-MS analysis revealed accumulation of pregnenolone (**2**), consistent with its function as the sterol side chain cleaving enzyme (Figure 3E-F). No substrate was co-infiltrated for *Ec*CYP87A126, as *N. benthamiana* accumulates potential substrates cholesterol, campesterol, stigmasterol, and sitosterol in its leaves (Suza & Chappell, 2016).

Using* cyp87a126 *mutant lines to investigate intermediates in cardenolide biosynthesis

cyp87a126 mutant lines provide a tool for investigation of intermediates in cardenolide biosynthesis, as only the first enzyme in the pathway is absent and the rest of the pathway remains intact. The following predicted intermediates were fed to *cyp87a126-2* plants: pregnenolone (**2**), isoprogesterone (**3**), progesterone (**4**), and 5β-pregnane-3,20-dione (**5**). All infiltrated substrates rescued cardenolide biosynthesis in *cyp87a126-2* plants (Figure 4), consistent with the pathway shown in Figure 1.

Leaf extracts of* cyp87a126 *mutant lines display decreased Na^+,K^+-ATPase inhibition

We assessed cardenolide knockout lines for Na^+,K^+-ATPase inhibitory activity using an *in vitro* assay with porcine Na^+,K^+-ATPase. Methanolic extracts of *cyp87a126-1* displayed on average 15-fold lower inhibitory activity than extracts of wildtype leaves, and *cyp87a126-2* showed 251-fold lower inhibitory activity, on par with the cardenolide-free *A. thaliana* control (Figure 5A-B). A one-way ANOVA on log-transformed half maximal inhibitory concentration (IC50) found differences between groups (one-way ANOVA, $F_{3,12}$=69.34, p<0.001; Tukey's HSD: WT-*cyp87a126-1* p<0.001, WT-*cyp87a126-2* p<0.001, *cyp87a126-1*-*cyp87a126-2* p=0.051). While we do see some apparent Na^+,K^+-inhibition at high concentrations of both *A. thaliana* and cardenolide-free *E. cheiranthoides* lines, this is likely non-specific inhibition or an artefact of the assay, making it difficult to assess the fold-change decrease in inhibitory activity attributed solely to cardenolides.

Insect performance on* cyp87a126 *mutant lines

To test the impact of cardenolides on insects feeding on *E. cheiranthoides*, we conducted insect choice and performance assays using two generalist insects: the green peach aphid (*Myzus persicae*) and the cabbage looper moth (*Trichoplusia ni*), and one crucifer-feeding specialist: the cabbage butterfly (*Pieris rapae*).

In the choice assays, the overall trend was a preference for cardenolide-free mutant lines, with varying levels of significance for each species. While more adult *M. persicae* aphids chose mutant lines over wildtype (*cyp87a126-1*: 56% chose mutant, *cyp87a126-2*: 67% chose mutant), this difference was not significant (Figure 5C, paired t-test: *cyp87a126-1* p=0.12, *cyp87a126-2b1*: p=0.08). *Trichoplusia ni* caterpillars showed a clear preference for *cyp87a126* mutant lines over wildtype, as measured by leaf area eaten (Figure 5D; paired t-test: *cyp87a126-1* p=0.001,

cyp87a126-2 p=0.013). Gravid adult *P. rapae* uniformly chose to oviposit on *cyp87a126* mutant plants, with the exception of a single egg laid on a wildtype plant. For *cyp87a126-1*, all 128 eggs were laid on mutant plants (χ^2=128, df=1, p<0.001), and for *cyp87a126-2*, 74 of 75 eggs were laid on mutant plants, with only one egg laid on wildtype (χ^2=71.1, df=1, p<0.001).

Results were less uniform in performance assays. For *M. persicae*, population growth over nine days from five adult aphids restricted to a single plant was not different between wildtype and either mutant line (Figure 5C, one-way ANOVA: $F_{2,33}$=0.27 p=0.77). When bagged on individual leaves, *T. ni* was more likely to refuse to feed on wildtype than on either mutant line (Figure 5D; χ^2=17.44, df=2, p<0.001). Among caterpillars that did begin feeding, caterpillars grew marginally better on *cyp87a126-1* than on wildtype after correcting for leaf age, but there was no difference between *T. ni* growth on *cyp87a126-2* and wildtype (Figure 5D; one-way ANOVA $F_{2,82}$=4.18 p=0.019, Tukey's HSD: WT-*cyp87a126-1* p=0.014, WT-*cyp87a126-2* p=0.22). None of the 14 *P. rapae* caterpillars placed on wildtype plants began feeding, while 29 of 30 caterpillars placed on the two mutant lines fed and produced substantial damage (Figure 5E, χ^2=39.8, df=2, p<0.001). While mortality of *P. rapae* caterpillars was high in general, four of those feeding on mutant plants reached adulthood, demonstrating their suitability as a host plant.

DISCUSSION

Convergent evolution of CYP87A126 as a sterol side chain cleaving enzyme and the first committed step in cardenolide biosynthesis

The presence of sterol side-chain cleaving enzymes in cardenolide-producing plants has been the subject of speculation for decades (Iino et al., 2007; Lindemann, 2015; Lindemann & Luckner, 1997; Pilgrim, 1972; Stohs & El-Olemy, 1971) and have only recently been identified in *Digitalis* spp., *Calotropis procera*, and *E. cheiranthoides* (Carroll et al., 2023; Kunert et al.,

2023). In this study, we independently confirm that *Ec*CYP87A126 possesses sterol side chain cleaving activity and generate knockout lines showing that this activity is necessary for cardenolide production *in planta*. This discovery is an important first step towards establishing the full cardenolide biosynthetic pathway in *Erysimum*. Based on our knockout lines and substrate feeding, it is now clear that *Erysimum* cardenolide biosynthesis proceeds through pregnane intermediates much like in *Digitalis* (Kunert et al., 2023).

In addition, the characterization of *Ec*CYP87A126 demonstrates convergent evolution of enzyme activity between *Ec*CYP87A126 and *Dl*CYP87A4 (Carroll et al., 2023). Unlike animals, in which the unrelated cytochrome P450 CYP11A1 is responsible for sterol side chain cleavage (John et al., 1984; Miller & Auchus, 2011), both *Erysimum* and *Digitalis* evolved sterol side chain cleaving activity in enzymes from the CYP87A subfamily, in apparently independent instances of gene duplication and neofunctionalization. Notably, the two amino acid substitutions identified by Carroll et al. (Carroll et al., 2023) that are necessary for sterol side chain cleaving activity in *Dl*CYP87A4, V355A and A357L, are also present in *Ec*CYP87A126. While the ancestral function of the CYP87A clade is unknown, other related enzymes from the CYP87 family are known to act on triterpenoids (Ghosh, 2017; Zhou et al., 2016). The specifics behind how these enzymes acquired this novel activity, whether through shifts in substrate preference, activity, or both, will be the basis of future study.

Identification of other cytochrome P450 monooxygenases involved in cardenolide biosynthesis

Our screen of cytochrome P450 monooxygenases revealed a second P450 that is involved in cardenolide modification. *Ec*CYP716A418 mutant lines still make high quantities of cardenolides, but they almost exclusively accumulate digitoxigenin glycosides, which are not oxygenated at carbons 4 and 19 (Figure 1). Based on the predicted pathway, we hypothesize that

*Ec*CYP716A418 hydroxylates digitoxigenin at C19. However, we did not see this activity when co-infiltrating *Ec*CYP716A418 with digitoxigenin in *N. benthamiana*. This result leaves open the possibility that hydroxylation by *Ec*CYP716A418 occurs earlier in the pathway, for example prior to lactone ring formation, or only after glycosylation. It is also possible the observed phenotype is somewhat more cryptic, and there is not a direct link between this enzyme and cardiac glycoside hydroxylation. It is not surprising that *Ec*CYP716A418 is involved in cardenolide biosynthesis, as other members of CYP716A are well known for the modification of triterpenoid scaffolds, including β-amyrin (Carelli et al., 2011; Ghosh, 2017). In addition, *Ec*CYP716A418 is duplicated several times in *E. cheiranthoides* relative to *A. thaliana* (Figure 2D), a pattern that is often observed in the evolution of specialized metabolic pathways (Moghe & Last, 2015).

A knockout line of the third cytochrome P450 discussed in this paper, *Ec*CYP71B132, did not have an altered cardenolide phenotype. The lack of a phenotype does not conclusively exclude involvement of *Ec*CYP71B132 in cardenolide biosynthesis. For example, a potential alternative start codon 50 base pairs after the Cas9-induced deletion may result in a functional protein with an N-terminal truncation of 66 amino acids that leaves the active site intact. Even if *cyp71b132-1* is a complete functional knockout, it is possible that its role in cardenolide biosynthesis is complemented by a functionally redundant enzyme. Nonetheless, the lack of known CYP71B family members acting on steroid-like compounds indicates that the involvement of *Ec*CYP71B132 in cardenolide biosynthesis is less likely—the most closely related characterized enzyme, *At*CYP71B15, acts on indolic intermediates in camalexin biosynthesis (Böttcher et al., 2009).

Cardenolides as escape from herbivory

While it has long been understood that *Erysimum* represents a unique instance of the co-occurrence of two potent defensive compounds, the overall benefit to the plant for investing in two distinct but potentially redundant defenses has been difficult to test. Our insect feeding and performance assays highlight that plant-insect relationships can be highly species specific. The generalist green peach aphid performed similarly on *E. cheiranthoides* regardless of the presence of cardenolides, and the generalist cabbage looper, *T. ni*, preferred the cardenolide-free mutant but grew only slightly better when feeding on it. This is perhaps unsurprising as generalist insects are known for their ability to tolerate a wide range of defensive metabolites. The much clearer impact of the loss of cardenolides is the reversal of *Erysimum*'s escape from a glucosinolate-feeding specialist herbivore. Consistent with previous reports involving exogenously applied cardiac glycosides (Renwick et al., 1989; Rothschild et al., 1988; Sachdev-Gupta et al., 1993), we found that *P. rapae* refuses to oviposit or feed on wildtype *Erysimum*. However, with the loss of cardenolide biosynthesis in the *cyp87a126* mutant lines, *P. rapae* was able to complete its entire life cycle on *E. cheiranthoides*, from oviposition to caterpillar growth and pupation.

While our results are strong evidence that cardenolide production allowed *Erysimum* to escape from feeding pressure of specialist herbivores and provides a marginal advantage against some generalists, it is possible that a different pattern would emerge with other herbivores. However, complete escape from even a subset of specialist insects represents a distinct ecological advantage, as specialist insects are observed to cause the majority of damage a plant suffers in certain contexts (Bidart-Bouzat & Kliebenstein, 2008; Coley & Barone, 1996). Despite this apparently clear defensive advantage, cardenolide production has been lost or drastically reduced in the accession of *E. collinum* screened in this study (Züst et al., 2020). Whether this

loss has become fixed would require more thorough sampling in its native range in Iran, but the persistence of even some individuals with a complete lack of cardenolides points to context-dependent benefits and likely substantial costs of cardenolide production. *Erysimum collinum* also accumulates high levels of glucoerypestrin, a glucosinolate unique to *Erysimum* (Blažević et al., 2020; Fahey et al., 2001; Kjær & Gmelin, 1957; Züst et al., 2020) that may have allowed an alternative escape route from glucosinolate specialists, rendering cardenolides unnecessary as a second line of defense. By contrast, there are no known cases of the loss of glucosinolates in *Erysimum*, perhaps because glucosinolates are involved in non-defensive processes such as signaling and development (Katz et al., 2015). Future experiments, including observation of mutant performance in a more natural ecological context will more fully dissect the complex role cardenolides play in allowing *Erysimum* to escape herbivory.

ACKNOWLEDGEMENTS

We thank Prof. Tobias Züst for his insightful comments on the manuscript, Prof. David Nelson for cytochrome P450 naming, and Julia Dahl for her assistance with insect experiments. This research was funded by United States Department of Agriculture award 2020-60713-33565 to GJ; a Chemistry Biology Interface Training Program fellowship under National Institutes of Health/National Institute of General Medical Sciences (T32GM138826) and a US National Science Foundation Graduate Research Fellowship (DGE–2139899) to GCY; and a Summer Undergraduate Research Fellowship from the American Society of Plant Biologists and a Rawlings Cornell Presidential Research Scholar award to MLA.

REFERENCES

Alani, M. L., Younkin, G. C., Mirzaei, M., Kumar, P., & Jander, G. (2021). Acropetal and basipetal cardenolide transport in Erysimum cheiranthoides (wormseed wallflower).

Phytochemistry, *192*(September), 112965. https://doi.org/10.1016/j.phytochem.2021.112965

Bach, S. S., Bassard, J.-É., Andersen-Ranberg, J., Møldrup, M. E., Simonsen, H. T., & Hamberger, B. (2014). *High-Throughput Testing of Terpenoid Biosynthesis Candidate Genes Using Transient Expression in Nicotiana benthamiana BT - Plant Isoprenoids: Methods and Protocols* (M. Rodríguez-Concepción (ed.); pp. 245–255). Springer New York. https://doi.org/10.1007/978-1-4939-0606-2_18

Bidart-Bouzat, M. G., & Kliebenstein, D. J. (2008). Differential levels of insect herbivory in the field associated with genotypic variation in glucosinolates in Arabidopsis thaliana. *Journal of Chemical Ecology*, *34*, 1026–1037. https://doi.org/10.1007/s10886-008-9498-z

Blažević, I., Montaut, S., Burčul, F., Olsen, C. E., Burow, M., Rollin, P., & Agerbirk, N. (2020). Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry*, *169*, 112100. https://doi.org/10.1016/j.phytochem.2019.112100

Böttcher, C., Westphal, L., Schmotz, C., Prade, E., Scheel, D., & Glawischnig, E. (2009). The multifunctional enzyme CYP71b15 (PHYTOALEXIN DEFICIENT3) converts cysteine-indole-3-acetonitrile to camalexin in the indole-3-acetonitrile metabolic network of Arabidopsis thaliana. *Plant Cell*, *21*(6), 1830–1845. https://doi.org/10.1105/tpc.109.066670

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Brock, A., Herzfeld, T., Paschke, R., Koch, M., & Dräger, B. (2006). Brassicaceae contain nortropane alkaloids. *Phytochemistry*, *67*(18), 2050–2057. https://doi.org/10.1016/j.phytochem.2006.06.024

Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., Graham, N., Odoardi,

M., Piano, E., Arcioni, S., May, S., Scotti, C., & Calderini, O. (2011). Medicago truncatula CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell*, *23*(8), 3070–3081. https://doi.org/10.1105/tpc.111.087312

Carroll, E., Gopal, B. R., Raghavan, I., Mukherjee, M., & Wang, Z. Q. (2023). A cytochrome P450 CYP87A4 imparts sterol side-chain cleavage in digoxin biosynthesis. *Nature Communications*, *14*, 4042. https://doi.org/10.1038/s41467-023-39719-4

Clough, S. J., & Bent, A. F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal*, *16*(6), 735–743. https://doi.org/10.1046/j.1365-313X.1998.00343.x

Coley, P. D., & Barone, J. A. (1996). Herbivory and plant defenses in tropical forests. *Annual Review of Ecology and Systematics*, *27*, 305–335. https://doi.org/10.1146/annurev.ecolsys.27.1.305

Dong, L., Almeida, A., Pollier, J., Khakimov, B., Bassard, J. E., Miettinen, K., Stærk, D., Mehran, R., Olsen, C. E., Motawia, M. S., Goossens, A., & Bak, S. (2021). An Independent Evolutionary Origin for Insect Deterrent Cucurbitacins in Iberis amara. *Molecular Biology and Evolution*, *38*(11), 4659–4673. https://doi.org/10.1093/molbev/msab213

Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., Heckel, D. G., Der, J. P., Wafula, E. K., Tang, M., Hofberger, J. A., Smithson, A., Hall, J. C., Blanchette, M., Bureau, T. E., Wright, S. I., DePamphilis, C. W., Schranz, M. E., Barker, M. S., … Berenbaum, M. R. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(27), 8362–8366. https://doi.org/10.1073/pnas.1503926112

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and Plants : A Study in Coevolution. *Evolution*,

18(4), 586–608.

Fahey, J. W., Zalcmann, A. T., & Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, *56*, 5–51. https://doi.org/10.1016/S0031-9422(00)00316-2

Feeny, P. (1977). Defensive Ecology of the Cruciferae. *Annals of the Missouri Botanical Garden*, *64*(2), 221–234.

Feng, H., Chen, W., Hussain, S., Shakir, S., Tzin, V., Adegbayi, F., Ugine, T., Fei, Z., & Jander, G. (2023). Horizontally transferred genes as RNA interference targets for aphid and whitefly control. *Plant Biotechnology Journal*, *21*, 754–768. https://doi.org/10.1111/pbi.13992

Fraenkel, G. (1959). The Raison d'Etre of Secondary Plant Substances. *Science*, *129*, 1466–1470. https://doi.org/10.1007/978-3-662-65961-8_25

Frisch, T., & Møller, B. L. (2012). Possible evolution of alliarinoside biosynthesis from the glucosinolate pathway in Alliaria petiolata. *FEBS Journal*, *279*(9), 1545–1562. https://doi.org/10.1111/j.1742-4658.2011.08469.x

Gatto, L., Gibb, S., & Rainer, J. (2020). MSnbase, efficient and elegant R-based processing and visualisation of raw mass spectrometry data. *BioRxiv*.

Gatto, L., & Lilley, K. (2012). MSnbase - an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics*, *28*, 288–289.

Getman-Pickering, Z. L., Campbell, A., Aflitto, N., Grele, A., Davis, J. K., & Ugine, T. A. (2020). LeafByte: A mobile application that measures leaf area and herbivory quickly and accurately. *Methods in Ecology and Evolution*, *11*, 215–221. https://doi.org/10.1111/2041-210X.13340

Ghosh, S. (2017). Triterpene structural diversification by plant cytochrome P450 enzymes. *Frontiers in Plant Science*, *8*(November), 1–15. https://doi.org/10.3389/fpls.2017.01886

Gordon, H. T. (1961). Nutritional Factors in Insect Resistance to Chemicals. *Annual Review of Entomology*, *6*(1), 27–54. https://doi.org/10.1146/annurev.en.06.010161.000331

Graves, S., Piepho, H.-P., & with help from Sundar Dorai-Raj, L. S. (2023). *multcompView: Visualizations of Paired Comparisons*. https://cran.r-project.org/package=multcompView

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Iino, M., Nomura, T., Tamaki, Y., Yamada, Y., Yoneyama, K., Takeuchi, Y., Mori, M., Asami, T., Nakano, T., & Yokota, T. (2007). Progesterone: Its occurrence in plants and involvement in plant growth. *Phytochemistry*, *68*, 1664–1673. https://doi.org/10.1016/j.phytochem.2007.04.002

John, M. E., John, M. C., Ashley, P., MacDonald, R. J., Simpson, E. R., & Waterman, M. R. (1984). Identification and characterization of cDNA clones specific for cholesterol side-chain cleavage cytochrome P-450. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(18 I), 5628–5632. https://doi.org/10.1073/pnas.81.18.5628

Katz, E., Nisani, S., Sela, M., Behar, H., Chamovitz, D. A., Katz, E., Nisani, S., Sela, M., Behar, H., & Chamovitz, D. A. (2015). The effect of indole-3-carbinol on PIN1 and PIN2 in Arabidopsis roots. *Plant Signaling and Behavior*, *2324*, 16–19. https://doi.org/10.1080/15592324.2015.1062200

Kjær, A., & Gmelin, R. (1957). isoThiocyanates XXV. Methyl 4-isoThiocyanatobutyrate, a New Mustard Oil Present as a Glucoside (Glucoerypestrin) in Erysimum Species. In *Acta*

Chemica Scandinavica (Vol. 11, pp. 577–578). https://doi.org/10.3891/acta.chem.scand.11-0577

Kunert, M., Langley, C., Lucier, R., Ploss, K., Rodriguez Lopez, C. E., Serna Guerrero, D. A., Rothe, E., O'Connor, S. E., & Sonawane, P. D. (2023). A promiscuous CYP87A enzyme activity initiates cardenolide biosynthesis in plants. *Nature Plants*, https://doi.org/10.1038/s41477-023-01515-9

Lindemann, P. (2015). Steroidogenesis in plants – Biosynthesis and conversions of progesterone and other pregnane derivatives. *Steroids*, *103*, 145–152. https://doi.org/10.1016/j.steroids.2015.08.007

Lindemann, P., & Luckner, M. (1997). *Biosynthesis of pregnane derivatives in somatic embryos of Digitalis lanata*. *46*(3), 507–513.

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50*, W276–W279. https://doi.org/10.1093/nar/gkac240

Makarevich, I. F., Zhernoklev, K. V., Slyusarskaya, T. V., & Yarmolenko, G. N. (1994). Cardenolide-containing plants of the family Cruciferae. *Chemistry of Natural Compounds*, *30*(3), 275–289. https://doi.org/10.1007/BF00629957

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288–4297. https://doi.org/10.1093/nar/gks042

Miller, W. L., & Auchus, R. J. (2011). The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocrine Reviews*, *32*(1), 81–151.

https://doi.org/10.1210/er.2010-0013

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., Lanfear, R., & Teeling, E. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Moghe, G., & Last, R. L. (2015). Something old, something new: Conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiology*, *169*(November), pp.00994.2015. https://doi.org/10.1104/pp.15.00994

Monroe, J. G. (2017). *genemodel: Gene Model Plotting in R*. https://cran.r-project.org/package=genemodel

Nielsen, J. (1978a). Host Plant Discrimination Within Cruciferae: Feeding Responses of Four Leaf Beetles (Coleoptera: Chrysomelidae) To Glucosinolates, Cucurbitacins and Cardenolides. *Entomologia Experimentalis et Applicata*, *24*(1), 41–54. https://doi.org/10.1111/j.1570-7458.1978.tb02755.x

Nielsen, J. (1978b). Host plant selection of monophagous and oligophagous flea beetles feeding on crucifers. *Entomologia Experimentalis et Applicata*, *24*, 362–369.

Okamura, Y., Dort, H., Reichelt, M., Tunstrom, K., Wheat, C. W., & Vogel, H. (2022). Testing hypotheses of a coevolutionary key innovation reveals a complex suite of traits involved in defusing the mustart oil bomb. *Proceedings of the National Academy of Sciences*, *119*(51). https://doi.org/10.1073/pnas

Petschenka, G., Züst, T., Hastings, A. P., Agrawal, A. A., & Jander, G. (2023). Quantification of plant cardenolides by HPLC, measurement of Na+/K+-ATPase inhibition activity, and characterization of target enzymes. *Methods in Enzymology*, *680*, 275–302.

https://doi.org/10.1016/bs.mie.2022.08.003

Pilgrim, H. (1972). “Cholesterol side-chain cleaving enzyme” aktivität in keimlingen und in vitro kultivierten geweben von Digitalis purpurea. *Phytochemistry*, *11*(5), 1725–1728. https://doi.org/10.1016/0031-9422(72)85026-X

Pinheiro, J, & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer. https://doi.org/10.1007/b98882

Pinheiro, Jose, & Bates, D. (2023). *nlme: Linear and nonlinear mixed effects models* (R package version 3.1-162). https://cran.r-project.org/package=nlme

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Ramsey, J. S., Elzinga, D. A., Sarkar, P., Xin, Y. R., Ghanim, M., & Jander, G. (2014). Adaptation to Nicotine Feeding in Myzus persicae. *Journal of Chemical Ecology*, *40*, 869–877. https://doi.org/10.1007/s10886-014-0482-5

Ramsey, J. S., Wilson, A. C. C., de Vos, M., Sun, Q., Tamborindeguy, C., Winfield, A., Malloch, G., Smith, D. M., Fenton, B., Gray, S. M., & Jander, G. (2007). Genomic resources for Myzus persicae: EST sequencing, SNP identification, and microarray design. *BMC Genomics*, *8*(423). https://doi.org/10.1186/1471-2164-8-423

Renwick, J., Radke, C., & K, S.-G. (1989). Chemical consituents of Erysimum cheiranthoides deterring oviposition by cabbage butterfly, Pieris rapae. *Journal of Chemical Ecology*, *15*, 2161–2169.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rothschild, M., Alborn, H., Stenhagen, G., & Schoonhoven, L. M. (1988). A strophanthidin glycoside in siberian wallflower: A contact deterrent for the large white butterfly. *Phytochemistry*, *27*(1), 101–108. https://doi.org/10.1016/0031-9422(88)80598-3

Sachdev-Gupta, K., Radke, C., Renwick, J. A. A., & Dimock, M. B. (1993). Cardenolides from Erysimum cheiranthoides: Feeding deterrents to Pieris rapae larvae. *Journal of Chemical Ecology*, *19*(7), 1355–1369. https://doi.org/10.1007/BF00984881

Sainsbury, F., Thuenemann, E. C., & Lomonossoff, G. P. (2009). PEAQ: Versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal*, *7*, 682–693. https://doi.org/10.1111/j.1467-7652.2009.00434.x

Shinoda, T., Nagao, T., Nakayama, M., Serizawa, H., Koshioka, M., Okabe, H., & Kawai, A. (2002). Identification of a triterpenoid saponin from a crucifer, Barbarea vulgaris, as a feeding deterrent to the diamondback moth, Plutella xylostella. *Journal of Chemical Ecology*, *28*(3), 587–599. https://doi.org/10.1023/A:1014500330510

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(539). https://doi.org/10.1038/msb.2011.75

Stohs, S. J., & El-Olemy, M. M. (1971). Pregnenolone and Progesterone from 20a-hydroxycholesterol by Cheiranthus cheiri leaf homogenates. *Phytochemistry*, *10*(1967), 3053–3056.

[dataset] Strickler, S. R., Powell, A. F., Mueller, L. A., Zust, T., & Jander, G. (2019). *NCBI BioProject ID PRJNA563696. Rapid and independent evolution of ancestral and novel chemical defenses in a genus of toxic plants (Erysimum, Brassicaceae)*. NCBI.

https://www.ncbi.nlm.nih.gov/bioproject/PRJNA563696/

Suza, W. P., & Chappell, J. (2016). Spatial and temporal regulation of sterol biosynthesis in Nicotiana benthamiana. *Physiologia Plantarum*, *157*(2), 120–134. https://doi.org/10.1111/ppl.12413

Taussky, H. H., & Shorr, E. (1953). A Microcolorimetric Method for the Determination of Inorganic Phosphorus. *The Journal of Biological Chemistry*, *202*(2), 675–685. https://doi.org/10.1016/s0021-9258(18)66180-0

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, *44*, W232–W235. https://doi.org/10.1093/NAR/GKW256

Wang, W. C., Menon, G., & Hansen, G. (2003). Development of a novel Agrobacterium-mediated transformation method to recover transgenic Brassica napus plants. *Plant Cell Reports*, *22*, 274–281. https://doi.org/10.1007/s00299-003-0691-9

Weigel, D., & Glazebrook, J. (2006). Transformation of agrobacterium using the freeze-thaw method. *CSH Procols*, *7*, pdb.prot4666. https://doi.org/10.1101/pdb.prot4666

Wisecaver, J. H., Borowsky, A. T., Tzin, V., Jander, G., Kliebenstein, D. J., & Rokas, A. (2017). A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell*, *29*, 944–959. https://doi.org/10.1105/tpc.17.00009

Zhong, S., Joung, J. G., Zheng, Y., Chen, Y. R., Liu, B., Shao, Y., Xiang, J. Z., Fei, Z., & Giovannoni, J. J. (2011). High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harbor Protocols*, *6*(8), 940–949. https://doi.org/10.1101/pdb.prot5652

Zhou, Y., Ma, Y., Zeng, J., Duan, L., Xue, X., Wang, H., Lin, T., Liu, Z., Zeng, K., Zhong, Y.,

Zhang, S., Hu, Q., Liu, M., Zhang, H., Reed, J., Moses, T., Liu, X., Huang, P., Qing, Z., … Huang, S. (2016). Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nature Plants*, *2*(November), 1–8. https://doi.org/10.1038/nplants.2016.183

Züst, T., Strickler, S. R., Powell, A. F., Mabry, M. E., An, H., Mirzaei, M., York, T., Holland, C. K., Kumar, P., Erb, M., Petschenka, G., Gómez, J. M., Perfectti, F., Müller, C., Pires, J. C., Mueller, L. A., & Jander, G. (2020). Independent evolution of ancestral and novel defenses in a genus of toxic plants (Erysimum, brassicaceae). *ELife*, *9*, 1–42. https://doi.org/10.7554/eLife.51712

DATA ACCESSIBILITY

The data that support the findings of this study are available in the supplementary material of this article. Raw sequencing reads are openly available on NCBI (PRJNA1015726).

BENEFIT-SHARING

Benefits from this research accrue from the sharing of our data and results on public databases as described above. Seeds from mutant lines are available upon request.

AUTHOR CONTRIBUTIONS

GCY, MLA, HDF, MM, and GJ designed the research; GCY, MLA, HDF, MM, and APH performed the research; GCY and MLA analyzed data; GCY, AAA, and GJ wrote and edited the paper.

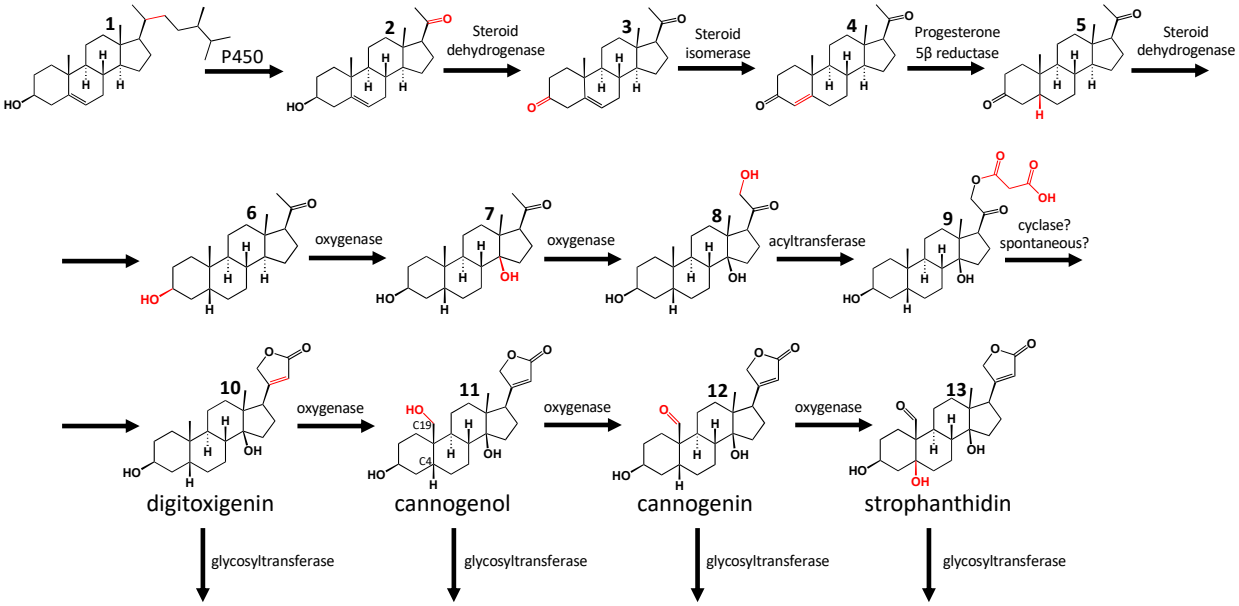

Figure 1. Proposed cardenolide biosynthetic pathway. The first step (sterol side chain cleavage), and the conversion of digitoxigenin **10** to cannogenol **11** are discussed in this paper. One to two sugars may be attached at the 3-hydroxyl group.

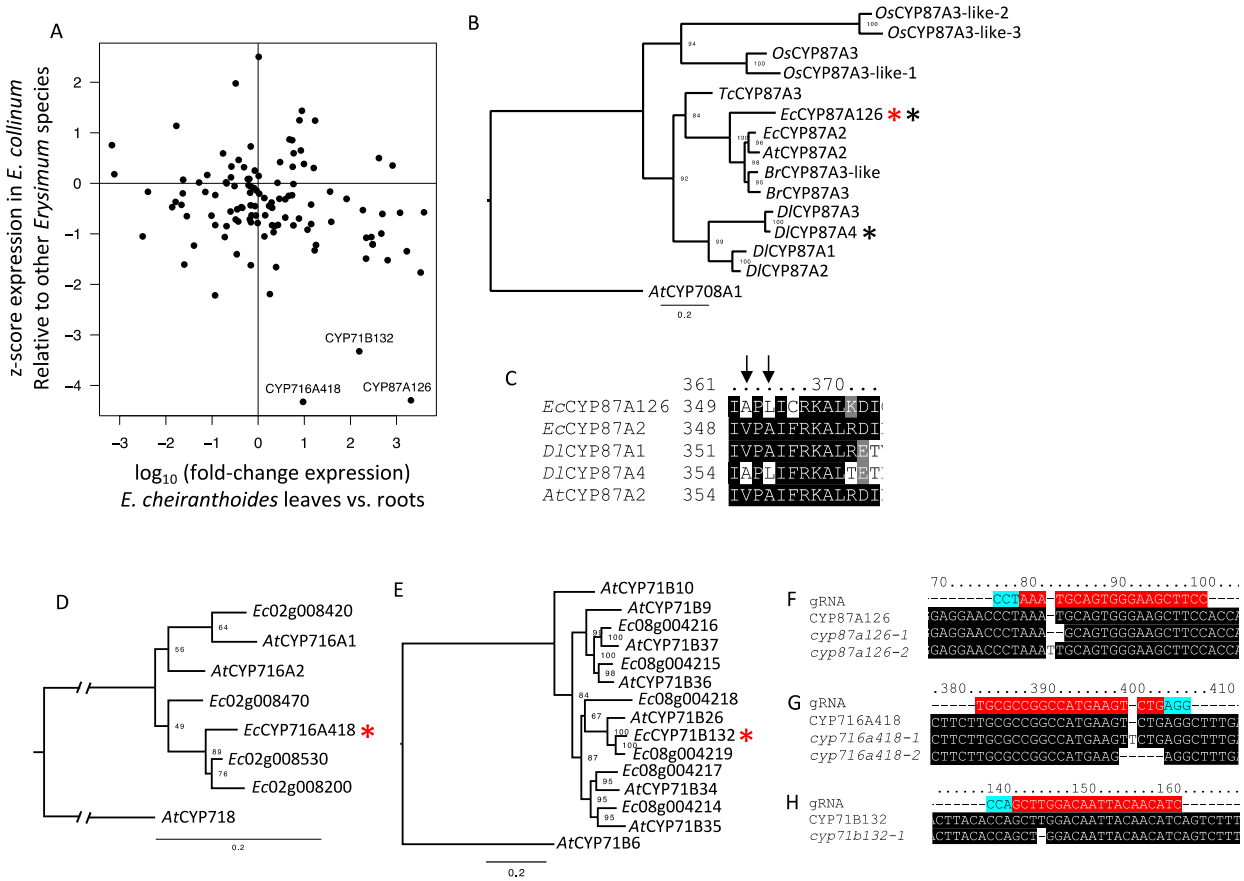

Figure 2. Identification and knockout of candidate cytochrome P450s. (A) Expression pattern of cytochrome P450 monooxygenases in *Erysimum.* Fold change in expression between young leaves and roots is plotted against z-score of expression in *E. collinum* relative to 48 *Erysimum* species. Genes with high expression in young leaves, where cardenolides are synthesized, and low expression in *E. collinum*, which does not produce cardenolides, are good candidates for involvement in cardenolide biosynthesis. (B, D, E) Gene trees of candidate cytochrome P450s. Species abbreviations: *At* (*Arabidopsis thaliana*), *Br* (*Brassica rapa*), *Dl* (*Digitalis lanata*), *Ec* (*Erysimum cheiranthoides*), *Os* (*Oryza sativa*), and *Tc* (*Theobroma cacao*). Candidate genes are marked with a red star. Black stars indicate genes that have been previously shown to have sterol side chain cleaving activity. (C) Alignment of selected CYP87A proteins. Convergent amino acid substitutions that are critical for sterol side chain cleaving activity are marked with an

arrow. (F,G,H) Location and sequence of Cas9 protospacers (red, with 3’ NGG PAM in turquoise) used for generation of mutant lines with wildtype and mutant sequences. Number of base pairs from the start of the coding sequence are indicated.

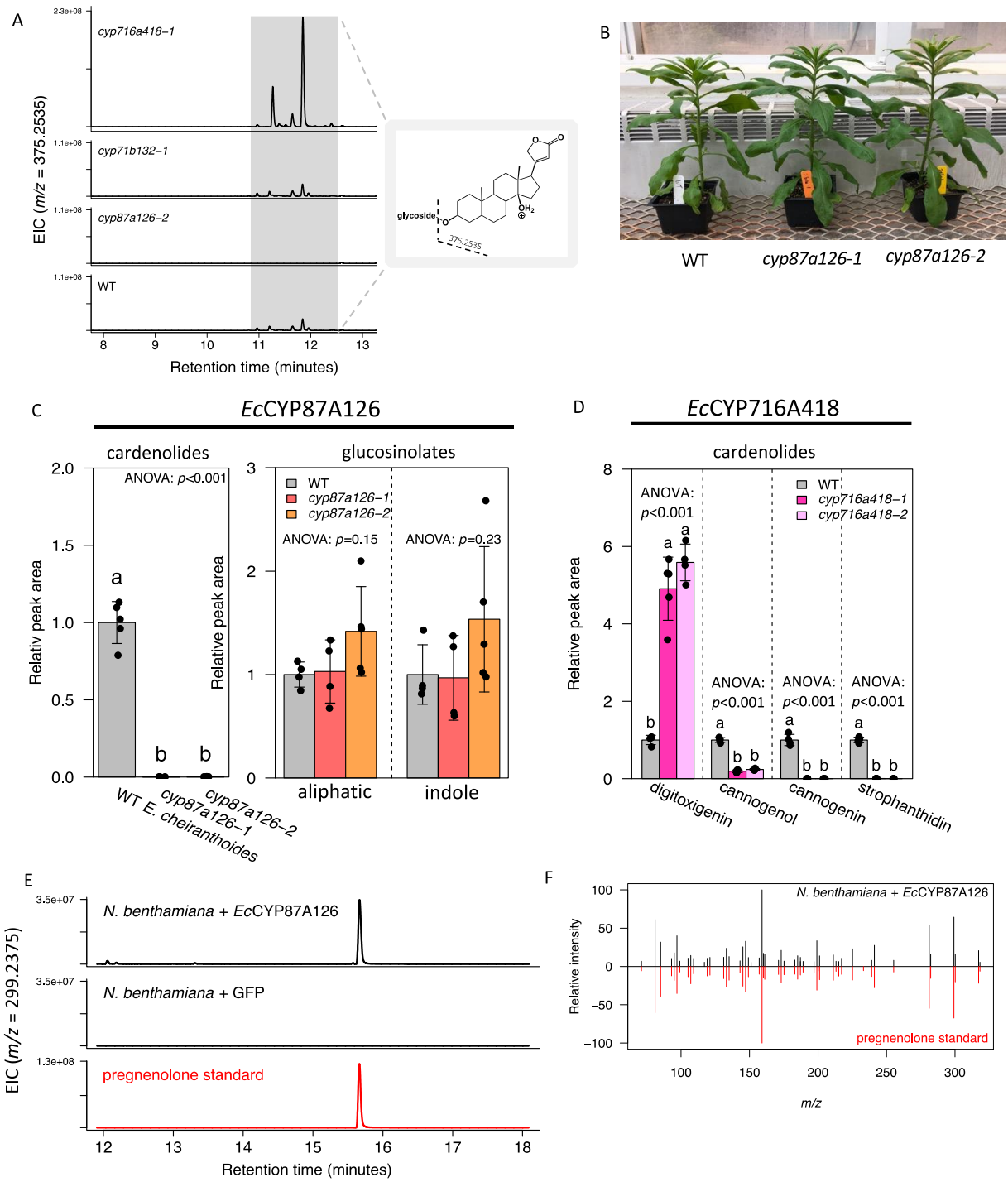

Figure 3. *Ec*CYP87A126 and *Ec*CYP716A418 mutant lines have altered cardenolide content. (A) Extracted ion chromatograms from a representative mutant plant for each candidate gene. *m/z* = 375.2535 is a fragment common to all digitoxigenin-containing cardenolides in positive

electrospray ionization. (B) Despite lacking cardenolides, *cyp87a126* mutant plants show no obvious growth phenotype. (C) Total cardenolide and glucosinolate-related peak area in *cyp87a126* mutants compared to wildtype. (D) Cardenolide abundance by genin in *cyp716a418* mutant lines compared to wildtype. For all plots: error bars indicate ±sd; letters indicate $p<0.001$, Tukey's HSD. (E) Extracted ion chromatograms of *m/z* = 299.2375 (pregnenolone [M-H2O+H]+) of *N. benthamiana* leaves expressing EcCYP87A126 or a GFP control. A pregnenolone standard (red) was infiltrated into a separate *N. benthamiana* leaf to account for any potential modifications made by endogenous enzymes. (F) MSMS spectra of pregnenolone standard (red) compared to product of expression of *Ec*CYP87A126 in *N. benthamiana* (black).

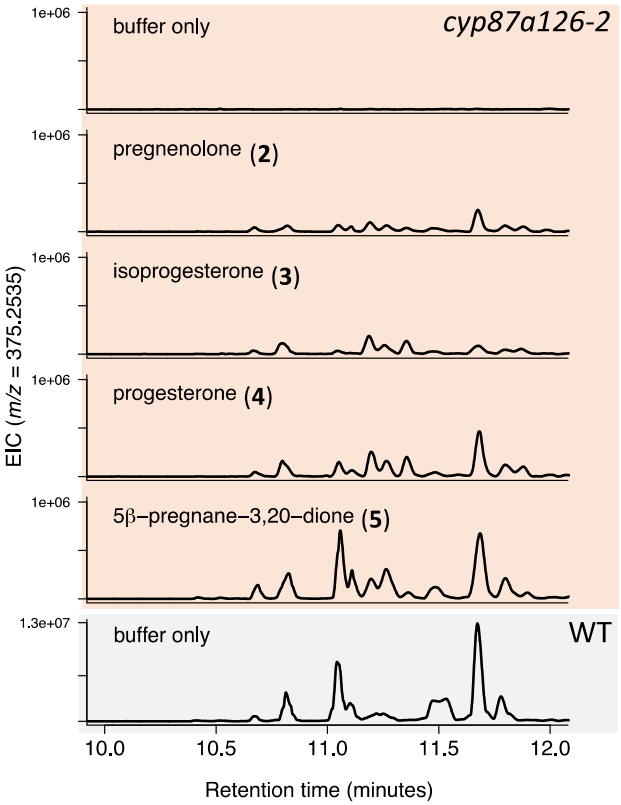

Figure 4. Rescue of cardenolide biosynthesis in *cyp87a126-2* mutant line. ESI+ extracted ion chromatograms of *m/z* = 375.2535 from *cyp87a126-2* plants two days after substrate infiltration (orange background) compared to wildtype (grey background). See Figure 1 for molecular structures.

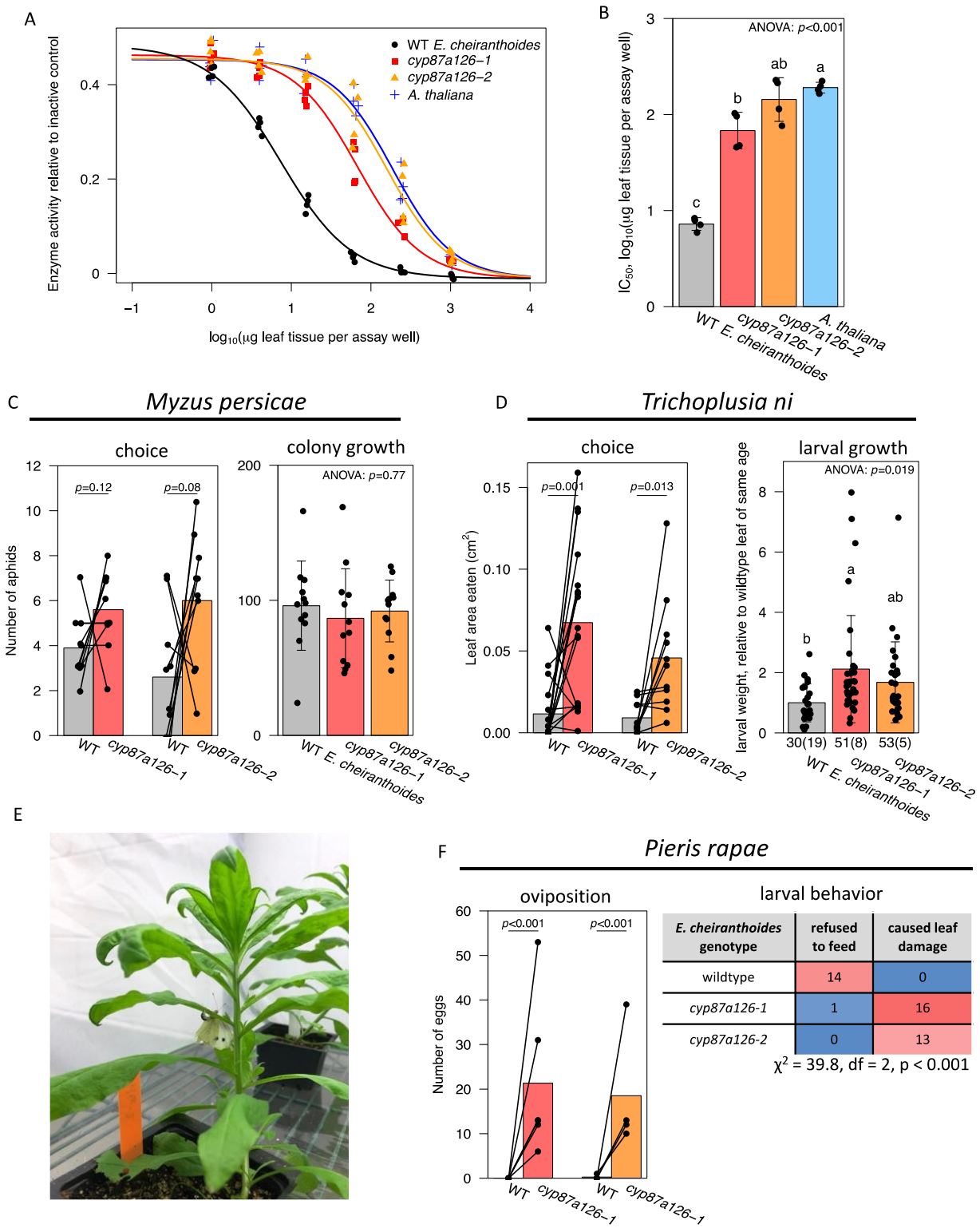

Figure 5. Functional implications of *cyp87a126* knockout. (A) Na+/K+-ATPase inhibition assay

for leaf extracts of *cyp87a126* mutant lines compared with wildtype (WT) *E. cheiranthoides* and *A. thaliana* Col-0 as a cardenolide-free control. Inhibition curves calculated from four replicates of each tissue. (B) Half-maximal inhibitory concentration of leaf extracts estimated from inflection point of inhibition curves. (C) *M. persicae* assays: binary choice as measured by aphid position after 24 hours; colony growth of five synchronized aphids after 9 days. (D) *T. ni* assays: binary choice, leaf area eaten after two days; growth, larval weight after 12 days of feeding, normalized by leaf position to remove effect of leaf age. Numbers below plot indicate caterpillars surviving or dying (in parentheses) after 8 days. (E) *P. rapae* butterfly ovipositing on cardenolide-free *E. cheiranthoides* plant. (F) *P. rapae* assays: oviposition in binary choice assay; larval behavior when confined to an individual leaf. For all plots: error bars indicate ±sd; letters indicate $p<0.05$, Tukey's HSD. P-values are from paired t-tests in choice assays or chi-squared test in oviposition and larval behavior assays.