Gene loss associated with plasticity-first evolution in Heliconius butterflies

Erika C. Pinheiro de Castro[1], Francesco Cicconardi[2], Ian A. Warren[1], Nicol Rueda-M[3,4], Camilo Salazar[4], Søren Bak[5], Stephen H. Montgomery[2] & Chris D. Jiggins[1]

[1]Department of Zoology, University of Cambridge.
[2]School of Biological Sciences, University of Bristol.
[3]Darwin Tree of Life, Welcome Sanger.
[4] School of Sciences and Engineering, Biology Program, Universidad del Rosario.
[5]Department of Plant and Environmental Science, University of Copenhagen.

ABSTRACT

Phenotypic plasticity occurs when a genotype can produce more than one phenotype under different environmental conditions. Genetic accommodation allows plastic phenotypes to be tuned to new environments and eventually to even be lost via canalization/assimilation. The colourful and toxic Heliconiini butterflies have biochemical plasticity – they either sequester their cyanogenic glucosides (CG) from their larval hostplant or biosynthesize them when compounds for sequestration are not available. Here, we traced the evolution of CG biosynthesis in Heliconiini butterflies, a fundamental component of this biochemical plasticity. We first CRISPR-edited the *CYP405s*, which encode the first enzyme of this pathway in *Heliconius erato,* and confirmed that *CYP405*-knockout caterpillars do not biosynthesize CGs. Then, we identified the *CYP405*s and *CYP332*s (the second gene in this pathway) in 63 species of the subfamily Heliconiinae, as well as in other lepidopterans. Most lepidopterans have a *CYP332*, but *CYP405* is mostly found in butterflies and has been duplicated in all *Heliconius* species. Both genes were independently co-opted into CG biosynthesis in the Heliconiinae butterflies and *Zygaena* moths, in which they were first functionally characterized. Finally, we performed ancestral reconstruction analyses of biochemical plasticity using data from over 700 Heliconiini butterflies. We demonstrated that although plasticity was ancestral in the whole tribe and allowed these butterflies to increase their hostplant range, it might have been lost in few specialized clades, such as the Sapho clade that just sequester CG from their hostplants. This clade has several *CYP405* copies, but they lack structurally important P450 domains, which may explain why biosynthesized CGs are virtually lost in species of the Sapho clade. Our findings represent one of the few examples of plasticity-first evolution in which the genetic mechanisms associated with its accommodation/assimilation are known. This study also emphasizes the importance of co-option in the evolution of complex traits, such as toxicity.

INTRODUCTION

Adaptive phenotypic plasticity occurs where one genotype produces multiple phenotypes depending on environmental conditions, allowing species and populations to maximise fitness across different environments. Such plasticity could also promote diversification under the 'plasticity first' hypothesis, by allowing populations to colonise novel environments and niches more readily (Levis & Pfennig, 2019). Although most fitness traits have some degree of plasticity, phenotypic plasticity has been mainly studied in morphological and physiological traits, which produce a visible phenotype (Sommer, 2020). Phenotypic plasticity in chemical traits has been understudied.

Chemical defences create an "enemy-free" niche for protected organisms to radiate and speciate (Ehrlich & Raven, 1964). Yet, chemically protected species need to evolve not only mechanisms to acquire toxins, but to also tolerate/resist them and both mechanisms needs to be present for toxicity to translate into an adaptive advantage (Beran et al., 2019). To make the evolutionary trajectory of chemical defences even more complicated, most mechanisms to acquire and tolerate toxins are often biochemical pathways – such that all substrates, enzymes and intermediates need to be present for the advantageous product to be formed. Chemical defence is therefore a complex trait which may be difficult to evolve, due to the high number of elements. As such, gene co-option (Petschenka & Agrawal, 2016), or horizontal gene-transfer (Piel et al., 2004; Wybouw et al., 2016) might facilitate the acquisition of toxicity. Furthermore, the evolution of toxicity and its complexity goes beyond genetics: For most heterotrophic organisms, diet also plays a crucial role in chemical defences, by either providing toxins that can be directly sequestered or the precursors for their biosynthesis (Nishida, 2014). This implies that: (1) the evolution of toxicity often cannot be separated from the evolution of diet specialization; and (2) there is substantial phenotypic plasticity in animal toxicity in response to diet, although this is rarely considered as an example of adaptive plasticity.

Cyanide is a toxin that disrupts the mitochondrial electron transport chain and is a breakdown product of cyanogenic glucosides (CG). Some Lepidoptera have evolved to detoxify cyanide and feed on plants protected by CG (Zagrobelny et al., 2018). While lepidopteran adaptations to detoxify cyanide came from both horizontal gene transfer (i.e. β-cyanoalanine synthase - Wybouw et al., 2014) and co-option of genes with existing functions (i. e. Rhodanase - Steiner et al., 2018), the evolution of the mechanisms used to acquire CG are less well understood. Possibly as a counter-adaptation against these specialized lepidopterans, some plant lineages/populations have disabled sequestration by changing the structure of their CGs - e. g. certain *Passiflora* species against heliconiine butterflies (de Castro et al., 2019) - or even by becoming acyanogenic - e.g. *Turnera ulmifolia* in Mexico against *Euptoieta* butterflies (Schappert & Shore, 1999); and *Lotus corniculatus* against *Zygaena* moths (Smith

et al., 1954). These changes in the CG composition of the hostplant would affect the biochemical phenotype of the lepidopteran that sequester these toxins. In return, some lepidopterans have evolved mechanisms to maintain their toxicity regardless of the CG profile of their host plants by biosynthesizing their own CGs.

Both higher plants and lepidopterans *de novo* biosynthesize CGs using the same three enzymatic steps: they convert amino acids into oximes and then into cyanohydrins, which are then stabilized with a glycosylation - forming a CG (Figure 1). To assemble this pathway, they have recruited cytochrome P450s (P450s/CYPs) and UDP-glycosyltransferases (UGTs) (Figure 1). Despite having the same precursors, intermediates, products, and catalytic reactions, *de novo* biosynthesis of CGs evolved independently in lepidopterans and plants (Jensen et al., 2011). Moths of the family Zygaenidae and butterflies of the subfamily Heliconiinae are the only insects currently known to *de novo* biosynthesize CGs and both clades produce the same aliphatic compounds. In lepidopterans, CG biosynthesis has been fully characterized in the six-spot burnet moth *Zygaena filipendulae* (Jensen et al., 2011), and is composed of CYP405A2, CYP332A3, and UGT33A1. Three *CYP405* and one *CYP332* homologues have been found in the genome of the butterfly *Heliconius melpomene*: *HmCYP405A5*, *HmCYP405A5, HmCYP405A6,* and *HmCYP332A1* (Chauhan et al., 2013). These copies have all the conserved domains/motifs of a functional P450 and are expressed in life stages of intense CG biosynthesis (de Castro, Demirtas, et al., 2019), but their involvement in CG biosynthesis in heliconiines has not yet been empirically confirmed. Since moths and butterflies split ~100 MYA (Kawahara et al., 2019), if CG biosynthesis in these two clades involves orthologous genes it is likely that these were independently co-opted to this pathway, but this hypothesis remains to be tested.

While biosynthesis of CG allows them to maintain toxicity regardless of hostplant chemistry, it may have higher fitness costs than CG sequestration. Cyanogenic lepidopterans only biosynthesize CGs when they cannot sequester these compounds during larval feeding (de Castro et al., 2020; Fürstenberg-Hägg et al., 2014). The CG sequestered by heliconiines (cyclopentenic) have a different structure to those that are biosynthesized (aliphatic) (de Castro et al., 2018). The difference in structure between sequestered and synthesised compounds enables the use of target metabolomics to identify which strategy wild-caught heliconiine individuals used to acquire their CG reservoir (de Castro et al., 2019; de Castro et al., 2020). This creates a unique opportunity to study the importance of phenotypic plasticity in the evolution of toxicity. Moreover, phenotypic plasticity can itself evolve when selection acts to change reaction norms. The different degrees of hostplant specialization within Heliconiini, in which different clades tend specialized on specific subgenera of *Passiflora*, with some feeding on several hostplants, while others are completely monophagous, creates an unique system in which to understand how specialization affects biochemical plasticity. Finally, heliconiine butterflies

have been studied for over 150 years, with recent efforts resulting in assembled reference genomes (Cicconardi et al., 2022) and target-metabolomics data for most species of this tribe (de Castro, Zagrobelny, et al., 2019; Sculfort, Castro, et al., 2020), which allows powerful comparative analyses that can reveal the genetic basis of the maintenance and loss of biochemical plasticity in specific lineages.

We first utilize the gene-editing technology CRISPR-Cas9 to confirm the role of *CYP405As* in CG biosynthesis in *Heliconius*. Next, we explored the evolutionary history of these genes across 58 heliconiine genomes (Cicconardi et al., 2023) and used molecular evolution analysis, combined with target-metabolomic data for over 700 individuals (de Castro, Zagrobelny, et al., 2019; Sculfort, Castro, et al., 2020) to evaluate the importance of "biosynthesis versus sequestration" in each clade and test for selection on specific branches of the phylogenetic tree. We hypothesize that *CYP405s* are only duplicated in *Heliconius* which are more toxic than other heliconiines and that these copies might have been lost in some *Heliconius* clades that have specialized in sequestration and might have lost biochemical plasticity.

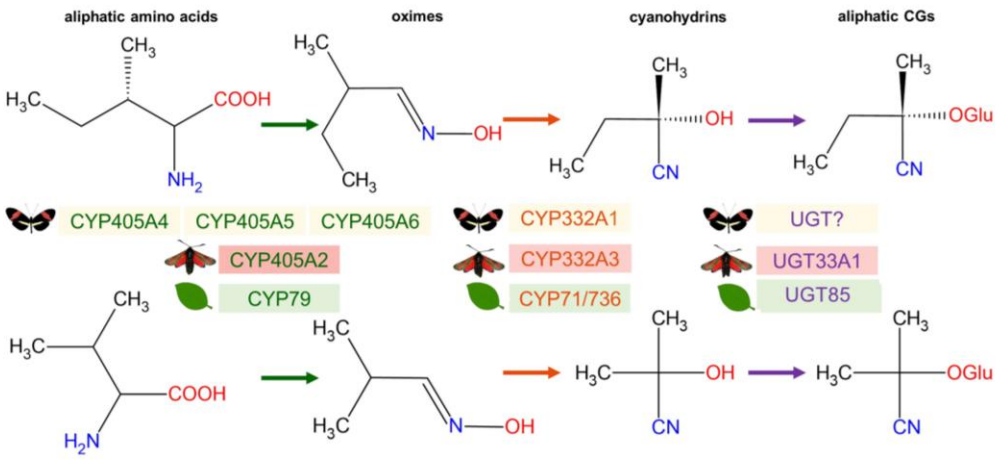


Figure 1. De novo biosynthesis of aliphatic cyanogenic glucosides (CGs) linamarin (bottom) and lotaustralin(top), which are catalyzed by the same three enzymatic steps and products are determined by which amino acid is used as subtract (valine or isoleucine, respective). Colourful boxes correspond to taxa in which this pathway has been characterized/described: *Heliconius melpomene* (yellow boxes) (Chauhan et al., 2013), *Zygaena filipendulae* (red boxes) (Jensen et al., 2011), and plants (green boxes), such as *Lotus* and cassava. The colours of the arrows correspond to the tree catalytic steps of CG biosynthesis and match the colour of the enzyme names that catalyse them. This pathway has the same enzymatic steps in both plants and lepidopterans because of convergent evolution.

METHODS

CRISPR-editing of *CYP405As* in *Heliconius erato demophoon*

CRISPR-Cas9 gene editing was used to confirm that *CYP405A*, previously characterized in *Zygaena* moths, is also involved in the *de novo* biosynthesis of CG in heliconiines. This experiment was performed using a population of *H. erato demophoon* kept under insectary conditions (Supplementary Methods) This species has three copies of *CYP405As*: *HedCYP40A5A4*, *HedCYP40A5A6a* and *HedCYP405A6b*. Single-guides RNAs (sgRNA) targeting these genes were designed using the software Geneious (v2020.1) by identifying N_{20}NGG sites within exons of the target-genes and internally blasting these sequences against the genome of *H. erato demophoon* v1 (http://ensembl.lepbase.org/Heliconius_erato_demophoon_v1/Info/Index) to remove sgRNA candidates with possible off-targeting sites. sgRNA4 (GCCTAACGGTGATCCTCAGAAGG) was selected targeting a region in Exon 7 of *HedCYP40A5A4,* and sgRNA6ab (GCCTAACGGTGACCCCATGAAGG) targeting the same region in both *HedCYP40A5A6a* and *HedCYP405A6b.* Unfortunately, it was not possible to design a sgRNA that could distinguish between the CDS of these two orthologues due to their similarity. sgRNAs were ordered from Synthego (https://www.synthego.com/) and dissolved following the manufacturer's instructions. Freshly-laid eggs of *H. erato* (within 1-2h of oviposition) were glued on microscope slides and injected with a mix containing 200 ng Cas9 (TrueCut Cas9 ProteinV2, Invitrogen) with 200 ng sgRNA per µL. As a control for each batch, some eggs were glued on microscope slides and injected with Cas9 only (200 ng/ul) or kept non-injected. Larvae from control and CRISPR-eggs were individually raised in tubes and fed *ad libitium* with *Passiflora biflora* leaves, which does not contain CGs that can be sequestered by heliconiines (forcing them to biosynthesize their own CGs). Due to high mortality in the experimental group (injected with the sgRNAs), larvae were sampled after 7-8d of hatching. For the phenotypic analyses, larvae were weighed at the day of sampling and part of their tissue used for quantification of CGs using LC-MS following the extraction and analytical methods described in de Castro et al. (2019). ANOVA was performed in R (R Core Team, 2018) to test if the weight and CG content of CRISPR-edited larvae was lower than the control.

Genotyping the CRISPR-edited CYP405s and control larvae

Nanopore Minion sequencing of PCR products was used for genotyping using the four primer PCR kit (Oxford Nanopore Technologies plc., Oxford UK, Kit: SQK-PBK004). DNA was extracted using an in house bead extraction protocol ((Kučka & Chan, 2022)). The following primers were used to amplify DNA from all three copies of the gene and add adapters for sequencing (Forward: TTTCTGTTGGTGCTGATATTGCAAGATTGACTCCATTGCACACT, Reverse : ACTTGCCTGTCGCTCTATCTTCCAATGACATCAGGTACATTGCG). PCR products were visualised on an

agarose gel prior to sequencing library preparation and barcoding, which was performed using the manufacturer's protocol (kit: SQK-PBK004). Sequencing was carried out until each library had 10,000 reads. Basecalling was done by the guppy basecaller (Wick et al., 2019), adapter trimming and quality control was performed using pore-chopper (https://github.com/rrwick/Porechop) and quality control was performed using MinionQC.r (Lanfear et al., 2019). Minimap2 (Li, 2018) was used to align the reads to a contig containing all three gene copies (Nadeau et al., 2014). Samtools (Danecek et al., 2021) was used to convert resulting output to a BAM file, and then to sort and index the files. The alignments were visualized in Geneious v2020.1.

Biosynthesis versus sequestration profile of heliconiines and ancestral reconstruction analyses

Published datasets from de Castro et al. (2019), Sculfort et al. (2020) and Rueda-M et al. (*in prep*) on the CG profile of different heliconiine species were used for this analysis, comprising 743 butterflies of 43 species. The concentration of each CG (µg/mg), as well as the total CG concentration (µg/mg), found in each examined individual was extracted from the original datasets. The aliphatic CGs linamarin, lotaustralin and epilostaustralin were labelled as 'biosynthesized CG' (Table S1). The cyclopentenyl CGs deidaclin/tetraphyllin A, tetraphyllin B, epivolkenin, dynocardin, and dyhidrogynocardin (Table S1) were labelled as 'sequestered CGs' (Table S1). The sum of the concentration of all biosynthesized CGs and all sequestered CGs in each individual was divided by its total CG concentration to calculate '% of biosynthesis and % of sequestration, respectively. An average of % of biosynthesis and % of sequestration for each species was calculated using all related individual entries and plotted in bi-directional bar charts per species following the tree hypothesis from Cicconardi et al. (2022). This dataset was also used to perform ancestral reconstruction analysis on biochemical plasticity, using the function "anc.ML" the package "Phytools"(Revell, 2024).

Cytochrome P450 CYP405s and CYP332: phylogeny & evolutionary pressure analyses

To evaluate the phylogenetic relationship of *CYP405s* and *CYP332s* and test whether the co-opted genes in *Zygena filipendula* and Heliconiinae belonged to the same orthogroups (OGs), we first downloaded all P450 genes of the CYP4 and CYP3 clades from insectp450.net database (Wu et al. 2025) and selected all sequences from Lepidoptera annotated as complete, for a total of 9230 sequences from 26 families. Amino-acid sequences were aligned with ClustalW and used to generate the full maximum likelihood phylogenetic tree using IqTree2 (Minh et al. 2020) using the LG+G4+F and 1000 ultrafast bootstrap replicates. Once the OGs were identified, we added all those genes to our list to generate the full OG phylogenetic trees using nucleotide sequences (see below). The protein sequence CYP405s and CYP332s from *Heliconius melpomene* and *H. erato* annotations were obtained from NCBI, as well as the sequences from *Bombyx mori*. All these sequences were used as queries in

EXONERATE (Slater & Birney, 2005) [setting: -n 3 --model protein2genome --percent 50 --refine full] to annotate coding exons and introns in 58 heliconiini and other two heliconiinae (*Speyeria mormonia* and *Brenthis ino*) species. The resultant nucleotide sequences were subsequently mapped using MINIMAP2 (LI, 2018) to explore genome assemblies where no record was found in the EXONERATE mapping to minimise false negative hits. The results were used as a template to manually check and correct the sequences in all species. All the amino acid sequences obtained were processed with CDSEARCH webserver (Lu et al., 2020) and BLASTP webserver (Camacho et al., 2013) to verify the presence and the completeness of the conserved domain (PFAM00067) and possible gene-fusions. All nucleotide sequences were aligned using MACSE2 (Ranwez et al., 2018), and an initial maximum likelihood phylogenetic gene tree reconstruction was generated as implemented in IQ-TREE2 [-M MFP -B 5000]. At this point all sequences with putative stop-codons were removed and alignment and ML gene tree recalculated and used as input for follow-up analyses.

Several tests for selective forces were carried out for *CYP405* and *CYP332* in the tribe Heliconiini: Initially, all sequences from each locus were realigned separately and model test and the overall ω (hyphy acd Universal $FASTA MG94CUSTOMCF3X4 Global) were calculated as implemented in HYPHY v2.5.32(MP) for Linux on x86_64 (Kosakovsky Pond et al., 2020) batch language. In addition, the Single-Likelihood Ancestor Counting (SLAC) method (Kosakovsky Pond & Frost, 2005), as implemented in the HYPHY batch language, was used to evaluate the total number of sites under purifying and positive selection,with a *P* value of 0.05 as threshold. To detect signatures of pseudogenisation, a scan for relaxed selection was initially performed for all branches of the gene trees, using thein RELAX.BF (Wertheim et al., 2014) code (https://github.com/veg/hyphy-analyses) as implemented in the HYPHY batch language. Signatures of diversifying positive selection were detected by scanning all internal branches of the whole CYP405 and CYP332 phylogeny using Adaptive Branch-Site Random Effects Likelihood (ABSREL) method (Kosakovsky Pond et al., 2011; Smith et al., 2015) as implemented in the HYPHY batch language, correcting for multiple tests and using a final adjusted *P* value threshold of 0.05.

Previous studies on the CG profile of heliconiines, i.e. studies de Castro et al. (2019) and Sculfort et al. (2020), found that some *Heliconius* species in the Sara-sapho clade have specialized on sequestration and might even have lost their biosynthetic ability (e.g. *H. sapho*, *H. antiochus* and *H hewitsoni*). In order to detect possible shifts in the selection acting upon these biosynthetic genes in this group, two test were performed: At first, we applied BUSTED-PHENOTYPE (BUSTED-PH.bf at https://github.com/veg/hyphy-analyses), a method that detects evidence of episodic diversifying selection in association with feature/phenotype/trait. This is built on BUSTED (Murrell et al., 2015) and it is implemented under HYPHY batch language. BUSTED-PHENOTYPE detects evidence of episodic diversifying selection within groups and between groups. Nevertheless, because BUSTED-PHENOTYPE

does not describe the direction of the selection, we performed the RELAX (Wertheim et al., 2014) test to investigate if the phenotype/trait branches experienced relaxation of intensification, therefore the direction of the selection.

RESULTS

H. erato with CRISPR-edited CYP405s do not biosynthesize CGs

We first used gene knockout experiments with CRISPR/Cas9 to confirm the role of *CYP405* genes in CG biosynthesis. Fresh eggs of *H. erato demophoon* were injected with a sgRNA targeting either *HedCYP405A4* or the duplicated *HedCYP405A6* (experimental group), with controls injected with just the Cas9 enzyme. Individuals in the experimental group showed reduced weight as compared to the control in the first week (Kruskal-Wallis, X^2 = 38.065, Df=2, p= 5.423e-09) and most failed to develop to final instar (Figure 2A). Target-metabolomics showed that larvae with a deletion in the CDS of *CYP405A4 or CYP405A6s* lacked or had very low biosynthesized CG in comparison with others (Figure 2B). Larvae in the experimental group that did not have deletions in the *CYP405A4 or CYP405A6s* also had more CG than those in which gene-editing was successful. Nanopore sequence of CRISPR-edited and control individuals also confirmed that the designed sgRNAs were specific and only caused deletions in the target *CYP405* orthologue (Figure S1)

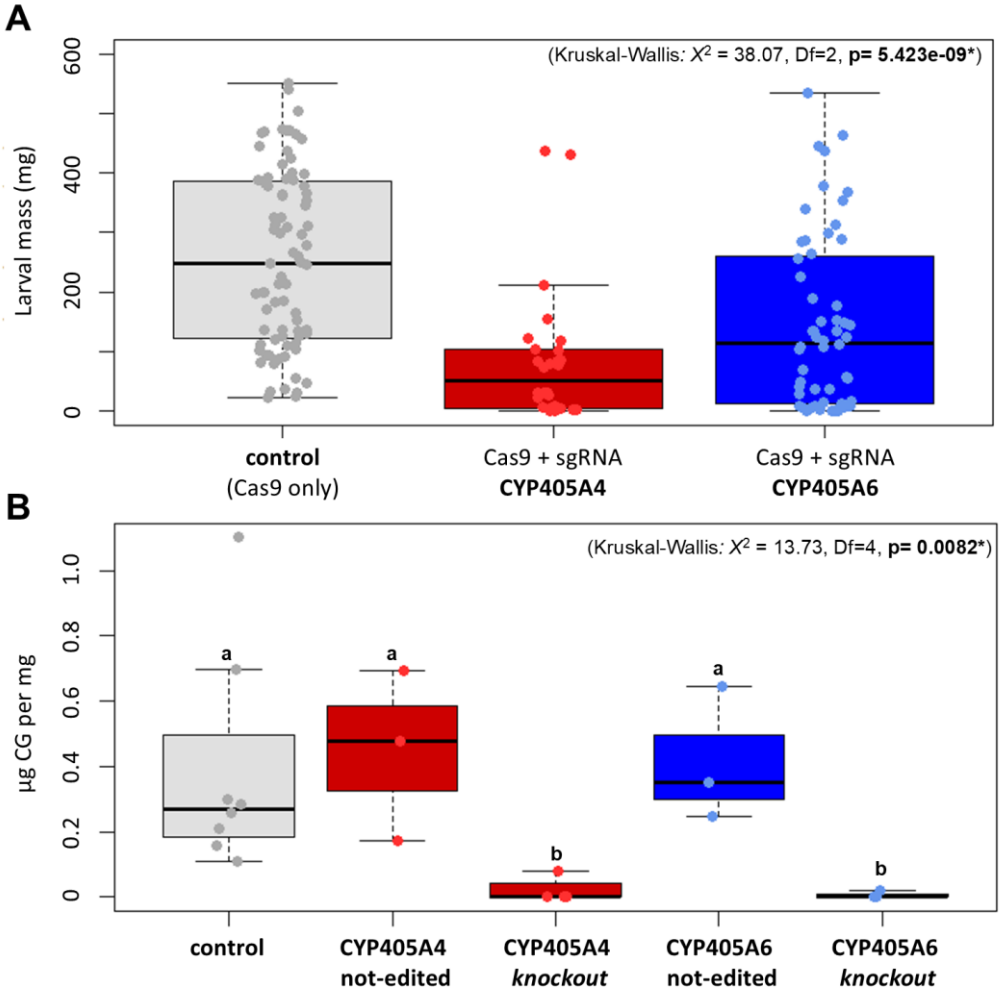


Figure 2 A) Larval mass (mg) of all 7d old larvae from injected eggs of *H. erato*. **B)** CG concentration (ug/mg) of genotyped larvae. Controls (grey) were injected with Cas9 only, whereas others were injected with Cas9 and the sgRNA specifically targeting either *CYP405A4* (red) or *CYP405A6* (blue). *CYP405A* knockouts had a deletion in the target copies, whereas not-edited larvae had the same genotype as control, even though they were injected with the respective sgRNAs.

CYP405 and CYP332 were coopted twice into CG biosynthesis in Lepidoptera

To evaluate whether *CYP405s*, as well as *CYP332s,* were independently coopted into CG biosynthesis in butterflies and moths, we constructed a maximum likelihood tree using the amino acid sequences of all complete CYP3 and CYP4 cytochrome P450s in lepidopterans (~9230 genes) (Figure S2). Whereas *CYP332* is present in most lepidopterans (Figure 3B), *CYP405* is found only in a few moth species (Figure 3A). All *CYP405s* and *CYP332s* in the subfamily Heliconiinae are very conserved and are monophyletic relative to other Lepidoptera. The two *CYP405s* from *Zygaena filipendulae* cluster with the orthologues in all butterflies instead of with other moths, and the same pattern is observed for CYP332, supporting our hypothesis of independent cooption

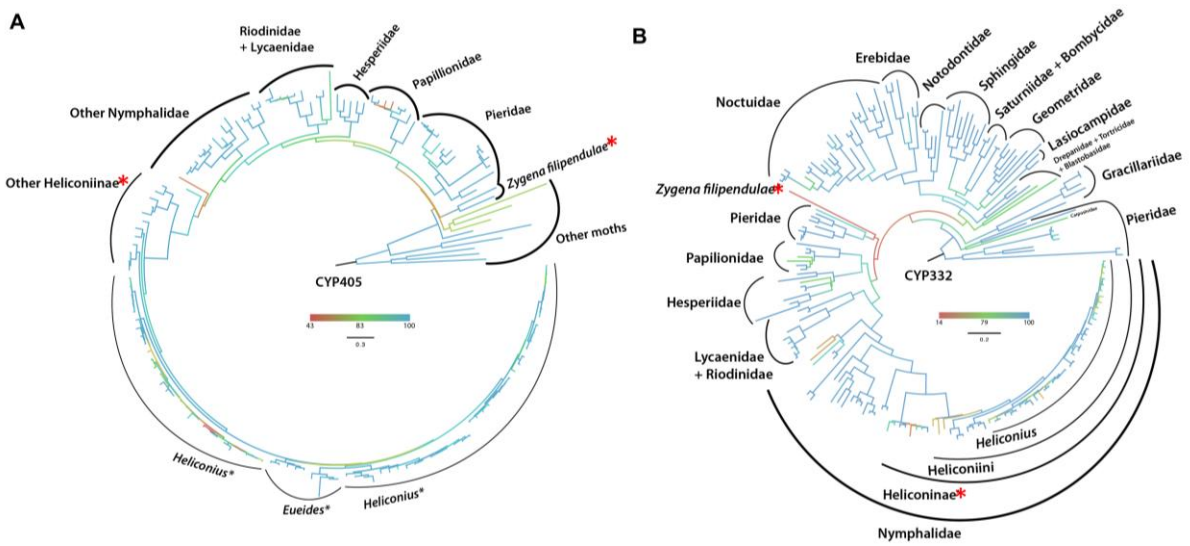


Figure 3. Maximum-likelihood tree of **A)** *CYP405* and **B)** *CYP332* in Lepidoptera. Taxa that are known to biosynthesize CGs is marked with *. Branch colour correspond to bootstrap values (Red=low to Blue=High). These two trees were subset from a tree including all complete CYP3 and CYP4 P450 in Lepidoptera.

Variation in the balance of CG biosynthesis and sequestration across the phylogeny

We next reconstructed the evolutionary history of biochemical plasticity in the tribe Heliconiini using a phylogenetic approach, using datasets on the cyanogenic glucoside (CG) profile of 743 heliconiine butterflies from other studies (Figure 4). At the base of the heliconiine phylogeny (Figure 4, "Other genera") CG biosynthesis is the dominant strategy. Nevertheless, sequestered CGs were found in a few individuals of all basal species (*Philaethria dido*: 1 out of 6*; Dryas iulia:* 15 of 45; *Agraulis vanillae:* 7 out 12; and *Dryadula phaetusa:* 1 out of 15), except for *Dione juno* (0 out of 15). This suggests that although biosynthesis is the dominant strategy, the ancestral state involved some degree of plasticity. Similarly, plasticity was also present in the ancestor of the genus *Eueides,* but biosynthesis is the most common strategy, with sequestration observed in only two *Eueides isabella* butterflies out of 34 *Eueides* individuals sampled, and in none of the other four species analysed.

CG sequestration is more frequent in the ancestor of *Heliconius,* the most speciose genus of the tribe. CG sequestration was observed in almost all species within the Most Recent Common Ancestor (MRCA) Erato + Sara-Sapho, except *H. telesiphe* (out of 16 species). Indeed, the ancestor of the Sara-sapho clade favoured CG sequestration over biosynthesis and in some species biosynthesis is rarely seen (i.e. *H. sapho*, *H. hewitsoni*, and *H. antiochus*). CG sequestration was also common in the MRCA Aoede + Doris + Wallacei + Melpomene. *Heliconius aoede*, the only representative of the Aoede clade, mostly contains sequestered compounds (small amounts of biosynthesis was found in 4 individuals out of 25). Nevertheless, CG biosynthesis regains importance in the MRCA Doris + Wallacei + Melpomene, especially in the Melpomene clade, in which some species (4 out of 12 – i.e. *H. pachinus*, *H. timareta*, *H. ethila* and *H. atthis*) only biosynthesize CGs and appear to have lost biochemical plasticity. Sequestered compounds were also virtually absent in *H. doris* (N=35) and *H. wallacei* (n= 14).

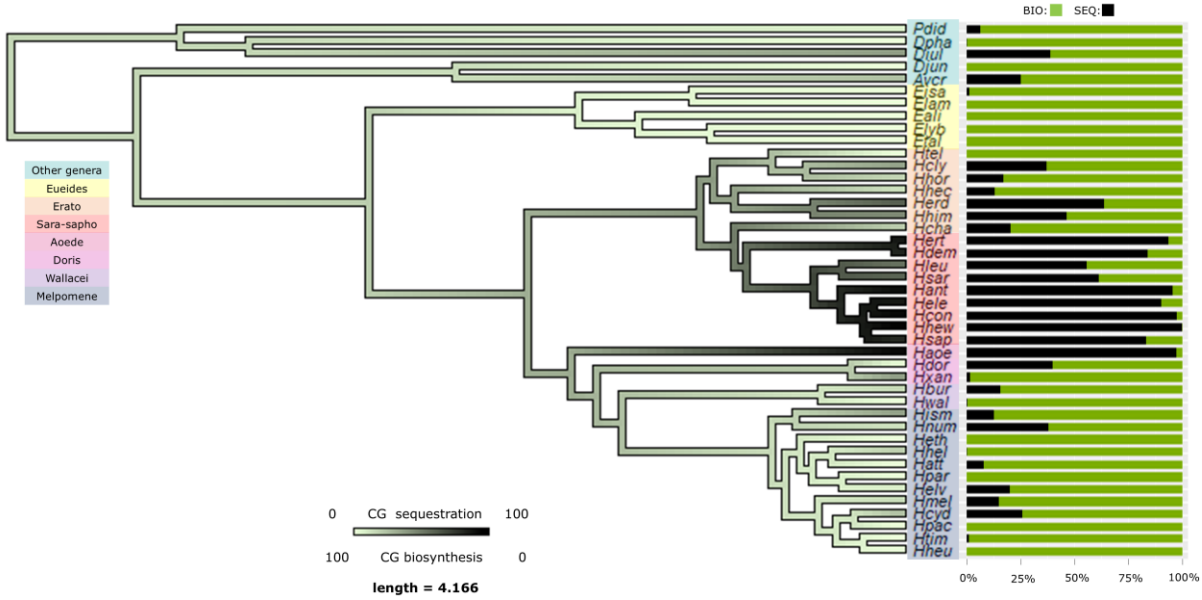


Figure 4. Ancestral reconstruction of biochemical plasticity in the Heliconiini tribe. This analysis used the tree hypothesis from Cicconard et al 2023 and a chemical dataset (743 butterflies) built based on the raw data from Castro et al 2018, Sculffold et al 2020, and Rueda-M et al. (*in prep.*). Green correponds to average percentage of biosynthesized CG and black to average percentage of sequestered CG per species. Species names are highlighted according to their clades. Legend: OTHER genera - *Pdid= Philaethria dido; Dpha= Dryadula phaetusa; Diul= Dryas iulia; Djun= Dione juno; Avcr: Agraulis vanillae.* genus *EUEIDES - Eisa= E.isabella; Elam: E. lampeto; Eali: E. aliphera; Elyb= E. lybia; Etal= E. tales.* genus *Heliconius:* ERAT0 – *Htel= H. telesiphe; Hcly= H. clysonymus; Hhor= H. hortense; Hhec= H. hecalesia; Herd= H. erato; Hhim= H. himera;* SARA-SAPHO*: Hcha= H. charithonia; Hert= H. eratosignis; Hdem= H. demeter; Hleu= H. leucadia; Hsar= H. sara; H.ant= H. antiochus; Hele= H. eleuchia; Hcon: H. congener; Hhew: H. hewitsoni; Hsap= H. sapho.* AOEDE*: Haoe: H. aoede;* DORIS: *Hdor= H. doris; Hxan= H. xanthocles; WALLACEI: Hbur= H. burney; H. wallacei;* MELPOMENE*: Hism= H. ismenius; Hnum= H. numata; Heth= H. ethila; Hhel= H. hecale; Hatt= H. atthis; Hpar= H. pardalinus; Helv= H. elevatus; Hmel=H. melpomene; Hcyd= H. cydno; Hpac= H. pachinus; Htim= H. timareta; Hheu= H. heurippa*.

CYP405s duplicated in Heliconius and Eueides, but not in other heliconiines

We used the available *Heliconius* CYP405A4 and CYP332A1 protein sequences as a starting query to identify homologues across Heliconiinae genomes, identifying 144 *CYP405* (Figure 5, Table S2) and 66 *CYP332A* loci (Figure 6, Table S2). Of the 210 recovered sequences, the majority (~80%) contain the full cytochrome p450 domain (pfam00067), while 10% and 8% have either the C-terminus or N-terminus missing, respectively, with only 2% missing both ends.

CYP405 was duplicated in the genera *Heliconius* and *Eueides*, but not in the other Heliconiini species (Figure 5). A few independent duplications and losses have occured in specific *Heliconiu*s clades/taxa, but overall, most *Heliconius* species have three or more *CYP405*s. We hypothesize that an ancestral *CYP405* loci was present in the outgroups (*CYP405A30*), other Heliconiini (*CYP405A29*), *Dione/Agraulis* (*CYP405A28*) and *Heliconius* (*CYP405A4*) (Figure 5). The ancestral locus seems to have undergone a duplication (*CYP405A5*) at the MRCA *Heliconius* + *Eueides*, followed by a loss of it in all the species of the *Erato* clade and in half of the 10 *Sara-Sapho* clade species. The ancestral locus likely duplicated a

second time at the stem of all *Heliconius* (*CYP405A6*). The synteny of *CYP405* is highly conserved with all loci located on the ancestral chromosome 15 (which corresponds to chromosome 17 in *Heliconius*).

CYP332 is the second enzyme of the CG biosynthetic pathway, and two copies of this gene were found in *Speyeria mormonia* and *Brenthis* (Figure 6).The majority of Heliconiini have a single *CYP332* locus, with the exception of *H. hortense* and *H. hermethena* (within the Erato clade), *H. aoede*, and *H. elevatus*, which have two. The locus is located on the ancestral chromosome 3 (which correspond to chromosome 11 in *Heliconius*), without any evidence of translocations and/or inversions.

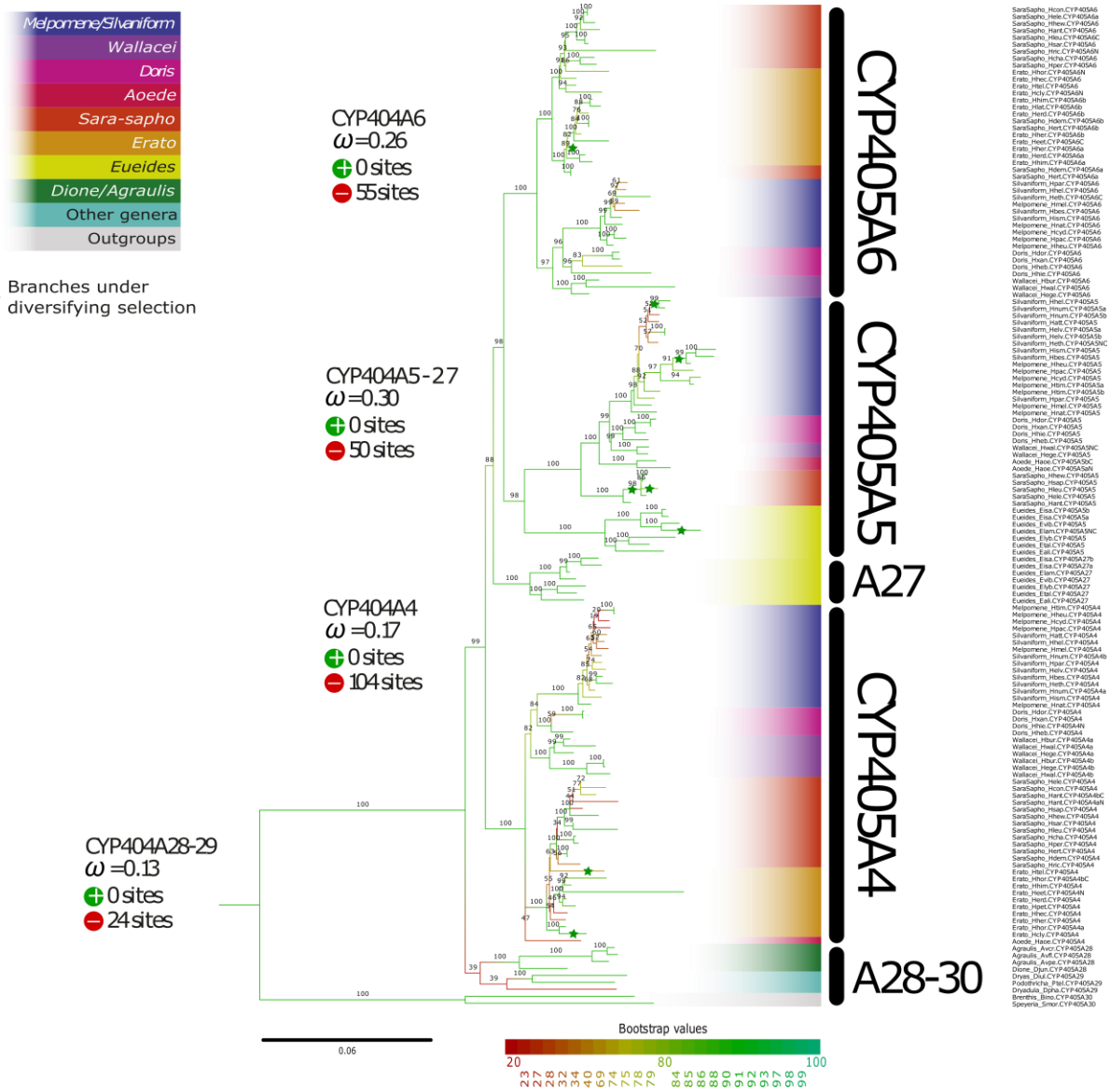


Figure 5 Gene-tree of *CYP405As* in the tribe Heliconiini, with *Speyeria mormonia* and *Brenthis* ino were used as outgroups. Species are colour coded according to their clades. Branch colours correspond to bootstrap value (Red=low, Green=high). Purifying selection (ω) on each orthologues was quantified, as well as number of sites under purifying (negative) and diversifying (positive) selection identified. Branches found under episodic diversifying positive selection in the internal branch scan analysis are marked with a green star. Legend: OTHER genera - *Pdid= Philaethria dido; Dpha= Dryadula phaetusa; Diul= Dryas iulia; Djun= Dione juno; Avcr: Agraulis vanillae.* genus *EUEIDES - Eisa= E.isabella; Elam: E. lampeto; Eali: E. aliphera; Elyb= E. lybia; Etal= E. tales. genus Heliconius:* ERATO – *Htel= H. telesiphe; Hcly= H. clysonymus; Hhor= H. hortense; Hhec= H. hecalesia; Herd= H. erato; Hhim= H. himera;* SARA-SAPHO: *Hcha= H. charithonia; Hert= H. eratosignis; Hdem= H. demeter; Hleu= H. leucadia; Hsar= H. sara; H.ant= H. antiochus; Hele= H. eleuchia; Hcon: H. congener; Hhew: H. hewitsoni; Hsap= H. sapho.* AOEDE: *Haoe: H. aoede*; DORIS: *Hdor= H. doris; Hxan= H. xanthocles;* WALLACEI: *Hbur= H. burney; H. wallacei;* MELPOMENE: *Hism= H. ismenius; Hnum= H. numata; Heth= H. ethila; Hhel= H. hecale; Hatt= H. atthis; Hpar= H. pardalinus; Helv= H. elevatus; Hmel=H. melpomene; Hcyd= H. cydno; Hpac= H. pachinus; Htim= H. timareta; Hheu= H. heurippa.*

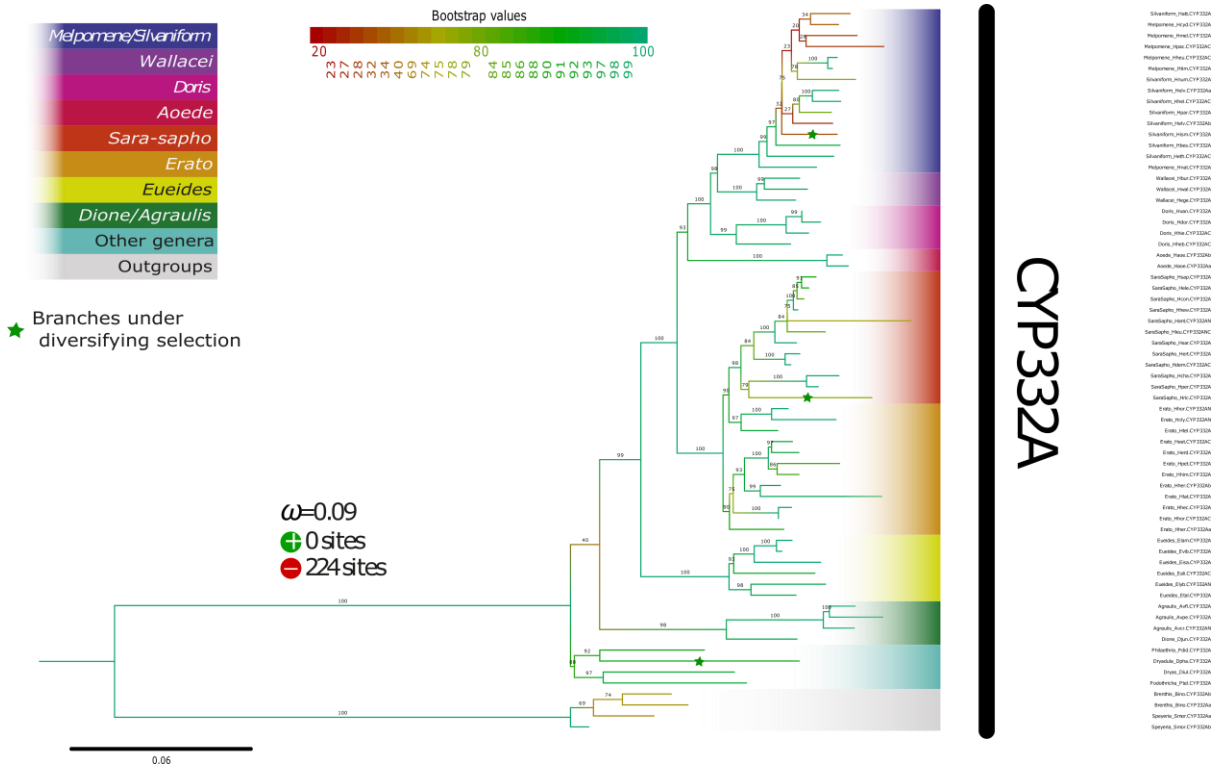


Figure 6. Phylogenetic tree of *CYP332* in the Heliconiini tribe. *Speyeria mormonia* and *Brenthis ino* were used as outgroups Species are coloured per clade. Branch colours correspond to bootstrap value (Red=low, Green=high). Purifying selection (ω), as well as number of sites under purifying (negative) and diversifying (positive) selection identified. Branches found under episodic diversifying positive selection in the internal branch scan analysis are marked with a green star. Legend: OTHER genera - *Pdid= Philaethria dido; Dpha= Dryadula phaetusa; Diul= Dryas iulia; Djun= Dione juno; Avcr: Agraulis vanillae.* genus *EUEIDES - Eisa= E.isabella; Elam: E. lampeto; Eali: E. aliphera; Elyb= E. lybia; Etal= E. tales. genus Heliconius:* ERATO – *Htel= H. telesiphe; Hcly= H. clysonymus; Hhor= H. hortense; Hhec= H. hecalesia; Herd= H. erato; Hhim= H. himera;* SARA-SAPHO: *Hcha= H. charithonia; Hert= H. eratosignis; Hdem= H. demeter; Hleu= H. leucadia; Hsar= H. sara; H.ant= H. antiochus; Hele= H. eleuchia; Hcon: H. congener; Hhew: H. hewitsoni; Hsap= H. sapho.* AOEDE: *Haoe: H. aoede*; DORIS: *Hdor= H. doris; Hxan= H. xanthocles;* WALLACEI: *Hbur= H. burney; H. wallacei;* MELPOMENE: *Hism= H. ismenius; Hnum= H. numata; Heth= H. ethila; Hhel= H. hecale; Hatt= H. atthis; Hpar= H. pardalinus; Helv= H. elevatus; Hmel=H. melpomene; Hcyd= H. cydno; Hpac= H. pachinus; Htim= H. timareta; Hheu= H. heurippa.*

Association between CYP405A-copy number, chemical phenotype and hostplant specialization

Heliconius species have up to four *CYP405* genes, but some sequences were smaller than expected (~1500 bp). Translated sequences were therefore manually checked for the presence of the six conserved motifs in insect P450s (Feyereisen, 1999): Cluster of Pro/Gly, C-Helix, I-Helix, ExLR, PERF, and Heme-binding loop. This resulted in 98 complete *CYP405s* out of 144, with most of the incomplete copies found in the *Heliconius sapho clade* (14 incomplete sequences) (Figure 7).

There is a clade-specific pattern in *CYP405*-copy number: the majority of the Silvaniform/Melpomene have three complete *CYP405s*, most Erato and Sara have two and Sapho have none. This pattern matches the chemical profile of these clades. CG sequestration in the Heliconiini tribe gained more importance in the MRCA *Heliconius* Erato + Sara + Sapho, especially for Sapho species, and in the Aoede clade. Interestingly, this clade-specific pattern is associated with their hostplant specialization: while other *Heliconius* (MRCA Silvanifom/Melpomene + Doris + Wallacei) have hostplants in the

subgenus *Passiflora,* in which most species lack CGs that can be sequestered (de Castro, Zagrobelny, et al., 2019), the Erato and Sara clades have hostplants in the subgeneus *Decaloba* and Sapho in the subgenus *Astrophea,* where CG for sequestration is common. Aoede species use the genera *Mitostemma* and *Dilkea* as hostplants from which they sequester CGs, instead of *Passiflora.* Our results suggests that there is an association between *CYP405*-copy number, biosynthesis versus sequestration plasticity, and hostplant specialization.

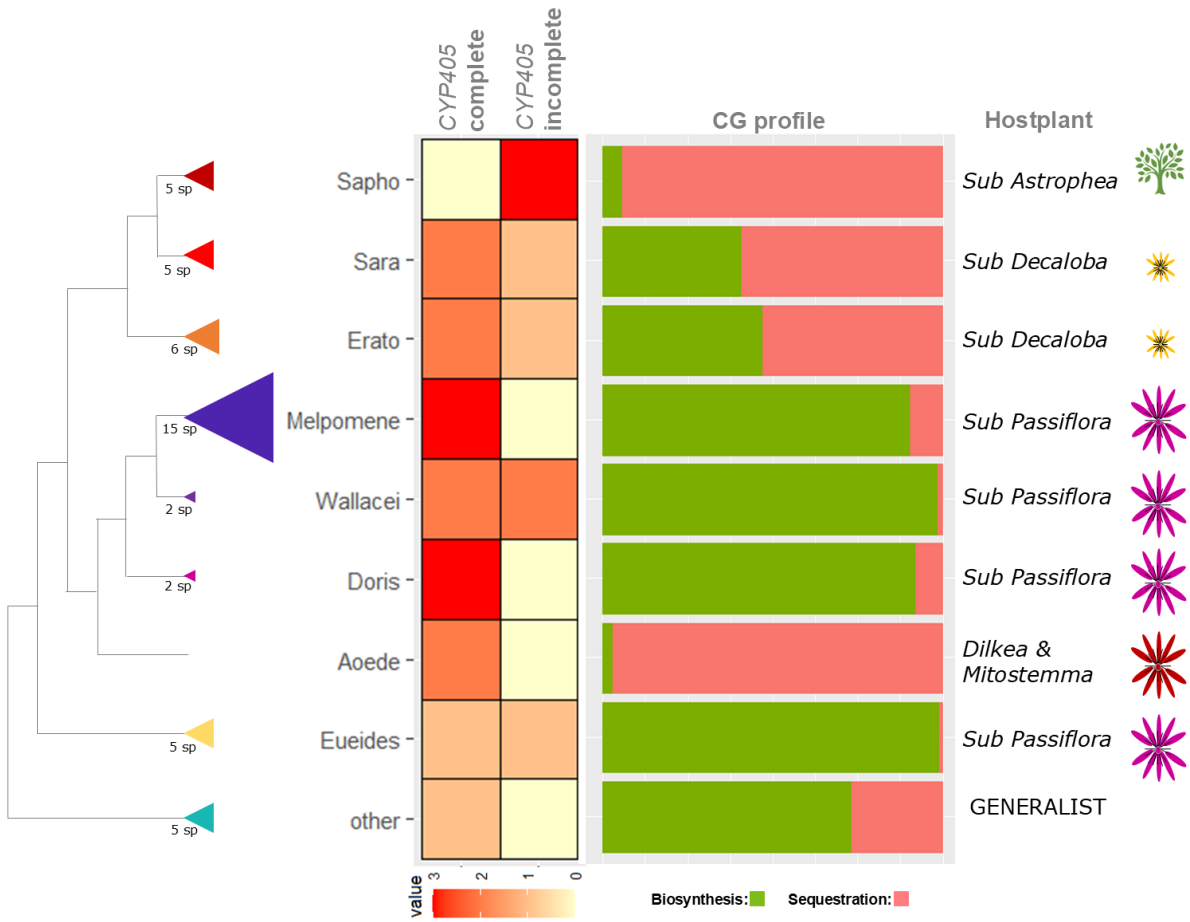


Figure 7. *CYP405*-copy number, chemical profile and hostplant specialisation of different Heliconiini clades. In this figure, the clade Sara-sapho was split into two distinct clades: Sara (*H. sara*, *H. leucadia*, *H. demeter*, *H. eratosignis* and *H. charithonia*) and Sapho (*H. sapho*, *H. antiochus, H. eleuchia, H. congener* and *H. hewitsoni* - as they have different hostplant specialization, different chemistry and different number of complete *CYP405*.

<u>P450s involved in CG biosynthesis are under purifying selection</u>

We quantified the levels of purifying selection acting across the *CYP405* orthologues and *CYP332* and found that ω varies substantially. The ancestral *CYP405A* locus, *CYP405A28-30* in the outgroups and other Heliconiini and *CYP405A4* in *Heliconius*, show purifying selection (ω = 0.13 and ω = 0.17, consistent with divergent selection following gene duplication. None of the orthologues had sites

under positive selection. *CYP332* instead shows a stronger signature of purifying selection (ω = 0.09; 224 sites under purifying selection; no sites under positive selection) than the *CYP405As*.

Although strong purifying selection was found in all *CYP405s* and *CYP332 loci*, we wanted to investigate the overall magnitude and pattern of diversifying positive selection testing all branches (internal and terminal) in the two gene trees. Diversifying positive selection in *CYP405A5* was identified in nine branches (~8% of all branches). Three internal branches of Silvaniform/Melpomene branches, three involved terminal braches of Erato species, two branches of Sara/Sapho, and one branch of Erato (future Figure 5; $\omega^+ > 128$; adjusted P-value < 0.0002). On average the signature of selection involved ~2% of the protein coding genes, with the highest proportion in the *H. telesiphe* and an internal branch of Silvaniform/Melpomene (~3.7).

The same analysis for *CYP332A* identified three branches, ~3% of all branches (adjusted P-adjusted < 0.0001). Namely, *H. ricini* (Sara/Sapho clade; $\omega^+ > 1000$; affecting 3.4% of the branch-site; adjusted P-value < 0.00001), *H. ismenius* (Silvaniform clade; $\omega^+ = 278$; affecting 1.2% of the branch-site; adjusted P-value = 0.00018), and in *Dr. phaetusa* ($\omega^+ = 46$; affecting 3.9% of the branch-site; adjusted P-value = 0.00002).

We next tested whether "sequestration-specialist" species (part of the Sara-sapho clade and *Heliconius aoede*), show evidence for reduced purifying selection on genes involved in CG biosynthesis. BUSTED-PH was used to test whether *CYP405s and CYP332s* have experienced episodic positive diversifying selection in "sequestration-specialist" in comparison to the others ("plastic species"), and the RELEX test infer the direction of the shift between these two groups. For *CYP405,* we performed these analyses separately for each orthologue and found that in all orthologues but in *CYP4056* episodic diversifying positive selection were detected in both "sequestration-specialist" and "plastic-species" *,* but no significant difference were detected. For *CYP405A6* episodic selection was not found in "sequestration-specialist", but no significant difference were detected by BUSTED-PH in comparison with "plastic-species". Regarding RELAX test, neither intensification nor relaxation were detected for each OGs. For *CYP332*, only the "plastic-species" show episodic diversifying positive selection, but no evidence of differential selection regimes nor intensification or relaxation of selection was found. (P value < 0.0001) in the RELAX test .

DISCUSSION

Phenotypic plasticity plays an important role in allowing organisms adapt to novel or extreme environments. However, plasticity is mainly studied in visible phenotypes. Our findings emphasize the importance of phenotypic plasticity for the evolution of toxicity, a complex molecular trait. Here, we first demonstrated, using a phylogenetic reconstruction, that there is a pattern of ancestral biochemical plasticity and derived loss of plasticity in Heliconiini butterflies. We also use CRISPR to characterize *CYP405s* in *Heliconius erato*, demonstrating that these genes are involved in CG biosynthesis, which enables biochemical plasticity. Finally, we have shown that gene-loss is associated with the loss of plasticity in a CG sequestration-specialist clade (Sapho), providing a novel example of plasticity-first evolution.

Passion vine butterflies and burnet-moths are the only lepidopterans known to biosynthesize CGs (Zagrobelny et al., 2018). Yet, most butterflies and some moths have a copy of *CYP405*, suggesting that this gene was independently co-opted as the first-step of CG *de novo* biosynthesis at least twice and likely has another function in other Lepidoptera (Zagrobelny, Jensen, et al., 2018). The high mortality and poor development of the *CYP405*-knockout larvae in our experiments might be associated with this other function and/or the importance of CG biosynthesis for nitrogen metabolism. Similar to lepidopterans, several distantly-related cyanogenic plants have co-opted *CYP79s* into the first step of CG biosynthesis (Sánchez-Pérez & Neilson, 2024), while ferns have recruited a FMO (Thodberg et al., 2020). Indeed, although plants and lepidopterans biosynthesize CGs using the same enzymatic steps, the genes involved are not homologous and have evolved independently. In higher plants, different P450s have been recruited into the second step of CG biosynthesis (Takos et al., 2011; Lai et al., 2020). In lepidopterans, this second step is catalysed by CYP332, a gene characterized in *Zygaena* moths and found across all Heliconiini butterflies. While the ancestral function of *CYP405* in lepidopterans remains unknown, *CYP332* were likely involved in detoxification of plant toxins. Most lepidopterans have one *CYP332,* but interestingly, *Helicoverpa amigera* has two copies: while *CYP332A1* has undergone neofunctionalization and is associated with resistance to fenvalerate, the function of *CYP332A2* is still unknown (Huang et al., 2022).Biochemical constraints likely underlie the extensive co-option and repeated evolution in CG biosynthesis, highlighting the critical role of P450 gene turnover in driving diversification (Sánchez-Pérez & Neilson, 2024).

By reconstructing the ancestral state of biochemical plasticity in the tribe Heliconiini, we demonstrated that plasticity was ancestral in the tribe. Yet, some specific clades might have undergone specialization either in CG biosynthesis (genus *Eueides* and the clade Doris) or sequestration (Erato clade), and even virtually lost biochemical plasticity (Aoede and Sapho clades). This derived accommodation of plasticity suggests that the dual strategy system used by Heliconiini

to acquire their toxins represents a novel example of plasticity-first evolution - in which ancestral plasticity allows colonization of new environments/ ecological niches, but the alternative phenotype can be tuned (genetic accommodation) or even get fixed (genetic assimilation/canalization) as lineages specializes in the novel conditions. Although the ability to biosynthesize CGs allowed Heliconiini butterflies to maintain their toxicity regardless of the CG composition of their larval hostplants, widening their diet-breadth, as lineages evolved preference for a specific hostplant, selection favoured specialization towards one biochemical strategy over the other. The loss of plasticity in the Sapho clade, is associated with the disruption of their *CYP405*, which lacks regions that code for some of the conserved motifs that are crucial for the catalytic activity of a cytochrome P450. Species in this clade tended to have several incomplete copies of this gene and only sequester CGs. We also identified episodic diversifying selection on CYP405 and CYP332 in the Sapho clade. When a plastic phenotype is exposed to selection more often, there is greater opportunity for adaptive refinement (Lafuente & Beldade, 2019). Supporting this, the Sapho clade tend to be monophagous and have preference for a small group of *Passiflor*a in which CG for sequestration is common (Engler-Chaouat & Gilbert, 2007), which could have favoured the loss of biochemical plasticity in this lineage.

To date, examples of plasticity-first evolution in which the associated genetic changes are understood are scarce. For example, snowshoe hares (*Lepus americanus*) in North American often have white colouration during winter and molt to brown in spring (Jones et al., 2018). This plasticity is ancestral in the species, allowing the hares to have an optimal camouflage at difference seasons. Nevertheless, some populations in warmer areas, such as British Columbia, lost this plasticity displaying a winter-brown phenotype. This loss of plasticity via genetic assimilation/canalization was caused by the introgression of an *Agouti* allele from black-tailed jackrabbits to this population of snowshoe hare (Jones et al., 2018). Similarly, side-blotched lizards (*Uta stansburiana*) living on the Pisgah Vulcan are cryptic and have a plastic colouration (Corl et al., 2018). The Pisgah phenotype is melanic and inhabits the lava flow, whereas the off-lava phenotype is lighter and cryptic in nearby light-coloured soils. Although both lizard populations can produce the on- and off-lava phenotypes, genetic accommodation has led them to specialize on the phenotype cryptic in their local environment. It has been shown that the Pisgah population can develop darker coloration than the off-lava populations on dark substrate, because they have specific variants of the genes *PREP* and *PRKAR1A* which control melanin formation (Corl et al., 2018). These are two clear examples of the importance of ancestral plasticity in morphological traits for the colonization of new environments, as well as the role of refinement in specific lineages towards local specialisation for diversification. Our results on the biochemical plasticity in Heliconiini butterflies provide another example of plasticity-first evolution in which the genetic basis for the trait is well understood.

Phenotypic plasticity requires a molecular switch mechanism for sensing the environment and controlling the expression of alternative phenotypes (Sommer, 2020). While biochemical plasticity is controlled by the presence of CG for sequestration in the larval hostplant, how butterflies perceive these compounds in their diet and modulate the level of CG biosynthesis in response to that remains unknown. The vast majority of *Heliconius* clades evolved to have at least two complete *CYP405As.* We initially hypothesized that these were redundant orthologues, but our CRISPR experiments suggest that both *CYP405A4* and *CYP405A6* are essential for CG biosynthesis in *H. erato*. We predict that the multiple *CYP405* copies might be necessary to fine-tune this pathway according to diet and modulate biochemical plasticity. In fact, the metabolism of toxins is often composed of enzymes that though very similar, are not redundant and are involved in modulating responses to the environment. For example, CYP83A1 and CYP83B1 in *Arabdopsis thaliana* are very similar and catalyse the same step in the biosynthesis of glucosinolates (Naur et al., 2003). Despite their similarity, CYP83A1 is more efficient in using aliphatic oximes, whereas CYP83B1 is better for aromatic oximes, and the two orthologues enable *A. thaliana* to precisely modulate their glucosinolate profile in response to stress and developmental changes (Naur et al., 2003). To detoxify glucosinolates from their Brassicaceae hostplants, *Pieris* caterpillars utilize the enzymes NSP (Nitrile-Specifier Protein) and MA (Major Allergen) that convert these compounds into inert nitriles (Okamura et al., 2022). Although NSP and MA derived from the same ancestor gene, NSP expression is upregulated by high concentrations of aliphatic glucosinolates, in contrast with MA that is upregulated by aromatics (Okamura et al., 2019).

In summary, we have provided the first functional evidence for the genetic basis of CG biosynthesis in heliconiine butterflies and performed ancestral reconstruction analyses on their biochemical plasticity. Our results bring empirical evidence for an example of plasticity-first evolution, with plasticity ancestral in the tribe and virtually lost in one clade specialized in sequestration. This loss was associated with genes involved in CG biosynthesis losing key functional domains. It also corroborates the hypothesis that biochemical constrains could be the cause of high incidence of parallel evolution in mechanisms to detoxify and synthesize toxins.

ACKNOWLEDGEMENTS

We thank the Zoology Department for hosting this project, specially Glennis Julian who maintains the butterfly and plant stocks used in the CRISPR experiment. We also thanks Mika Zagrobelny and Rene Feyereisen for their amazing work that set the foundation for this project, as well as discussions over the year about the evolution of cyanogenesis in Lepidoptera. We thank the staff at the PLBIO-PLEN-KU, for the assistance in the target-metabolomic analyses, especially Mariela Alejandra González Ramírez for the maintenance of the analytical instruments and Mohammed Saddik Motawie for the chemical standards used in these analyses. We thank Davi Nelson for naming the P450s identified in this project. EC is thankful for the H2020 Marie Skłodowska-Curie Actions (841230) which funded this project. EC, CJ and IW thank the financial support from UKRI – NERC (NE/W005131/1). FC and SHM thank the financial support from UKRI - NERC (NE/N014936/1) and ERC (758508).

REFERENCES

Beran, F., Köllner, T. G., Gershenzon, J., & Tholl, D. (2019). Chemical convergence between plants and insects: biosynthetic origins and functions of common secondary metabolites. *New Phytologist*, *223*(1), 52–67. https://doi.org/10.1111/NPH.15718

Bui Quang Minh, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D Woodhams, Arndt von Haeseler, Robert Lanfear (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era, *Molecular Biology and Evolution*, 37. https://doi.org/10.1093/molbev/msaa015

Camacho, C., Madden, T., Ma, N., Tao, T., Agarwala, R., & Morgulis, A. (2013). *BLAST command line applications user manual*.

Chauhan, R., Jones, R., Wilkinson, P., Pauchet, Y., & Ffrench-Constant, R. H. (2013). Cytochrome P450-encoding genes from the Heliconius genome as candidates for cyanogenesis. *Insect Molecular Biology*, *22*(5), 532–540. https://doi.org/10.1111/imb.12042

Corl, A., Bi, K., Luke, C., Challa, A. S., Stern, A. J., Sinervo, B., & Nielsen, R. (2018). The Genetic Basis of Adaptation following Plastic Changes in Coloration in a Novel Environment. *Current Biology*, *28*(18), 2970-2977.e7. https://doi.org/10.1016/J.CUB.2018.06.075

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/GIGASCIENCE/GIAB008

de Castro, É. C., Zagrobelny, M., Zurano, J. P., Zikan Cardoso, M., Feyereisen, R., & Bak, S. (2019). Sequestration and biosynthesis of cyanogenic glucosides in passion vine butterflies and consequences for the diversification of their host plants. *Ecology and Evolution*, *9*(9), 5079-5093.

de Castro, É. C., Musgrove, J., Bak, S., McMillan, W. O., & Jiggins, C. D. (2021). Phenotypic plasticity in chemical defence of butterflies allows usage of diverse host plants. *Biology Letters*, *17*(3), 20200863.

de Castro, É. C. P., Zagrobelny, M., Cardoso, M. Z., & Bak, S. (2018). The arms race between heliconiine butterflies and Passiflora plants – new insights on an ancient subject. *Biological Reviews*, *93*(1). https://doi.org/10.1111/brv.12357

de Castro, É. C. P., Zagrobelny, M., Zurano, J. P., Cardoso, M. Z., Feyereisen, R., & Bak, S. (2019). Sequestration and biosynthesis of cyanogenic glucosides in passion vine butterflies and consequences for the diversification of their host plants. *Ecology and Evolution*, *9*(9), 5079–5093. https://doi.org/10.1002/ece3.5062

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and Plants - a Study in Coevolution. *Evolution*, *18*(4), 586–608. https://doi.org/10.2307/2406212

Engler-Chaouat, H. S., & Gilbert, L. E. (2007a). De novo synthesis vs. sequestration: Negatively correlated metabolic traits and the evolution of host plant specialization in cyanogenic butterflies. *Journal of Chemical Ecology*, *33*(1), 25–42. https://doi.org/10.1007/s10886-006-9207-8

Engler-Chaouat, H. S., & Gilbert, L. E. (2007b). De novo synthesis vs. sequestration: Negatively correlated metabolic traits and the evolution of host plant specialization in cyanogenic butterflies. *Journal of Chemical Ecology*, *33*(1), 25–42. https://doi.org/10.1007/S10886-006-9207-8/FIGURES/4

Feyereisen, R. (1999). Insect P450 enzymes. *Annual Review of Entomology*, *44*, 507–533. https://doi.org/10.1146/ANNUREV.ENTO.44.1.507

Francesco Cicconardi, Edoardo Milanetti, Érika C. Pinheiro de Castro, Anyi Mazo-Vargas, Steven M. Van Belleghem, Angelo Alberto Ruggieri, Pasi Rastas, Joseph Hanly, Elizabeth Evans, Chris D Jiggins, W Owen McMillan, Riccardo Papa, Daniele di Marino, Arnaud Martin, & Stephen H Montgomery. (2022). Evolutionary dynamics of genome size and content during the adaptive radiation of Heliconiini butterflies. *BioRxiv*. https://doi.org/10.1101/2022.08.12.503723

Fürstenberg-Hägg, J., Zagrobelny, M., Jørgensen, K., Vogel, H., Møller, B. L., & Bak, S. (2014). Chemical defense balanced by sequestration and de novo biosynthesis in a lepidopteran specialist. *PLoS ONE*, *9*(10). https://doi.org/10.1371/JOURNAL.PONE.0108745

Huang, Y., Qin, J., Wu, P., Zheng, J., & Qiu, L. (2022). Comparison of cytochrome P450 CYP332A1 gene in resistant and susceptible strains of the cotton bollworm, Helicoverpa armigera (Lepidoptera: Noctuidae). *Applied Entomology and Zoology*, *57*(2), 171–181. https://doi.org/10.1007/s13355-022-00777-7

Jensen, N. B., Zagrobelny, M., Hjernø, K., Olsen, C. E., Houghton-Larsen, J., Borch, J., Møller, B. L., & Bak, S. (2011). Convergent evolution in biosynthesis of cyanogenic defence compounds in plants and insects. *Nature Communications*, *2*, 273. https://doi.org/10.1038/ncomms1271

JONES, D. A. (1962). Selective Eating of the Acyanogenic Form of the Plant Lotus corniculatus L. by Various Animals. *Nature*, *193*(4820), 1109–1110. https://doi.org/10.1038/1931109a0

Jones, M. R., Scott Mills, L., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J. R., Jiggins, F. M., Jensen, J. D., Melo-Ferreira, J., & Good, J. M. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science*, *360*(6395), 1355–1358. https://doi.org/10.1126/SCIENCE.AAR5273/SUPPL_FILE/AAR5273_JONES_SM.PDF

Kariyanna, B., Prabhuraj, A., Asokan, R., Ramkumar, G., Venkatesan, T., Gracy, R. G., & Mohan, M. (2021). Genome mining and functional analysis of cytochrome P450 genes involved in insecticide resistance in Leucinodes orbonalis (Lepidoptera: Crambidae). *Biotechnology and Applied Biochemistry*, *68*(5), 971–982. https://doi.org/10.1002/BAB.1997

Kawahara, A. Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E. F. A., Donath, A., Gimnich, F., Frandsen, P. B., Zwick, A., dos Reis, M., Barber, J. R., Peters, R. S., Liu, S., Zhou, X., Mayer, C., Podsiadlowski, L., Storer, C., Yack, J. E., Misof, B., & Breinholt, J. W. (2019). Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(45), 22657–22663. https://doi.org/10.1073/PNAS.1907847116/-/DCSUPPLEMENTAL

Kosakovsky Pond, S. L., & Frost, S. D. W. (2005). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, *22*(5), 1208–1222. https://doi.org/10.1093/molbev/msi105

Kosakovsky Pond, S. L., Murrell, B., Fourment, M., Frost, S. D. W., Delport, W., & Scheffler, K. (2011). A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution*, *28*(11), 3033–3043. https://doi.org/10.1093/molbev/msr125

Kosakovsky Pond, S. L., Poon, A. F. Y., Velazquez, R., Weaver, S., Hepler, N. L., Murrell, B., Shank, S. D., Magalis, B. R., Bouvier, D., Nekrutenko, A., Wisotsky, S., Spielman, S. J., Frost, S. D. W., & Muse, S. V. (2020). HyPhy 2.5 - A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Molecular Biology and Evolution*, *37*(1), 295–299. https://doi.org/10.1093/molbev/msz197

Kučka, M., & Chan, Y. F. (2022). *HMW DNA extraction using magnetic beads*. https://doi.org/dx.doi.org/10.17504/protocols.io.b46bqzan

Lafuente, E., & Beldade, P. (2019). Genomics of developmental plasticity in animals. *Frontiers in Genetics*, *10*(JUL), 456044. https://doi.org/10.3389/FGENE.2019.00720/PDF

Lai, D., Maimann, A. B., Macea, E., Ocampo, C. H., Cardona, G., Pičmanová, M., Darbani, B., Olsen, C. E., Debouck, D., Raatz, B., Møller, B. L., & Rook, F. (2020). Biosynthesis of cyanogenic glucosides in Phaseolus lunatus and the evolution of oxime-based defenses. *Plant Direct*, *4*(8), e00244. https://doi.org/10.1002/PLD3.244

Lanfear, R., Schalamun, M., Kainer, D., Wang, W., & Schwessinger, B. (2019). MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics*, *35*(3), 523. https://doi.org/10.1093/BIOINFORMATICS/BTY654

Levis, N. A., & Pfennig, D. W. (2019). Plasticity-led evolution: Evaluating the key prediction of frequency-dependent adaptation. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1897). https://doi.org/10.1098/RSPB.2018.2754

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, *48*(D1), D265–D268. https://doi.org/10.1093/nar/gkz991

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, a., Eren, K., Pollner, T., Martin, D. P., Smith, D. M., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Gene-Wide Identification of Episodic Selection. *Molecular Biology and Evolution*, *32*(5), 1365–1371. https://doi.org/10.1093/molbev/msv035

Naur, P., Petersen, B. L., Mikkelsen, M. D., Bak, S., Rasmussen, H., Olsen, C. E., & Halkier, B. A. (2003). CYP83A1 and CYP83B1, Two Nonredundant Cytochrome P450 Enzymes Metabolizing Oximes in the Biosynthesis of Glucosinolates in Arabidopsis. *Plant Physiology*, *133*(1), 63–72. https://doi.org/10.1104/pp.102.019240

Nishida, R. (2014). Chemical ecology of insect–plant interactions: ecological significance of plant secondary metabolites. *Bioscience, Biotechnology, and Biochemistry*, *78*(1), 1–13. https://doi.org/10.1080/09168451.2014.877836

Okamura, Y., Dort, H., Reichelt, M., Tunström, K., Wheat, C. W., & Vogel, H. (2022). Testing hypotheses of a coevolutionary key innovation reveals a complex suite of traits involved in

defusing the mustard oil bomb. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(51), e2208447119. https://doi.org/10.1073/PNAS.2208447119/SUPPL_FILE/PNAS.2208447119.SD01.XLSX

Okamura, Y., Sato, A., Tsuzuki, N., Sawada, Y., Hirai, M. Y., Heidel-Fischer, H., Reichelt, M., Murakami, M., & Vogel, H. (2019). Differential regulation of host plant adaptive genes in Pieris butterflies exposed to a range of glucosinolate profiles in their host plants. *Scientific Reports 2019 9:1*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-43703-8

Petschenka, G., & Agrawal, A. A. (2016). How herbivores coopt plant defenses: Natural selection, specialization, and sequestration. *Current Opinion in Insect Science*, *14*, 17–24. https://doi.org/10.1016/j.cois.2015.12.004

Piel, J., Höfer, I., & Hui, D. (2004). Evidence for A Symbiosis Island Involved in Horizontal Acquisition of Pederin Biosynthetic Capabilities by the Bacterial Symbiont of Paederus fuscipes Beetles. *Journal of Bacteriology*, *186*(5), 1280–1286. https://doi.org/10.1128/JB.186.5.1280-1286.2004/FORMAT/EPUB

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N., Delsuc, F., & Wilke, C. (2018). MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular Biology and Evolution*, 2–4. https://doi.org/10.1093/molbev/msy159

Revell, L. J. (2024). phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*, *12*, e16505. https://doi.org/10.7717/PEERJ.16505

Rueda-M, N., de Castro, E C. P., Pardo-Diaz, C. de Martino, F., Jiggins, C. D., Arias, M. and Salazar, C. (*in prep.)* Chemical defense variation in *Heliconius* butterflies: testing the role of mimicry rings and ecoregions.

Schappert, P. J., & Shore, J. S. (1999). Effects of cyanogenesis polymorphism in Turnera ulmifolia on Euptoieta hegesia and potential Anolis predators. *Journal of Chemical Ecology*, *25*(6), 1455–1479. https://doi.org/Doi 10.1023/A:1020995329980

Sculfort, O., Castro, E. C. P., Kozak, K. M., Bak, S., Elias, M., Nay, B., & Llaurens, V. (2020). Variation of chemical compounds in wild Heliconiini reveals ecological factors involved in the evolution of chemical defenses in mimetic butterflies. *Ecology and Evolution*, *10*(5), 2677–2694. https://doi.org/10.1002/ece3.6044

Sculfort, O., de Castro, E. C. P., Kozak, K. M., Bak, S., Elias, M., Nay, B., & Llaurens, V. (2020). Variation of chemical compounds in wild Heliconiini reveals ecological factors involved in the evolution of chemical defenses in mimetic butterflies. *Ecology and Evolution*, *10*(5). https://doi.org/10.1002/ece3.6044

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 1–11. https://doi.org/10.1186/1471-2105-6-31

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, *32*(5), 1342–1353. https://doi.org/10.1093/molbev/msv022

Sommer, R. J. (2020). Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges. *Genetics*, *215*(1), 1. https://doi.org/10.1534/GENETICS.120.303163

Steiner, A.-M., Busching, C., Vogel, H., & Wittstock, U. (2018). Molecular identification and characterization of rhodaneses from the insect herbivore Pieris rapae. *Scientific Reports*, *8*(1), 10819. https://doi.org/10.1038/s41598-018-29148-5

Takos, A. M., Knudsen, C., Lai, D., Kannangara, R., Mikkelsen, L., Motawia, M. S., Olsen, C. E., Sato, S., Tabata, S., Jørgensen, K., Møller, B. L., & Rook, F. (2011). Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway. *The Plant Journal*, *68*(2), 273–286. https://doi.org/10.1111/J.1365-313X.2011.04685.X

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., & Scheffler, K. (2014). RELAX: Detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution*, *32*(Templeton 2008), 1–13. https://doi.org/10.1093/molbev/msu400

Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, *20*(1), 1–10. https://doi.org/10.1186/S13059-019-1727-Y/FIGURES/4

Wu H, Li Z, Zhong Z, Guo Y, He L, Xu X, Mao Y, Tang D, Zhang W, Jin F, Pang R (2025). Insect Cytochrome P450 Database: An Integrated Resource of Genetic Diversity, Evolution and Function. Mol Ecol Resour. 2025;25(4):e14070. doi: 10.1111/1755-0998.14070.

Wybouw, N., Dermauw, W., Tirry, L., Stevens, C., Grbić, M., Feyereisen, R., & Van Leeuwen, T. (2014). A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. *ELife*, *3*, e02365. https://doi.org/10.7554/eLife.02365

Wybouw, N., Pauchet, Y., Heckel, D. G., & Leeuwen, T. Van. (2016). Horizontal Gene Transfer Contributes to the Evolution of Arthropod Herbivory. *Genome Biology and Evolution*, *8*(6), 1785–1801. https://doi.org/10.1093/GBE/EVW119

Zagrobelny, M., de Castro, É. C. P., Møller, B. L., & Bak, S. (2018). Cyanogenesis in arthropods: From chemical warfare to nuptial gifts. In *Insects* (Vol. 9, Issue 2). MDPI AG. https://doi.org/10.3390/insects9020051

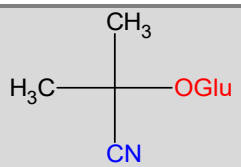
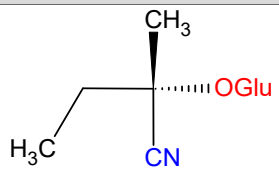
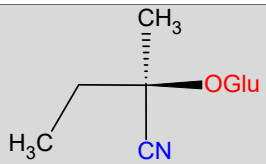
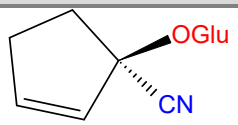
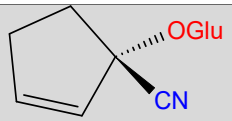
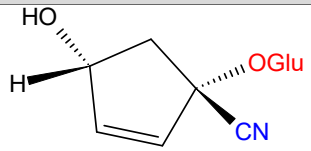
Supplementary Methods

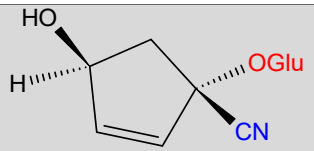
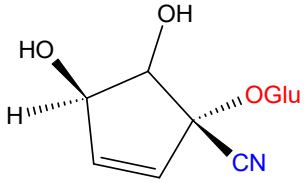
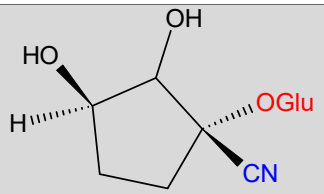
Butterfly husbandry

For the CRISPR experiments, fresh eggs were supplied from a insectary population of *Heliconius erato demophoon* keep at the Zoology Department. Butterflies are kept in the green house at 28 °C degrees, 80% Relative humidity and 12h/12h light cycle and fed *ad libitum* with artificial nectar as well as flower of *Lantana camara* and *Psiguria sp*. Plants of *Passiflora biflora* with fresh tips are provided for egg laying and eggs are collected to be raised in cages in a CT room at 28 °C degrees, 80% Relative humidity and 12h/12h light cycle. Larvae are fed *ad libitum* with fresh shoots of *P. biflora* and pupae are transferred to a pupal cage (40 x 40 x 60 cm) for eclosion. Freshly emerged butterflies are transferred to the green house for breeding.

Classification of CGs

Table S1. Cyanogenic glucosides (CGs) de novo biosynthesized and sequestered by heliconiini identified in de Castro et al. (2019), Sculfort et al. (2020) and Rueda-M et al. (*in prep.*).

Source (Type)	Compound name	Structure
Biosynthesized (aliphatic)	Linamarin	
	Lotaustralin	
	Epilotastralin	
Sequestered from Passifloraceae (cyclopentenic)	Deidaclin	
	Tetraphyllin A	
	Tetraphyllin B	

	Epivolkenin	
	Gynocardin	
	Dihydrogynocardin	

Statistical Analyses

The larval weight of all injected individuals did not have a normal distribution (Shapiro-Wilko, W = 0.91115, p-value = 1.956e-08). Thus, Kruskal-Wallis was used to investigate if differences in larval weight were significant, followed by Dunn' Test for pairwise comparison. The CG concentration of larvae that were successfully phenotyped and genotyped also did not have a normal distribution. Kruskal-Wallis and Dunn' Test were used to evaluate if the differences in CG concentration between control, non-edited (-) and edited larvae (+) were significant.

SUPPLEMENTARY FIGURES

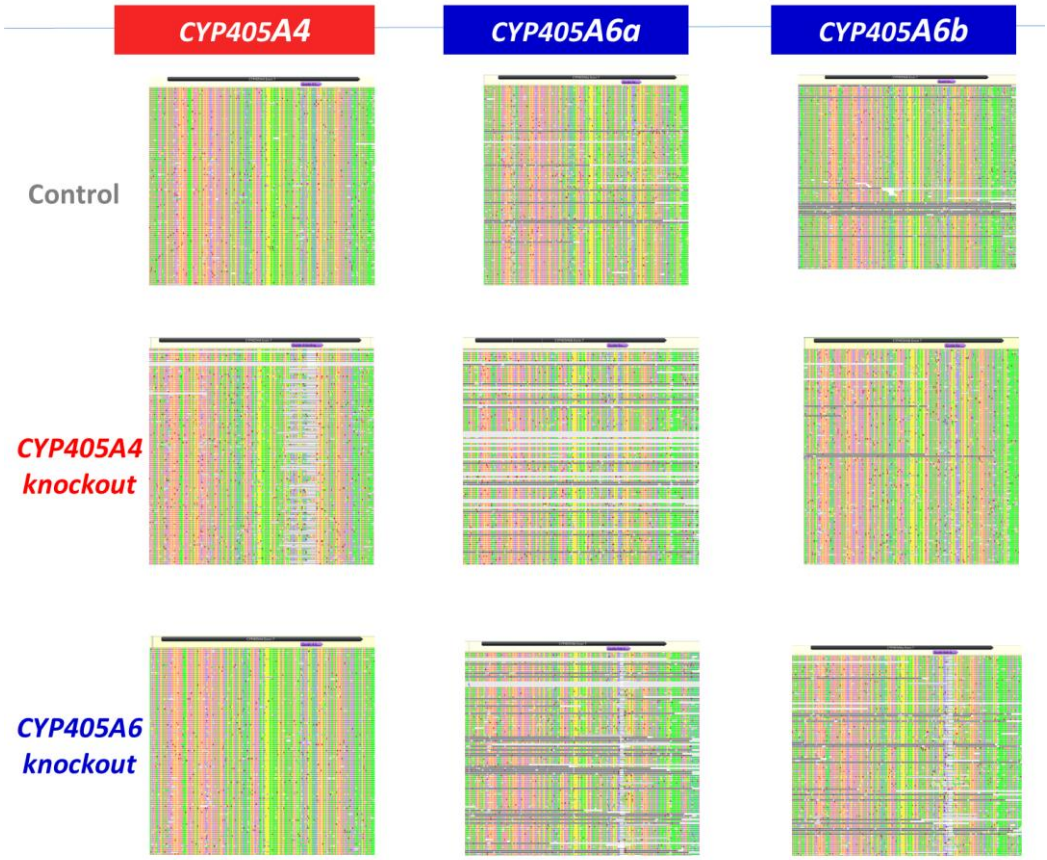


Figure S1. Each row correspond to a single genotyped individual and each column to a *CYP405* orthologue. Alignments show Nanopore sequencing results of exon 7 for each orthologues (black box) with the target-CRISPR (purple box).