PNAS RESEARCH ARTICLE | PLANT BIOLOGY CHEMISTRY

Structural innovation in the evolution of plant chemical defense

Paola Rubiano-Buitrago[a], Amy P. Hastings[a], Masaaki Uematsu[b], Jeremy M. Baskin[b], Anurag A. Agrawal[a,c,1], and Christophe Duplais[d,1]

Affiliations are included on p. 7.

Contributed by Anurag A. Agrawal; received December 10, 2025; accepted February 5, 2026; reviewed by Lee Dyer and Martin Volf

Chemical defenses are fundamental in organismal biology and widely used in medicine and agriculture. Plant defense chemistry evolves in response to various selective pressures, particularly herbivory, and theory has emphasized predicting toxin abundance and diversity. Here we test hypotheses about the evolution of structural innovation in chemical defense by combining molecular complexity metrics, metabolomics, molecular docking, and phylogenetic analyses, using milkweed cardenolides, steroidal glycosides that inhibit animal Na^+/K^+-ATPases. We identify the addition of a nitrogen–sulfur (N,S) heterocycle in highly substituted cardenolides as a major structural innovation that restores toxicity against coevolved natural enemies, such as the monarch butterfly. This toxicity is likely achieved by rigidifying the cardenolide scaffold and creating additional nonelectrostatic interactions within the Na^+/K^+-ATPase binding pocket, thereby enhancing binding affinity despite target-site resistance. Two biosynthetically distinct N,S-cardenolides, uscharin and labriformin, rank among the most complex structures in this chemical class and show divergent macroevolutionary histories: uscharin represents an ancestral character state with repeated losses, whereas labriformin has independently evolved multiple times in later-diverging lineages. This pattern across Asclepiadoideae indicates that the structural innovation evolved repeatedly, apparently limited by lineage-specific biosynthetic constraints among precursor pathways. N,S-cardenolides occur in over 75% of the 59 *Asclepias* species examined here, and species producing N,S-cardenolides exhibit greater cardenolide abundance, richness, metabolomic space, and toxicity against adapted organisms. More generally, structural innovation defines a distinct evolutionary axis in plant chemistry, enabling defense diversification and adaptive recovery of toxicity. Such innovations are predicted to build on existing molecular scaffolds in response to ecological challenges, here driven by coevolving specialist herbivores.

coevolution | plant defense theory | cardenolide toxins | chemical ecology | secondary compounds

Significance

This work provides the strongest empirical test to date of how structural innovation in plant metabolism creates an evolutionary axis of novel defensive bioactivity and ecological opportunity. In this case, biosynthetic modifications to cardenolide toxins introduce new functional groups, adding a nitrogen and sulfur (N,S) heterocycle, thereby modifying the existing chemical scaffolds. We show that these plant toxins evolved through repeated gains of novel structural attributes, challenging models of directional evolutionary escalation. In milkweeds, the independent evolution of N,S-cardenolides illustrates how structural innovation can restore potency against adapted herbivores such as monarch butterflies. The evolution of structural innovation in chemical toxins thus offers a general mechanism by which defenses diversify, gaining or regaining potency, under natural selection by herbivores.

The evolution of plant defensive chemistry is driven by a dynamic interplay between natural selection, ecological function, and biosynthetic innovation. Traditional metrics, such as phytochemical richness, abundance, or the number of chemical classes produced (1–3), capture only part of defensive chemical diversity. Structural analysis should be considered alongside these descriptive efforts, especially those that capture biosynthetic modifications resulting in innovations that have potential to create novel bioactivity (4–6). In evolutionary biology, such transformative shifts are recognized as key innovations, novel traits that open new adaptive niches and drive diversification across lineages (7). We argue that structural innovations in plant metabolism play an analogous role, representing an axis of chemical evolution (8, 9) that not only complements richness and abundance of plant defenses, but also represents a general mechanism of metabolite evolution, generating novel ecological function (1, 2, 10, 11).

Structural innovation refers to biosynthetic modifications, whether a single step or sequential, before or after scaffold elaboration, that introduces new functional groups creating a new chemical class. For example, during the radiation of the Brassicales, the evolution of glucosinolates emerged from multiple stepwise biosynthetic modifications (6), whereas steroidal alkaloids in the Solanaceae originated from the transamination of terpenoids, which generated novel and potent bioactivity (12). Such innovations are thought to be driven by strong coevolutionary interactions. Once these innovations appear, their widespread presence may be further modified or constrained by biosynthetic costs, especially in later diverging species (13), and existing models of defense evolution emphasize such trade-offs (14–18).

Chemical ecology theory further suggests that phytochemical diversity and structural novelty may arise from selective pressures outside pairwise coevolution or escalating arms races, including pathway constraints and enhanced defense against a breadth of enemies (e.g., interaction diversity) (2, 4, 11). Our framework builds on these perspectives by treating structural innovation as a distinct axis of chemical defense evolution. This study integrates

Author contributions: P.R.-B., A.P.H., M.U., A.A.A., and C.D. designed research; P.R.-B., A.P.H., M.U., A.A.A., and C.D. performed research; A.A.A. and C.D. contributed new reagents/analytic tools; P.R.-B., M.U., A.A.A., and C.D. analyzed data; and P.R.-B., A.P.H., M.U., J.M.B., A.A.A., and C.D. wrote the paper.

Reviewers: L.D., University of Nevada, Reno; and M.V., Biologicke centrum Akademie ved Ceske republiky.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: aa337@cornell.edu or c.duplais@cornell.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2535853123/-/DCSupplemental.

Published March 12, 2026.

diverse approaches and incorporates structural metrics capturing both incremental biosynthetic changes (e.g., successive oxidations on the same core structure) and structural innovations that introduce entirely novel moieties within the chemical class, thereby increasing ecological potential. This multidimensional insight into the evolution of chemical defense allows us to test classic hypotheses while also generating new predictions.

Milkweeds in the genus *Asclepias* provide an ideal system for testing plant defense theory because they produce cardenolides, steroidal toxins that inhibit the universal animal transmembrane protein Na^+/K^+ ATPase, with a well-documented mechanistic, ecological and evolutionary basis (19–24). While the molecular complexity of cardenolides vary, a subset of structures possesses a rare nitrogen- and sulfur-containing (N,S) heterocyclic moiety likely derived from cysteine (25–29). N,S-cardenolides were previously found in <8% of the 140 *Asclepias* spp. and have not been reported outside the Asclepiadoideae. Remarkably, cardenolides with a heterocycle, compared to those without, show no greater toxicity to sensitive animals, but they are the most potent cardenolides inhibitors of the Na^+/K^+-ATPase of highly adapted specialists like the monarch butterfly (*Danaus plexippus*) and large milkweed bug (*Oncopeltus fasciatus*) (25, 26). By combining molecular complexity metrics, metabolomics, molecular docking, and phylogenetic reconstructions across 52 milkweed species, we test 1) how the structural innovation of the N,S moiety restores potency against adapted herbivores, 2) whether this innovation has independent origins within the clade, and 3) and whether lineage-specific constraints and putative trade-offs shape the evolution of N,S-cardenolides. This work challenges simple models of defense escalation (19, 30) and offers a pathway forward for deciphering chemical evolution in species interactions.

Results

Molecular Complexity of Cardenolides and Target Site Inhibition. To assess incremental biosynthetic changes among cardenolides, we calculated DOX-g, a novel modification of the DOX index of molecular complexity that integrates the degree of oxidation state (31) adjusted by length of the glycosylated chain (32, 33). This index may be particularly useful for glycosylated natural products beyond cardenolides, as it reduces the effect of additional sugars on the index of complexity (See *Materials and Methods* for further explanation). Among the 184 cardenolides described from the genus *Asclepias,* N,S-compounds rank among the most structurally complex (Fig. 1). The most frequently reported N,S-compounds, voruscharin and uscharin, have DOX-g values ≥ 21, while labriformin is at 25, ranking in the top 4% owing to its highly substituted steroidal core (*SI Appendix*, *Supplementary Text*) (34–36). We propose biosynthetic pathways for uscharin and voruscharin, derived from coroglaucigenin, and labriformin and reduced labriformin, derived from syriogenin (Fig. 1). These two distinct routes both involve rare dioxane formation followed by the condensation of cysteine to form the thiazolidine heterocycle and its desaturation to install a 3-thiazoline ring; this last step is a hallmark of structural innovation in this metabolite class (cardenolides).

Cardenolide toxicity is commonly assessed using in vitro Na^+/K^+-ATPase inhibition assays, quantified as half of the maximal inhibitory concentration (IC_{50}), defined as the concentration required to reduce enzyme activity by 50%. Thus in a general sense, structural complexity could help explain the specific interaction with an enzymatic target (37). Indeed, DOX-g captures unique aspects of molecular complexity compared to existing indexes (38, 39) (*SI Appendix*, Figs. S1–S3 and Table S2) and reflects the scaffold elaboration that appears to be required before the addition of the N,S-ring. Across a limited set of purified compounds, we found little consistent relationship between DOX-g and IC_{50}, although we report an interaction between DOX-g and enzyme-type, suggesting that structural complexity may disproportionately impact toxicity to the adapted vs. sensitive enzyme ($t_{1,6}$ = 2.14, P = 0.07, *SI Appendix*, Fig. S5). In particular, cardenolides with increasing DOX-g scores from 17 to 25 showed increasing toxicity to the monarch enzyme by 14-fold, whereas toxicity to the sensitive enzyme remained unchanged, or slightly decreased, across the same complexity scores (*SI Appendix*, Fig. S5). Here, the key evolutionary step of adding a N,S-heterocycle restores potency against adapted Na^+/K^+-ATPases and thus explains why N,S-cardenolides are among the most potent inhibitors for highly resistant specialist herbivores (25, 26, 33).

Molecular Docking. Using AlphaFold 3 structural predictions of the monarch butterfly Na^+/K^+-ATPase, we modeled free energies and binding poses for N,S-cardenolides and their

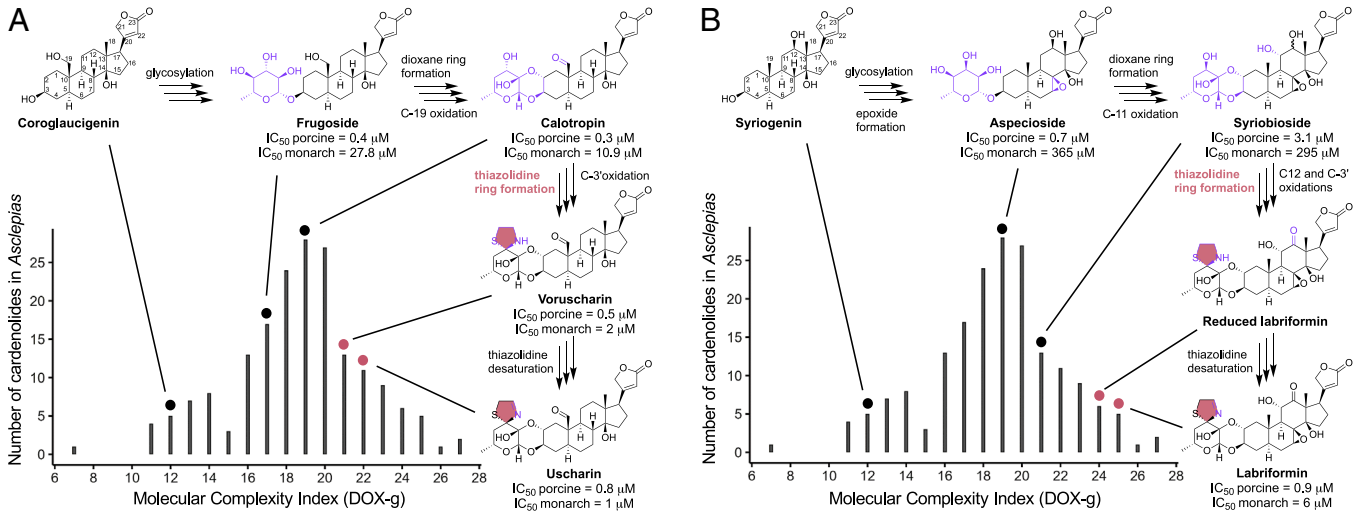


Fig. 1. Putative biosynthetic pathways for the four most common N,S-cardenolides. (*A*) From coroglaucigenin to uscharin and (*B*) from syriogenin to labriformin, highlighting sequential modifications (colored in purple), including thiazolidine ring formation (N,S-ring is colored in red), with enzyme inhibition values (IC_{50}) shown for a sensitive (porcine) and adapted (monarch) Na^+/K^+-ATPases. Toxicity values of syriobioside correspond to C12-α epimer. The histogram depicts the distribution of molecular complexity across 184 cardenolides.

non-N,S-precursors: calotropin and syriobioside. Separated by distinct aglycones (i.e., pathways starting with coroglaucigenin and syriogenin, respectively), the three ligands within each pathway showed differences in free energy (all pairwise P values < 0.001). Introduction of a five-membered thiazolidine or 3-thiazoline ring into the cardenolide scaffold markedly enhanced binding affinity to the monarch Na^+/K^+-ATPase relative to precursors. Among these compounds, voruscharin and reduced labriformin had the strongest predicted binding affinity (Fig. 2). This result suggests that the additional conversion of thiazolidine to 3-thiazoline, despite yielding uscharin and labriformin with weaker bonding affinity, may stabilize the molecule, facilitating storage and transport within the plant, or may expand ecological function by targeting a broader range of herbivores while maintaining sufficient potency to restore toxicity against the adapted monarch butterfly Na^+/K^+-ATPase.

Examination of the binding pocket revealed no obvious steric hindrance or residue-specific interactions that could explain the increased bioactivity of the N,S-cardenolides, and overall ligand conformation remained intact. Instead, an atom-by-atom examination of the energy contributions (*SI Appendix*, Fig. S6) suggests that the N,S-heterocycle ring improves binding primarily by adding favorable nonelectrostatic interactions (van der Waals, hydrogen bonds, and desolvation effects) around that portion of the molecule. Outside the N,S-heterocycle ring, neither nonelectrostatic nor electrostatic (charge-based) interactions changed substantially between N,S-cardenolides and their non-N,S-precursors. Across all molecules, nonelectrostatic forces were stronger on the lactone moiety and weaker on the glycosidic side, whereas electrostatic contributions displayed the opposite lactone–glycosidic side imbalance. Without the five-membered ring at the C3' substituent, calotropin showed greater conformational variability within the pocket, whereas the lactone moiety remained consistent (Fig. 2*A*).

Macroevolution of N,S-Cardenolide Toxins. To assess the distribution and maintenance of this structural innovation, we quantified the phylogenetic signal of four focal N,S-cardenolides (40) and used a SIMMAP model to simulate changes in trait states across a modern molecular phylogeny of *Asclepias* (41) (Fig. 3). Labriformin, reduced labriformin, uscharin, and voruscharin each showed phylogenetic signal (*SI Appendix*, Table S4), albeit with different levels of significance. Ancestral state reconstructions indicate that both voruscharin and uscharin represent an ancestral trait with several gains and losses (Fig. 3*B* and *SI Appendix*, Fig. S7), consistent with its occurrence in early diverging *Asclepias* (e.g., *Asclepias subulata*) and related outgroup genera such as *Calotropis* and *Gomphocarpus* (42). Conversely, labriformin evolved independently many times in *Asclepias* with minimal losses (89% of the 1,000 simulations with ≤2 losses) and this compound is especially prevalent in the North American temperate clade, in which >90% of species produce labriformin (Fig. 3*C*). The opposing patterns for uscharin and labriformin suggest a phylogenetic turnover during the radiation of *Asclepias*.

Given the distinct biosynthetic pathways for the formation of uscharin and labriformin (Fig. 1), we examined if there are negative correlations for the presence and abundance of those N,S-cardenolides. Based on analyses of the cardenolide content of latex from 36 species that produce at least one of the N,S-cardenolides, a Spearman's rank correlation on the abundances of these compounds revealed a strong negative relationship between the expression of labriformin and uscharin (the two end-products, see Fig. 3*B*, rho = −0.75, P < 0.001). A nonparametric correlation using Phylogenetic Independent Contrasts (PICs) corrected for zero inflated data showed the same negative relationship (S = 34563, rho −0.39, P-value = 0.003*) (Fig. 3*E*). This pattern is consistent with a macroevolutionary constraint in which lineages tend to specialize in one of two alternative N,S-cardenolide biosynthetic routes. Perhaps this, along with costs of production and phylogenetic conservatism, makes high abundances of the two major types of N,S-cardenolides unlikely in the same species. As these compounds exhibit broadly overlapping inhibitory effects on the highly adapted monarch Na/K-ATPase (Fig. 2), we next assessed whether they are potent enough to

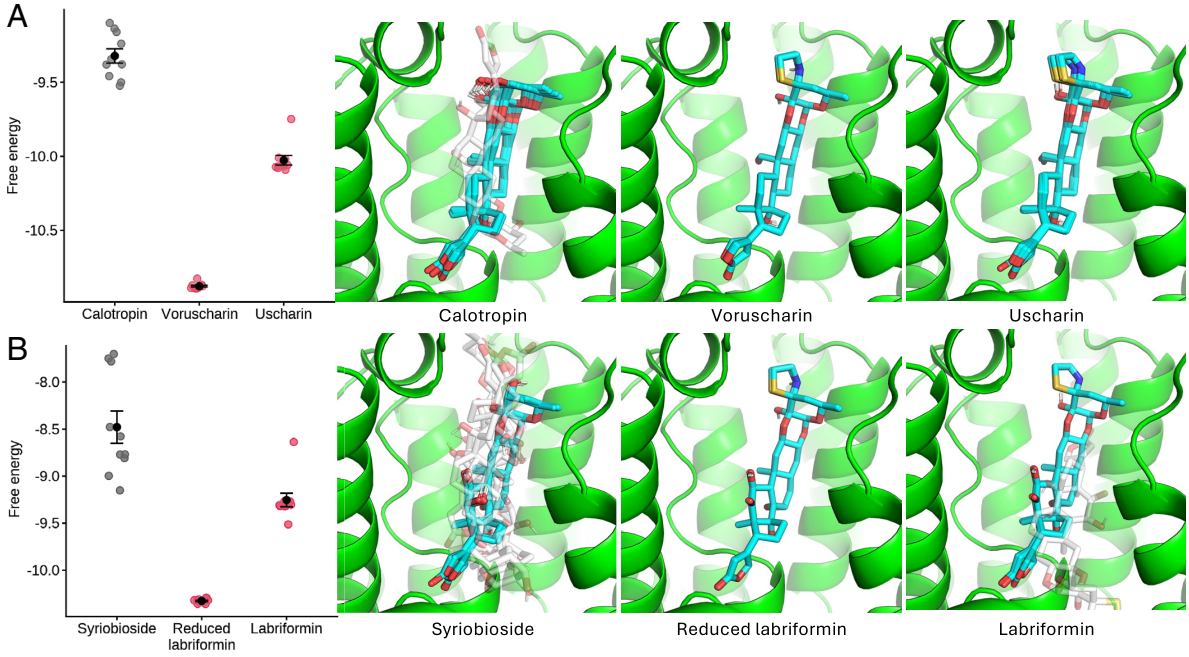


Fig. 2. Structural innovation of nitrogen-sulfur ring restores cardenolide toxicity in monarch butterfly. (*A*) Predicted binding free energies (black dot = mean ± SE across 10 replicate docking runs), and docking poses for N,S-cardenolides and their non-N,S precursors, coroglaucigenin and (*B*) syriogenin based ligands, using AlphaFold-based structural models of *D. plexippus* Na^+/K^+-ATPase sequence. Poses docking within the binding pocket are shown in cyan; nonbinding poses in semitransparent white.

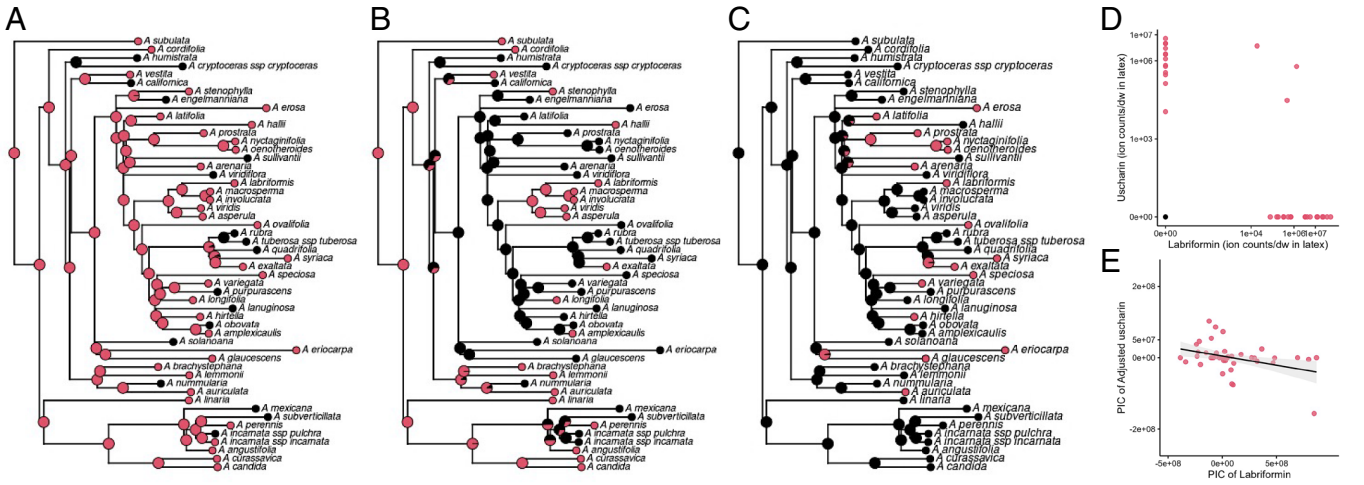


Fig. 3. Evolution and correlations of N,S-cardenolides in *Asclepias*. (*A*) Stochastic character mapping (SIMMAP) reconstruction of the evolutionary history of all N,S-cardenolides, (*B*) uscharin, and (*C*) labriformin in the latex of *Asclepias* species. Observed tip states are shown, with pie charts at internal nodes representing posterior probabilities of presence (red) or absence (black), inferred under an ER model using Bayesian MCMC sampling. Voruscharin and reduced labriformin are shown separately (*SI Appendix*, Fig. S7). (*D*) Correlation between raw ion abundances of labriformin and uscharin across the 36 *Asclepias* species that produce N,S-cardenolides; 16 species lacking both compounds collapse to a single point at the origin. (*E*) PICs reveal a negative association between labriformin and uscharin abundance after accounting for shared ancestry and zero inflation.

replace high concentrations of cardenolides or their high diversity in well-defended *Asclepias* species.

Remarkably, milkweed species that produce N,S-cardenolides exhibited >fourfold higher foliar cardenolide concentrations compared to non-N,S-cardenolide producers ($F_{1,52}$ = 7.42, phylogenetically corrected P = 0.003, Fig. 4*A*) and also have double the cardenolide richness in latex compared to species lacking these compounds ($F_{1,52}$ = 6.24, phylogenetically corrected P < 0.014, Fig. 4*B*; see *SI Appendix*, Fig. S8 for parallel results on foliar cardenolides). The latex metabolome of milkweeds producing N,S-cardenolides also showed expanded multivariate chemical space compared to non-N,S-cardenolide species, (phylogenetically corrected P = 0.001, Fig. 4*C* and *SI Appendix*, Fig. S9). Thus, species that produce N,S-cardenolides occupy more than twice the volume of metabolomic space covering secondary metabolite diversity (i.e., greater dispersion, mean distance to centroid = 20.17) compared to nonproducers (mean distance = 9.01) (F = 9.3, P = 0.001); this effect was not driven by the presence of N,S-cardenolides themselves (*Materials and Methods*).

The above results suggest that N,S-cardenolide producers are the most defended species. Accordingly, we tested leaf tissue extracts of 52 milkweed species in vitro on the monarch butterfly's Na^+/K^+-ATPase and found that species with N,S-cardenolides were 33% more inhibitive than species without N,S- compounds ($F_{1,52}$ = 6.33, phylogenetically corrected P = 0.009, Fig. 4*D*). To confirm that the observed patterns were related to the presence of complex metabolites (N,S-cardenolides) as part of plant defense, rather than being driven by a specific N,S-cardenolide, we also produced a group-level breakdown in *SI Appendix*, Fig. S10.

Discussion

Whereas cardenolides in milkweeds vary in terms of oxidation and glycosylation of the core structure, N,S-cardenolides represent an elaborate subset adding structural innovations through the incorporation of heterocycle rings. This structural novelty is not associated with increased toxicity to generalist herbivores; instead, these compounds restore potency against specialist Na^+/K^+-ATPases (25, 26). By rigidifying the cardenolide scaffold and enhancing nonelectrostatic interaction complementarity, these heterocycle rings reestablish strong binding affinity compared to precursors, thereby overcoming the target site insensitivity of highly specialized herbivores such as the monarch butterfly.

We show that two distinct N,S-cardenolide biosynthetic origins have evolved asymmetrically across the *Asclepias* phylogeny with contrasting macroevolutionary patterns, and are more widespread than previous phytochemical records suggested. N,S-cardenolides were produced in 69% of the species studied here, compared to <8% of the *Asclepias* spp. reported in previous studies. Uscharin, which was likely a trait in the ancestors of *Asclepias*, was frequently lost in early diversification of the genus, while labriformin evolved independently across multiple later-diverging *Asclepias* species. This pattern exemplifies nonparallel convergence, in which an adaptive advance against resistant herbivores is achieved not through modification of a single conserved trait, but via repeated innovations of distinct yet functionally similar metabolites (43). Our current results challenge the long-standing pattern of a unidirectional decline in cardenolide investment across *Asclepias* (21, 44), as the composition of the potent cardenolides has evolved nonlinearly.

Instead of following a phylogenetic escalation model (10, 45–47), where later-diverging species accumulate more potent or complex defenses, our results reveal a mosaic of gains and losses of innovative N,S-cardenolides built on increasingly complex scaffolds. Similar mosaic patterns occur in *Piper*, where correlated gains and losses among structurally related compound classes (e.g., flavonoids with chalcones; *p*-alkenyl phenols with kavalactones) reflect dependencies at the pathway level and early divergence in biosynthesis, resulting in lineage-specific chemical specialization rather than directional escalation. These pathway-specific structural modifications (e.g., successive prenylation, glycosylation, oxidation) generate novel bioactivity without increasing overall chemical abundance or diversity. Results from *Asclepias* and *Piper* together highlight biosynthetic innovation, rather than coevolutionary arms races, as a major driver of chemical defense evolution (48, 49) supporting the consensus view that chemical evolution involves iterative innovation and reconfiguration rather than monotonic increases in potency or complexity.

Our results suggest that the enhanced potency of N,S-cardenolides emerged from the stabilization of the ligand–enzyme complex, rather creating new residue-specific contacts. The increased

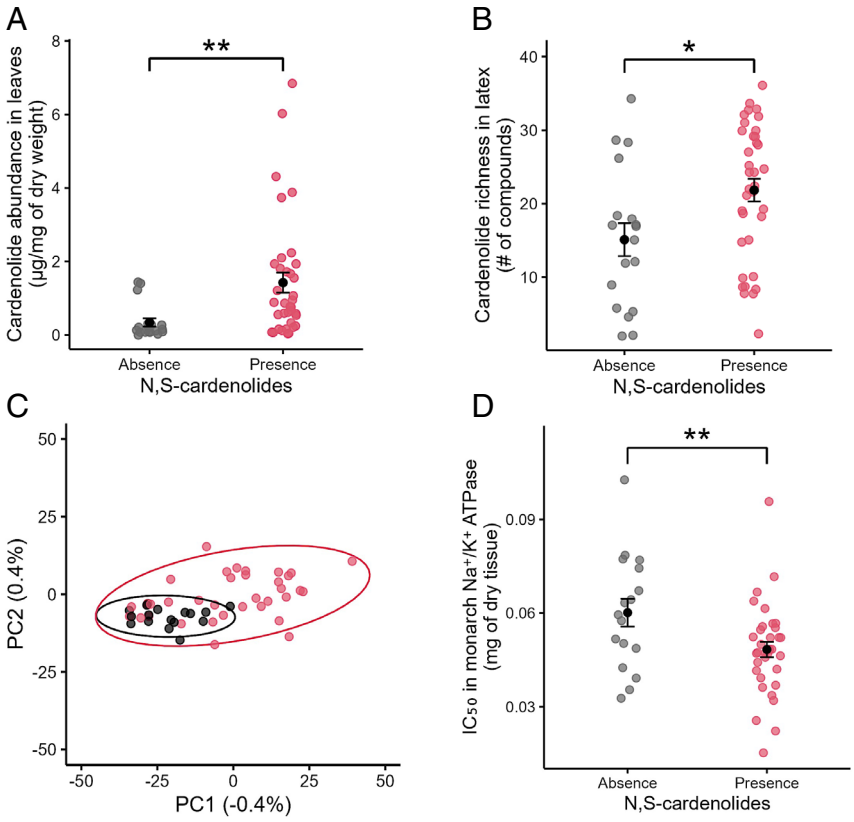


Fig. 4. Correlated trait evolution between milkweed species with (n = 39) and without (n = 12) N,S-cardenolides. (*A*) Cardenolide concentrations in leaves (measured by UV-HPLC), (*B*) cardenolide richness in latex (measured by targeted mass spectrometry), (*C*) latex metabolome composition based on 1,200 features with higher abundance (yellow = presence of N,S-cardenolides, blue = absence), and (*D*) IC_{50} of the monarch Na^+/K^+ ATPase, lower values represent higher toxicity. Panels *A*, *B*, and *D* display means ± SE; panel *C* shows and PC1 and PC2 from a Phy.PCA. All analyses incorporate phylogenetic correction to account for the shared evolutionary history among species.

conformational variability observed for N,S-cardenolides precursors, compared to N,S-cardenolides, suggests that the absence of the five-membered heterocycle allows for greater flexibility on the glycosidic side of the molecule. By adding additional surface contacts, the N,S-heterocycle ring likely helps lock the sugar side into place, reducing flexibility and making the complex more stable. This rigidification provides a plausible molecular mechanism by which N,S-cardenolides regain high binding affinity to the monarch Na^+/K^+-ATPase despite target-site resistance, an effect that has been born out in our physiological assays (Fig. 1).

The evolution of these specialized toxins reflects diversification constrained by lineage-specific biosynthetic precursors: species tend to produce either labriformin or uscharin, but rarely both compounds in high amounts, most likely influenced by metabolic costs. Demonstrating a true trade-off would require direct evidence of shared precursor limitation, enzymatic competition, or fitness consequences of simultaneous pathway expression, which remain to be tested. Importantly, the pattern of divergent evolutionary histories is not accompanied by reductions in overall phytochemical investment; species producing N,S-cardenolides exhibit higher total cardenolide abundance, greater richness, and occupy broader chemical space. Thus, the evolution of N,S-cardenolides reflects lineage-specific diversification, rather than a progressive escalation or decline. We hypothesize that the colonization of North America by *Asclepias*' ancestors from Africa led to selection for the losses of uscharin, as speciation proceeded from tropical central America to temperate climates with few specialist herbivores; nevertheless, as specialists accumulated on later diverging American *Asclepias*, natural selection favored the evolution of labriformin. Future research should examine the genetic basis of thiazolidine formation to explore whether labriformin biosynthesis originates from gene duplication, neofunctionalization, or from gene regulation of thiazolidine synthase (50).

Previous metrics of plant chemistry have emphasized diversity, composition, and functional breadth (1, 2), and recent work has suggested using chemical dissimilarity and complexity across plant metabolic profiles to understand plant defense (4, 51). In medicinal chemistry, increasingly sophisticated metrics have been developed to quantify molecular complexity, though their predictive value for bioactivity remains debated (37, 52). By contrast, structural innovation provides a distinct evolutionary axis, one that directly ties biosynthetic novelty to shifts in ecological function. Along with work in *Piper* and *Inga*, as well as other metabolically diverse clades, our results highlight structural innovation, compositional diversity, and pathway-level constraints as complementary and interacting axes in the evolution of plant defense.

In the case of N,S-cardenolides, structural innovation is an end product following the sequential biosynthesis of increasingly complex cardenolides. The nature of this study lies not in identifying the structural innovation per se, but in formalizing it within a pathway-resolved evolutionary framework that links biosynthetic sequence, chemical space, and ecological function in a model system, thereby reframing structural novelty as a dynamic process rather than an endpoint of escalation (53). Ultimately, evolutionary gains in structural innovation, to modify existing compounds or create new chemical classes, reveal how key biosynthetic steps generate new ecological opportunities in chemical evolution, and may represent an underappreciated mode of defensive adaptation.

Materials and Methods

Structural Complexity of Reported Milkweed Cardenolides. To assess the structural complexity of cardenolides in *Asclepias*, we generated a milkweed cardenolide database using CAS SciFinder®, retrieving entries reported through the fall of 2023. We identified compounds based on the cardenolide core structure [tetracyclic C (23)-steroid with methyl groups at C-10 and C-13 and a

five-membered lactone at C-17] and filtered for entries that explicitly reported isolated substances (excluding derivatives or referenced compounds) associated with the keyword "Asclepias." From this dataset, we calculated eight molecular descriptors to assess structural complexity using chemical formulas and SMILES representations: exact mass, fraction of sp^3 carbons (Fsp3), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), topological polar surface area (TPSA), and degree of oxidation (DOX).

Rationale for DOX-g as a Novel Structural Complexity Metric. The degree of oxidation (DOX) metric quantifies molecular oxidation states by considering the number of oxygen atoms in relation to the degree of unsaturation (DU) (54). While DOX is effective in assessing oxidation-driven complexity, it fails to account for structural modifications involving heteroatoms. Extensions to include heteroatoms in oxidation state calculations have been explored in organic chemistry (31), but their application to natural product complexity has not been developed. Other indexes of structural complexity could have been applied (34) (*SI Appendix*, Table S1). However, since we are comparing structures within a specific chemical family with restricted scaffolds differentiated by hydroxylation and heteroatom patterns, the adjusted degree of oxidation is a relevant metric for this study. Additionally based on Petschenka et al. (32) and Rubiano-Buitrago et al. (33), increasing the number of glycosyl units in cardenolides beyond the first glycosyl unit, does not imply a change in their inhibitory potential against Na^+/K^+-ATPase pump of animals. In general, mono- and diglycosylated genins are more toxic than nonglycosylated genins, whereas the presence of one vs. two sugars does not show a clear tendency to predict the toxicity of cardenolide glycosides. To take into account this glycosyl-related structure–activity relationship, we adjust glycosylation pattern by correcting for the additional oxygen contributions from extra sugar moieties. Consequently, we developed DOX-g, a modified degree of oxidation (DOX) formula that accounts for the structural effects of glycosylation and heteroatom incorporation to refine structural complexity assessments:

$$\mathrm{DOX-g} = \mathrm{DU} + (a/(1 + b)) + c.$$

DU: degree of unsaturation (31)
a: number of oxygen
b: number of sugars beyond 1st glycosyl unit
c: number of heteroatoms (i.e., N, S, P)

Using the calculated DOX-g values, we ranked all known milkweed cardenolides by complexity and analyzed their distribution across the dataset. Principal component analysis (PCA) was performed to determine how DOX-g compares to other structural descriptors in explaining variance among cardenolides (*SI Appendix*, Table S2). We also performed pairwise Pearson correlations between DOX-g and existing measures of structural complexity to validate its ability to capture molecular elaboration in cardenolide biosynthesis (*SI Appendix*, Figs. S1–S3).

Molecular Docking and Energy Decomposition Analysis. The three-dimensional structure of *D. plexippus* Na^+/K^+-ATPase (GenBank ID: JQ771507.1) was predicted using AlphaFold 3 (55). Ligand conformers were either downloaded from PubChem (3D conformer) or manually constructed using Avogadro (version 1.2.0), followed by geometry optimization with Universal Force Field (UFF) using the same software. Hydrogen atoms were added to the protein using PyMOL (version 3.1.0), and both protein and ligand structures were converted to docking-compatible *.pdbqt format using AutoDockTools in MGLTools (version 1.5.7) (56). This step also included the addition of Gasteiger charges and the assignment of rotatable bonds. Grid parameter files were generated using AutoDockTools, resulting in a grid box of 32 × 32 × 32 $Å^3$ with a spacing of 0.2 Å between grid points. The center of the grid box was defined by aligning the AlphaFold-predicted Na^+/K^+-ATPase structure to the cocrystal structure of *Sus scrofa* Na^+/K^+-ATPase in complex with ouabain (PDB ID: 4HYT), and extracting the geometric center of ouabain as the reference point. Docking simulations were carried out using AutoDock 4 (version 4.2.6) (56) with the Lamarckian genetic algorithm. For each ligand, docking was performed with 10 different random seeds using default settings. The resulting poses were visually inspected in PyMOL, and the estimated free energy of binding, per-atom contributions of nonelectrostatic and electrostatic interactions, and partial atomic charges for each pose were extracted and analyzed from the docking log files using a custom Python script.

Preparation of *Asclepias* Leaves and Latex Samples. We grew 52 *Asclepias* species from seed, based on our long-term collection of plant material from the field, native plant nurseries, and colleagues, all of which are also sampled in the newest molecular phylogeny of the genus (40). Seeds were surface sterilized in June 2019 by soaking in 10% bleach for 10 min, and rinsing well with deionized water. Seeds were nicked gently and placed between layers of moist paper towels, within petri dishes sealed with parafilm. They were stratified at 4 °C for 5 d, before being moved into a 30 °C chamber for germination. After 3 d, seedlings were planted in 500 mL pots filled with premoistened LM-111 All-purpose soil mix (Lamberts, Riviere-Ouelle, Quebec). Seedlings were grown in a growth chamber at 27 °C d/24 °C night with a 14 h day length, and watered as needed. After approximately 6 wk, two to six seedlings per species (176 total seedlings) were up-potted into 3.8L pots and moved outdoors, into a large mesh cage at a local field site in Tompkins Co., NY. Three weeks later, the plants were harvested for leaf and latex samples. First, the tip of the youngest undamaged fully expanded leaf was snipped to remove approximately 2 mm of the leaf, and all exuded latex was collected onto a 1 cm^2 circle of filter paper, and placed into a preweighed 2 mL microcentrifuge tube. Tubes containing latex samples were kept in a cooler, then reweighed fresh, to yield the mass of latex exuded. Four days later, we collected a sample of 10 representative leaves from each plant, starting at the apex and collecting approximately 1 leaf per node, to achieve even sampling down to the bottom of the plant. Immediately following leaf collection, we collected as much latex as possible, as droplets from the cut petioles, directly into one glass vial per plant. Leaves were weighed fresh and both leaves and latex were stored on dry ice in the field, until being moved into the freezer (−80 °C for leaves; −20 °C for latex).

Frozen leaf and latex samples were freeze-dried (Labconco, Kansas City, MO), and cardenolides were extracted as described in Petschenka et al. (32). Briefly, leaf tissue was ground to a fine powder in 2 mL tubes, using 3 mm stainless steel beads and a Retsch mixer mill (Haan, Germany). Typically, 50 mg of ground leaf tissue was moved into a 2 mL screw cap tube (Sarstedt, Nümbrecht, Germany) containing 20 zirconia-silica beads for extraction. We added 1.2 mL methanol, spiked with 15 μg hydrocortisone, to each sample. For plant samples with less than 50 mg leaf tissue available, we used either 20 mg (plus 0.48 mL spiked methanol) or 40 mg (plus 0.96 mg spiked methanol). Samples were extracted in a FastPrep (MP Biomedicals, Santa Ana, CA) twice at 6.5 m/s for 45 s, and centrifuged at 20,817 g for 12 min. An aliquot equivalent to 10 mg dry leaf tissue was removed from each sample for a separate analysis (see in vitro inhibition assays, below), and remaining supernatants were taken to dryness in a rotary evaporator (Labconco, Kansas City, MO).

Freeze-dried latex samples were weighed, in their respective glass vials, and empty vials were reweighed after extraction to give the dry mass of each latex sample. Dry latex samples ranged from near zero to over 40 mg in mass. One milliliter of methanol, spiked with 20 μg digitoxin, was added to each vial, and samples were vortexed and sonicated to bring as much latex as possible into solution. Samples were mixed on a rotary shaker at 200 rpm overnight, and transferred into a 2 mL screw cap tube (as above). Glass vials were washed once more with 0.5 mL methanol, and all remaining sample transferred to the 2 mL tube, with zirconia-silica beads. Samples were extracted, and supernatants taken to dryness, as above. Leaf and latex extract residues were brought up in 200 μL methanol, filtered using 0.2 mm syringe filters (Restek, Bellefonte, PA) and pipetted into glass LC vials with polyspring inserts (Chemglass Life Sciences, Vineland, NJ), for subsequent analysis.

Determination of Plant Extract Bioactivity. For each leaf extract, an aliquot representing 10 mg of starting dry leaf tissue (see above) was removed to a separate tube and taken to dryness for use in in vitro inhibition assays of Na^+/K^+-ATPase. Extracts were diluted in 25 μL DMSO and then brought up to 250 μL with millipore water for a final reconstitution in 10% DMSO. Three serial 10× dilutions were made of each stock, using 10% DMSO, to yield 4 concentrations of each extract, representing 0.8, 0.08, 0.008, and 0.0008 mg dry tissue per sample well. All extracts were randomized across assay plates, with 22 extracts (x4 concentrations) typically run per plate, alongside a set of ouabain controls and background wells. These preparations were tested against Na^+/K^+-ATPase from a nonadapted animal (porcine, MilliporeSigma, Burlington, MA) as well as the cardenolide-adapted monarch butterfly; inhibition of only the monarch enzyme is reported here. Monarch nervous tissues were dissected, homogenized, and freeze-dried, as the source of the Na^+/K^+-ATPase in the assays. Assay procedure followed that described in Petschenka et al. (26), except that all plant extracts and ouabain standards were prepared in 10% DMSO, as above. After comparison of the results of using individual backgrounds

(each extract dilution run in duplicate, with and without KCl) to those using a universal background (one reaction per plate lacking KCl and a fully inhibitive concentration of an ouabain standard), we found that the latter method gave similar results for this large dataset. Thus, all reaction absorbance values were corrected by subtracting the plate-specific background absorbance. These background-corrected absorbance values were used directly in subsequent model fitting.

Sigmoidal dose–response curves for each of our 176 plant extracts on each enzyme were fitted using a mixed model framework using the *nlme()* function in R (as described in Petschenka et al. 2022) (26). Model fitting was performed separately for each enzyme. Due to the small number of datapoints in each sigmoidal curve (4 points), we used an iteration procedure to estimate and fix the value for *scal* in the 4-parameter logistic model, while estimating the upper asymptote, lower asymptote, and *xmid* for each dilution series. This last parameter was exported as an estimate of the amount of dry plant tissue required to inhibit each enzyme type by 50% (IC_{50}). While each extract was tested at least once with each enzyme type, there were several cases where multiple replicates were run–in those cases, we calculated the mean IC_{50} for each plant extract and enzyme.

Chemical Analysis of Cardenolide Content. HPLC-MS: Samples were injected in a Thermo Fisher Scientific Vanquish UHPLC system coupled with a Thermo Q-Exactive HF hybrid quadrupole-orbitrap high-resolution mass spectrometer equipped with a HESI ion source. Metabolites were separated using acetonitrile containing 0.1% formic acid (organic phase) and 0.1% formic acid in water (aqueous phase) as solvents on a Thermo Fisher Scientific. Methanolic extracts were separated on an Agilent Zorbax Eclipse XDB-C18 column (150 mm × 2.1 mm; particle size: 1.8 μm), maintained at 40 °C with a flow rate of 0.5 mL/min. Solvent A contained 0.1% formic acid in water; solvent B contained 0.1% formic acid in acetonitrile. A/B gradient started at 5% B for 2 min after injection and increased linearly to 98% B at 11 min, followed by 3 min at 98% B, then back to 5% B over 0.1 min, and finally held at 5% B for an additional 2.9 min to re-equilibrate the column.

The mass spectrometer parameters were spray voltage of 3.5 kV, capillary temperature of 380 °C, probe heater temperature of 400 °C; 60 sheath flow rate, 20 auxiliary flow rate and 1 spare gas; and S-lens RF level of 50, resolution of 240,000 at m/z 200, automatic gain control AGC target of 3e6. The instrument was calibrated weekly with positive ion calibration solutions (Thermo Fisher). Each sample was analyzed in positive ionization modes using a m/z range of 100 to 1,100. Parameters for data-dependent tandem mass spectrometry (MS/MS) (dd-MS2) were MS1 resolution 60,000; AGC target 1e6. MS2 resolution was 30,000, AGC target 2e5, maximum injection time 50 ms, isolation window 1.0 m/z, stepped normalized collision energy (NCE) 10, 30; dynamic exclusion 1.5 s, top five masses selected for MS/MS per scan.

Cardenolide richness was quantified as the number of distinct retention times associated with m/z values that correspond to *Asclepias* spp. genins. These m/z genins appear in the feature tables at MS1 level due to ionization in the source and putative structures were confirmed by MS2 profile. The genin vector consists on 125 values of *mz* encompassing 25 $[M+H]^+$ adducts of 52 milkweed genins with a margin ±0.002 or 6 ppm. While the current genin vector in the R markdown is limited to samples of *Asclepias* origin (i.e., plant or insect-derived), the vector file can be readily modified to target any specific motif in MS1 data based on other cardenolides or other compounds of interest (*SI Appendix*, Fig. S4).

HPLC-DAD: Leaf extracts were injected into an Agilent 1100 system, and separated on a C18 reversed phase column, exactly as described in Petschenka et al. (26). Cardenolide abundance was estimated as the sum of the areas of chromatography peaks with an absorbance of λ 218 nm (214 to 222 nm) corrected by the internal standard, digitoxin (26).

Mining N,S-Cardenolides in Milkweed Chemistry. Voruscharin and uscharin, along with reduced labriformin and labriformin represent the two principal steroid cores of the N,S-cardenolides reported in Asclepiadoideae subfamily. Their functional role as inhibitors of the Na^+/K^+-ATPase pump in specialist herbivores within the milkweed system has been well documented. To assess their distribution, we screened for the presence of these four N,S-cardenolides in latex samples from 51 *Asclepias* species using UPLC-HRMS/MS data. Additional N,S-cardenolides such as thiazolidinone calotropin, S-oxythiazolidinone calotropin, and 16α-hydroxyuscharin have been reported, but were either below our concentration threshold or could not be annotated with confidence due to same molecular weight/formula with other cardenolides. The 52 species were analyzed by 165 LC-MS injections of latex samples, with 39 species having ≥3 replicates, nine species having two replicates, three species with one replicate.

We mined for N,S-cardenolides across samples using a merged feature table of the latex LC-MS data. Briefly, in R studio, we applied a predefined a m/z vector including common adducts of voruscharin, uscharin, reduced labriformin and labriformin to the feature table (.csv) obtained in for Wizard tool in MZmine 4.0. Parental mass for the four cardenolides was detected predominantly as $[M+H]^+$ adduct. The resulting matrix was filtered by ion counts (area) above 1e5 (before dried weight correction), and using a refined combination of retention time plus m/z for the four N,S-cardenolide analyzed.

We marked the presence of each N,S-cardenolides and the different possible combinations found in the latex samples of milkweeds. We consider the presence of the N,S-cardenolide where retention times and m/z were consistent for ≥2 replicates in the feature table. For the samples where only one replicate was evidence for one or several of the N,S-cardenolides inspection of the raw data was carried out thoroughly. Those putative matches were validated by overviewing if MS/MS spectra remained consistent to GNPS reference library for the reported cardenolides (voruscharin: CCMSLIB00006718039, uscharin: CCMSLIB00006718040, labriformin: CCMSLIB00006581614). And direct overlaying with species that hold a phytochemical record (i.e., *Asclepias syriaca* for labriformin, *Asclepias linaria* and *Asclepias curassavica* for uscharin). Thus, while ensuring that the *mz* with higher intensity at the proper retention time range corresponds to the expected parental mass. Overall, we validated the presence of N,S-cardenolides in 36 species for the 52 studied here (*SI Appendix*, Table S3).

N,S-Cardenolides Distribution in Milkweed Phylogeny. To assess whether the known cardenolides, voruscharin, uscharin reduced labriformin and labriformin, have a phylogenetic signal we treated their presence/absence as categorical trait and calculated Pagel's δ using the Delta function by Borges et al. (40). The phylogeny used is pruned from a nearly comprehensive and well-supported plastome tree of the genus (41); our sample covers nearly 40% of the species in the genus and all major clades. Values of δ near zero indicate little to no phylogenetic structure, whereas values greater than 1 suggest strong similarity among close relatives than expected under the Brownian motion evolutionary model.

Stochastic character mapping for presence/absence was estimated using an equal-rates (ER) model using Bayesian MCMC sampling using the SIMMAP function from "phytools" package.

Correlation Analysis. Phylogenetically Independent Contrasts (PIC), corrected for zero inflation to compare abundance of labriformin vs. uscharin was made using "ape" and "pscl" package. Phylogenetic corrections for correlated trait evolution between milkweed species with and without N,S-cardenolides in cardenolide concentration in leaves, cardenolide richness in latex and fresh mass of latex exudation was carried out with function aov.phylo of "geiger" package. Phylogenetic PCA was performed using phylogenetic principal component analysis (Phy.PCA) function from "phytools" package, and PERMANOVA with PC1 and PC2 loadings was made with "adonis2" function from package "vegan." For nonphylogenetic corrected PCA, loadings that are in the PC1 and PC2 used for PERMANOVA analysis were ranked for the top 20 features that influence the variance of those two principal components. Later, compared with the dataset of the latex LC-MS data, and filtered to search for the features corresponding to N,S cardenolides. On the top 40 features only uscharin appears. Top five or 10 features yield no hits for N,S cardenolides.

Data, Materials, and Software Availability. All MS data have been deposited in MassIVE https://massive.ucsd.edu: MSV000099124 (57) for leaves and MSV000099127 (58) for latex. Code for data analysis and figure generation can be found at Cornell e-commons: https://doi.org/10.7298/7sgs-jb78) (59).

ACKNOWLEDGMENTS. This work was improved by input from Andre Kessler, Georg Jander, Corrie Moreau, and Jennifer Thaler. We thank Mark Fishbein for providing the tree for the phylogeny of *Asclepias*, Chelsea Lee for assistance with plant care and sample collection, Maddy Oakes for help with latex cardenolide extractions, and Alfonso Aceves-Aparicio for support and feedback on data analysis. This work was supported by a grant from the US NSF (IOS-2209762) and fellowships to M.U. from the Japan Society for the Promotion of Science and the Human Frontiers Science Program.

Author affiliations: [a]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853; [b]Department of Chemistry and Chemical Biology and Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY 14853; [c]Department of Entomology, Cornell University, Ithaca, NY 14853; and [d]Department of Entomology, Cornell AgriTech, Cornell University, Geneva, NY 14456

1. M.-J. Endara, D. L. Forrister, P. D. Coley, The evolutionary ecology of plant chemical defenses: From molecules to communities. *Annu. Rev. Ecol. Evol. Syst.* **54**, 107–127 (2023).
2. A. Kessler, A. Kalske, Plant secondary metabolite diversity and species interactions. *Annu. Rev. Ecol. Evol. Syst.* **49**, 115–138 (2018).
3. W. C. Wetzel, S. R. Whitehead, The many dimensions of phytochemical diversity: Linking theory to practice. *Ecol. Lett.* **23**, 16–32 (2020).
4. C. S. Philbin, L. A. Dyer, C. S. Jeffrey, A. E. Glassmire, L. A. Richards, Structural and compositional dimensions of phytochemical diversity in the genus Piper reflect distinct ecological modes of action. *J. Ecol.* **110**, 57–67 (2022).
5. G. D. Moghe, B. J. Leong, S. M. Hurney, A. D. Jones, R. L. Last, Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *Elife* **6**, e28468 (2017).
6. K. A. Bird, A. A. Ramos, D. J. Kliebenstein, Phylogenetic and genomic mechanisms shaping glucosinolate innovation. *Curr. Opin. Plant Biol.* **85**, 102705 (2025).
7. M. Donoghue, Key innovations, convergence, and success: Macroevolutionary lessons from plant phylogeny. *Paleobiology* **31**, 77–93 (2005).
8. M. Wink, Plant secondary metabolism: Diversity, function and its evolution. *Nat. Prod. Commun.* **3**, 1934578X0800300 (2008).
9. M. Wink, Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3–19 (2003).
10. M. Berenbaum, P. Feeny, Toxicity of angular furanocoumarins to swallowtail butterflies: Escalation in a coevolutionary arms race? *Science* **212**, 927–929 (1981).
11. S. R. Whitehead, E. Bass, A. Corrigan, A. Kessler, K. Poveda, Interaction diversity explains the maintenance of phytochemical diversity. *Ecol. Lett.* **24**, 1205–1214 (2021).
12. D. Grzech *et al.*, Incorporation of nitrogen in antinutritional Solanum alkaloid biosynthesis. *Nat. Chem. Biol.* **21**, 131–142 (2025).
13. S. Desmet, K. Morreel, R. Dauwe, Origin and function of structural diversity in the plant specialized metabolome. *Plants* **10**, 2393 (2021).
14. D. Bajusz, A. Rácz, K. Héberger, Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **7**, 20 (2015).
15. M. Blanchard, L. M. Holeski, Consequences and costs of chemical complexity: The evolutionary ecology of direct phytochemical defense against herbivores. *Int. J. Plant Sci.* **185**, 3–14 (2024).
16. T. Garland, C. J. Downs, A. R. Ives, Trade-offs (and constraints) in organismal biology. *Physiol. Biochem. Zool.* **95**, 82–112 (2022).
17. C. D. K. Graham, E. J. Forrestel, A. L. Schilmiller, A. T. Zemenick, M. G. Weber, Evolutionary signatures of a trade-off in direct and indirect defenses across the wild grape genus, Vitis. *Evolution* **77**, 2301–2313 (2023).
18. D. Li, R. Halitschke, I. T. Baldwin, E. Gaquerel, Information theory tests critical predictions of plant defense theory for specialized metabolism. *Sci. Adv.* **6**, eaaz0381 (2020).
19. D. J. Futuyma, A. A. Agrawal, Macroevolution and the biological diversity of plants and herbivores. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18054–18061 (2009).
20. S. B. Malcolm, L. P. Brower, Evolutionary and ecological implications of cardenolide sequestration in the monarch butterfly. *Cell. Mol. Life Sci.* **45**, 284–295 (1989).
21. A. A. Agrawal, M. Fishbein, Phylogenetic escalation and decline of plant defense strategies. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10057–10060 (2008).
22. K. A. Mooney, R. Halitschke, A. Kessler, A. A. Agrawal, Evolutionary trade-offs in plants mediate the strength of trophic cascades. *Science* **327**, 1642–1644 (2010).
23. M. Karageorgi *et al.*, Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature* **574**, 409 (2019).
24. A. A. Agrawal, A. P. Hastings, C. Duplais, Testing the selective sequestration hypothesis: Monarch butterflies preferentially sequester plant defences that are less toxic to themselves while maintaining potency to others. *Ecol. Lett.* **27**, e14340 (2024).
25. A. A. Agrawal *et al.*, Cardenolides, toxicity, and the costs of sequestration in the coevolutionary interaction between monarchs and milkweeds. *Proc. Natl. Acad. Sci. U.S.A.* **118**, 2024463118 (2021).
26. A. A. Agrawal *et al.*, Functional evidence supports adaptive plant chemical defense along a geographical cline. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2205073119 (2022).
27. F. Brüschweiler, K. Stöckel, T. Reichstein, Calotropis -Glykoside, vermutliche Teilstruktur. Glykoside und Aglykone, 321. Mitteilung. *Helv. Chim. Acta* **52**, 2276–2303 (1969).
28. J. N. Seiber, S. M. Lee, J. M. Benson, "Cardiac glycosides (cardenolides) in species of Asclepias (Asclepiadaceae)" in *Handbook of Natural Toxins, Plant and Fungal Toxins*, R. F. Keeler, A. T. Tu, Eds. (Marcel Dekker, Amsterdam, 1983), vol. I, pp. 43–83.
29. J. N. Seiber, C. J. Nelson, S. M. Lee, Cardenolides in the latex and leaves of seven Asclepias species and *Calotropis procera*. *Phytochemistry* **21**, 2343–2348 (1982).
30. N. I. Cacho, D. J. Kliebenstein, S. Y. Strauss, Macroevolutionary patterns of glucosinolate defense and tests of defense-escalation and resource availability hypotheses. *New Phytol.* **208**, 915–927 (2015).
31. L. H. Klemm, A numerical measure of the degree of oxidation (DOX) of an organic molecule, with special attention to heteroatoms and heterocyclic compounds. *J. Heterocycl. Chem.* **32**, 1509–1512 (1995).
32. G. Petschenka *et al.*, Relative selectivity of plant cardenolides for Na+/K+-ATPases from the Monarch butterfly and non-resistant insects. *Front. Plant Sci.* **9**, 1424 (2018).
33. P. Rubiano-Buitrago, S. Pradhan, C. Paetz, H. M. Rowland, New structures, spectrometric quantification, and inhibitory properties of cardenolides from *Asclepias curassavica* seeds. *Molecules* **28**, 105 (2023).
34. L. Rodríguez-Hahn, G. Fonseca, The cardenolide content of *Asclepias linaria*. *Phytochemistry* **30**, 3941–3942 (1991).
35. G. H. Robinson, G. E. Burrows, E. M. Holt, R. J. Tyrl, A. D. Jones, "Investigation of the neurotoxic compounds in Asclepias subverticillata, western-whorled milkweed" in *Proceedings of the International Symposium on Poisonous Plants*, T. Garland, A. C. Barr, Eds. (CAB International, 1998), pp. 435–439.
36. M. C. Roy, F. R. Chang, H. C. Huang, M. Y. N. Chiang, Y. C. Wu, Cytotoxic principles from the Formosan milkweed, Asclepias curassavica. *J. Nat. Prod.* **68**, 1494–1499 (2005).
37. D. Lowe, Measuring molecular complexity (2023). https://www.science.org/content/blog-post/measuring-molecular-complexity.
38. S. H. Bertz, The first general index of molecular complexity. *J. Am. Chem. Soc.* **103**, 3599–3601 (1981).
39. W. Wei, S. Cherukupalli, L. Jing, X. Liu, P. Zhan, Fsp3: A new parameter for drug-likeness. *Drug Discov. Today* **25**, 1839–1845 (2020).
40. R. Borges, J. P. Machado, C. Gomes, A. P. Rocha, A. Antunes, Measuring phylogenetic signal between categorical traits and phylogenies. *Bioinformatics* **35**, 1862–1869 (2019).
41. M. Fishbein *et al.*, Evolution at the tips: Asclepias phylogenomics and new perspectives on leaf surfaces. *Am. J. Bot.* **105**, 514–524 (2018).
42. H. T. A. Cheung, F. C. K. Chiu, T. R. Watson, R. J. Wells, Cardenolide glycosides of the asclepiadaceae. New glycosides from asclepias fruticosa and the stereochemistry of uscharin, vorusharin and calotoxin. *J. Chem. Soc. Perkin Trans.* **1**, 2827 (1983), 10.1039/p19830002827.
43. E. Twomey *et al.*, Multiple routes to color convergence in a radiation of Neotropical poison frogs. *Syst. Biol.* **72**, 1247–1261 (2023).
44. A. A. Agrawal *et al.*, Evidence for adaptive radiation from a phylogenetic study of plant defenses. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18067–18072 (2009).
45. J. X. Becerra, K. Noge, D. L. Venable, Macroevolutionary chemical escalation in an ancient plant–herbivore arms race. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18062–18066 (2009).
46. P. P. Edger *et al.*, The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8362–8366 (2015).
47. C. W. Wheat *et al.*, The genetic basis of a plant-insect coevolutionary key innovation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20427–20431 (2007).
48. L. A. Richards *et al.*, Phytochemical diversity drives plant-insect community diversity. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10973–10978 (2015).
49. K. A. Uckele *et al.*, Phytochemistry reflects different evolutionary history in traditional classes versus specialized structural motifs. *Sci. Rep.* **11**, 17247 (2021).
50. C. P. Drummond, T. Renner, Genomic insights into the evolution of plant chemical defense. *Curr. Opin. Plant Biol.* **68**, 102254 (2022).
51. G. F. Schneider, D. Salazar, S. B. Hildreth, R. F. Helm, S. R. Whitehead, Comparative metabolomics of fruits and leaves in a hyperdiverse lineage suggests fruits are a key incubator of phytochemical diversification. *Front. Plant Sci.* **12**, 693739 (2021).
52. K. Martinez-Mayorga *et al.*, The pursuit of accurate predictive models of the bioactivity of small molecules. *Chem. Sci.* **15**, 1938–1952 (2024).
53. M. R. Berenbaum, A. R. Zangerl, "Phytochemical diversity: Adaptation or random variation?" in *Phytochemical Diversity and Redundancy in Ecological Interactions*, J. T. Romeo, J. A. Saunders, P. Barbosa, Eds. (Springer, Boston, MA, 1996), pp. 1–24.
54. J. B. Hendrickson, D. J. Cram, G. S. Hammond, *Organic Chemistry* (McGraw-Hill, ed. 3, 1970).
55. J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
56. G. M. Morris *et al.*, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
57. P. Rubiano-Buitrago *et al.*, Data of leaves extracts from 60sp of Asclepias sp. MassIVE Repository. https://doi.org/doi:10.25345/C5HX1641M. Deposited 10 September 2025.
58. P. Rubiano-Buitrago *et al.*, Data of latex extracts from 60sp of Asclepias sp. MassIVE Repository. https://doi.org/doi:10.25345/C54M91Q06. Deposited 10 September 2025.
59. P. Rubiano-Buitrago *et al.*, Data from: Chemical innovation and structural complexity in the macroevolution of plant defense [Data set]. Cornell University Library eCommons Repository. https://doi.org/10.7298/7SGS-JB78. Deposited 22 August 2025.

PNAS

Supporting Information for

Structural innovation in the evolution of plant chemical defense

Paola Rubiano-Buitrago, Amy P. Hastings, Masaaki Uematsu, Jeremy M. Baskin, Anurag A. Agrawal*, Christophe Duplais*

Corresponding authors:
Christophe Duplais c.duplais@cornell.edu
Anurag A. Agrawal aa337@cornell.edu

This PDF file includes:

Supporting Information Text

HPLC-MS data processing for cardenolide pipeline and list of mz values of genins

For the assessment of cardenolide content using a genin vector pipeline, LC-MS data requires the following preprocessing steps. Raw LC-MS files were first converted from .raw to .mzML using ProteoWizard MSconvert v3.0 and then processed in MZmine 4.2.0 (mzio GmbH) using the "mzwizard" tool with the predeterminate parameters for UPLC chromatography and Orbitrap spectrometry. Two feature tables were generated, one with leaves data and another for latex samples, both later exported for molecular network analysis. The resulting quantification file ("_quantif.csv") was refined by removing non-essential columns (D-M), renaming headers of column A (row ID), B (row mz), C (row retention time) with id, mz, and rt respectively. The number of decimals was reduced in mz (column B) to optimize filtering in R Markdown. The final processed dataset was saved as a .csv file in the "/data" folder within the appropriate R project environment. In R Studio, we filtered the feature table (.csv) obtained from the LC-MS analysis using a predefined mz vector containing the mz associated with cardenolide genins. The use of the genins vector in the R markdown is limited to samples from *Asclepias* related origin. To analyze abundance as a continuous variable, ion counts were normalized to dry weight to account for variation in sample concentration.

List of mz values of genins and theoretical genins based on reported cardenolides from milkweed used in genin vector (numbers in parenthesis reference structures in Fig. S4: 375.253(1); 391.248 (2-5-6-18); 357.242 (3), 389.232 (4-7-30), 373.238 (8-9-10-22), 407.243 (11-13-17-19-26), 405.228 (12-14-15-16-25-32) 421.223 (20-33-34) 403.212 (21-29) 393.228 (23) 437.218 (24-28) 435.202 (27) 423.238(31) 409.223 (35) 387.217 (36-31.1) 357.207 (23.1) 401.196 (24.1-28.1) 383.186 (24.2-28.2) 369.207 (25.1-31.2), 399.181 (27.1) 381.170 (27.2) 367.191 (29.1-33.2) 385.202 (33.1-34.1) 373.202 (35.1), 355.191(35.2)

Chemical complexity of milkweed cardenolides

Cardenolides exhibit low to high structural diversity, with oxidation and glycosylation as key drivers of chemical complexity. Most milkweed cardenolides have moderate polarity (TPSA 90-200 Å²), 3-6 hydrogen bond donors and 6-12 acceptors, reflecting common hydroxylation and lactone functionalities. The fraction of sp^3 carbons (Fsp^3) indicates a balance between steroidal rigidity and glycosylation-induced flexibility. Molecular weights cluster around 400-700 Da, with some exceeding 1,000 Da, highlighting extensive phenolic functionalization. The DOX distribution (20-30 range) highlights the role of oxidation in modulating bioactivity. The DOX metric is closest to a normal distribution, whereas HBD, HBA, and exact mass are right-skewed, and TPSA and Fsp^3 suggest possible bimodal distributions. High DOX-g cardenolides included the N,S cardenolides at the center of this study as well as structurally related compounds such as 16-hydroxyuscharin, and 2''-oxovoruscharin. Non- N,S-cardenolides with DOX-g ≥21 include diacetylated compounds (e.g., 16α- acetoxyasclepin and derivatives of calactinic acid. The most complex cardenolides (DOX-g ≥25) are those that incorporate phenolic-derived moieties and have been isolated from *A. linaria* and *A. subverticillata,* yet their biological activity remains unknown(*34*–*36*).

Figures

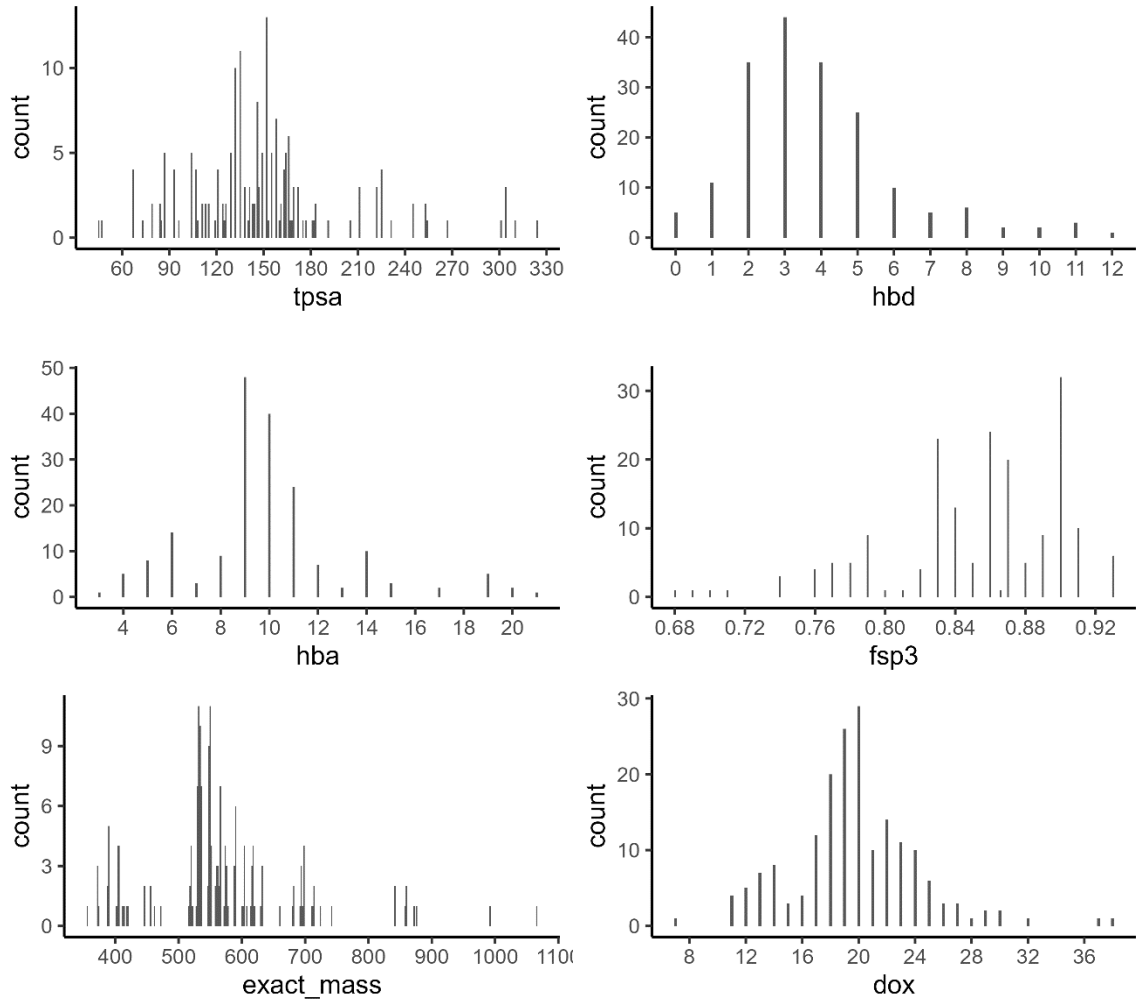


Fig. S1 Distribution of cardenolides reported in milkweed (*Asclepias* spp.) according to different metrics considered for chemical complexity.

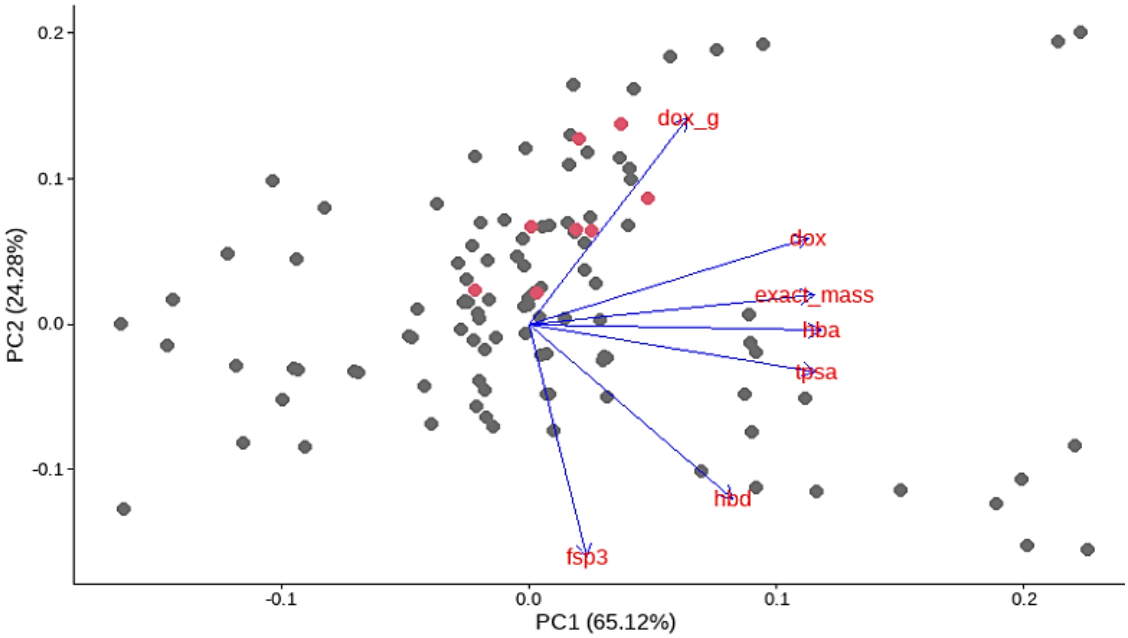


Fig. S2 Chemical space of structure complexity of milkweed cardenolides. Scores of the first and second principal components (PC1 and PC2, respectively) build from the values of different metrics obtained from 2d structures of reported cardenolides for *Asclepias* spp. N,S-cardenolides are present in the counts marked with red.

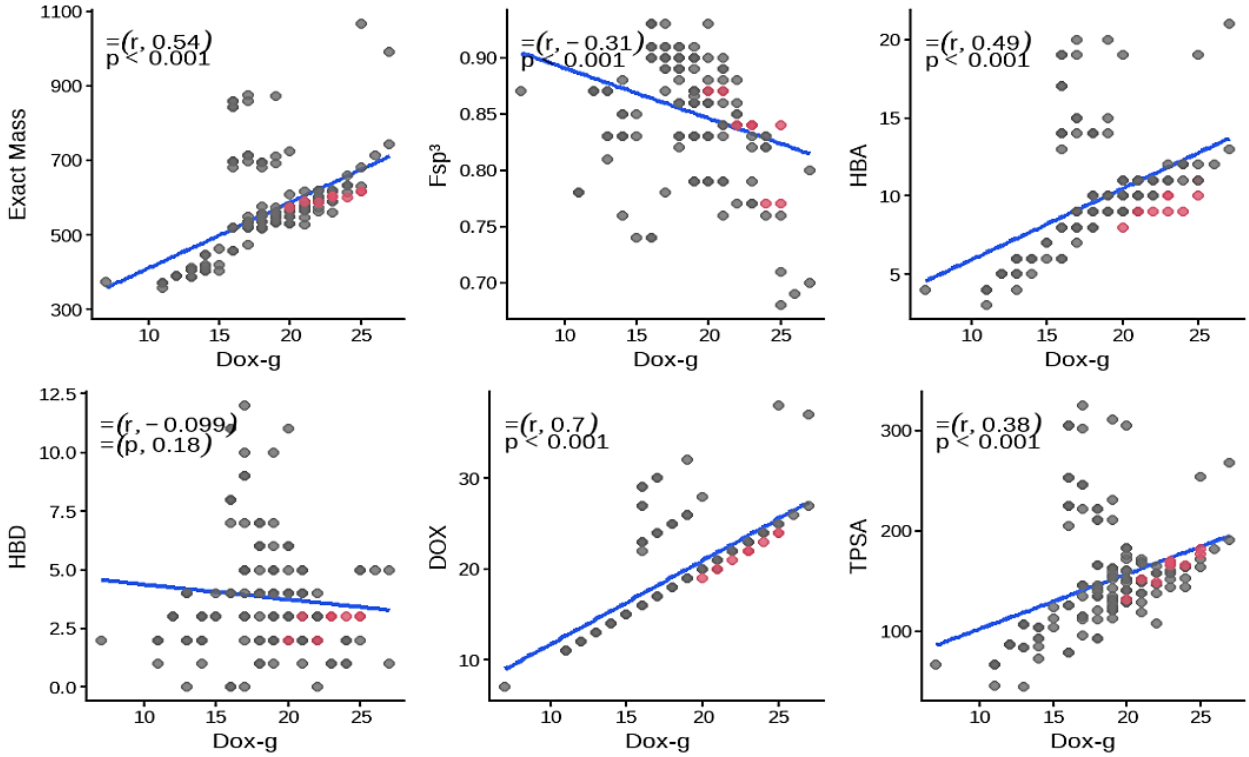


Fig. S3 Pearson's correlations between DOX-g and other chemical complexity metrics of reported milkweed cardenolides. N,S-cardenolides are present in the counts marked with red.

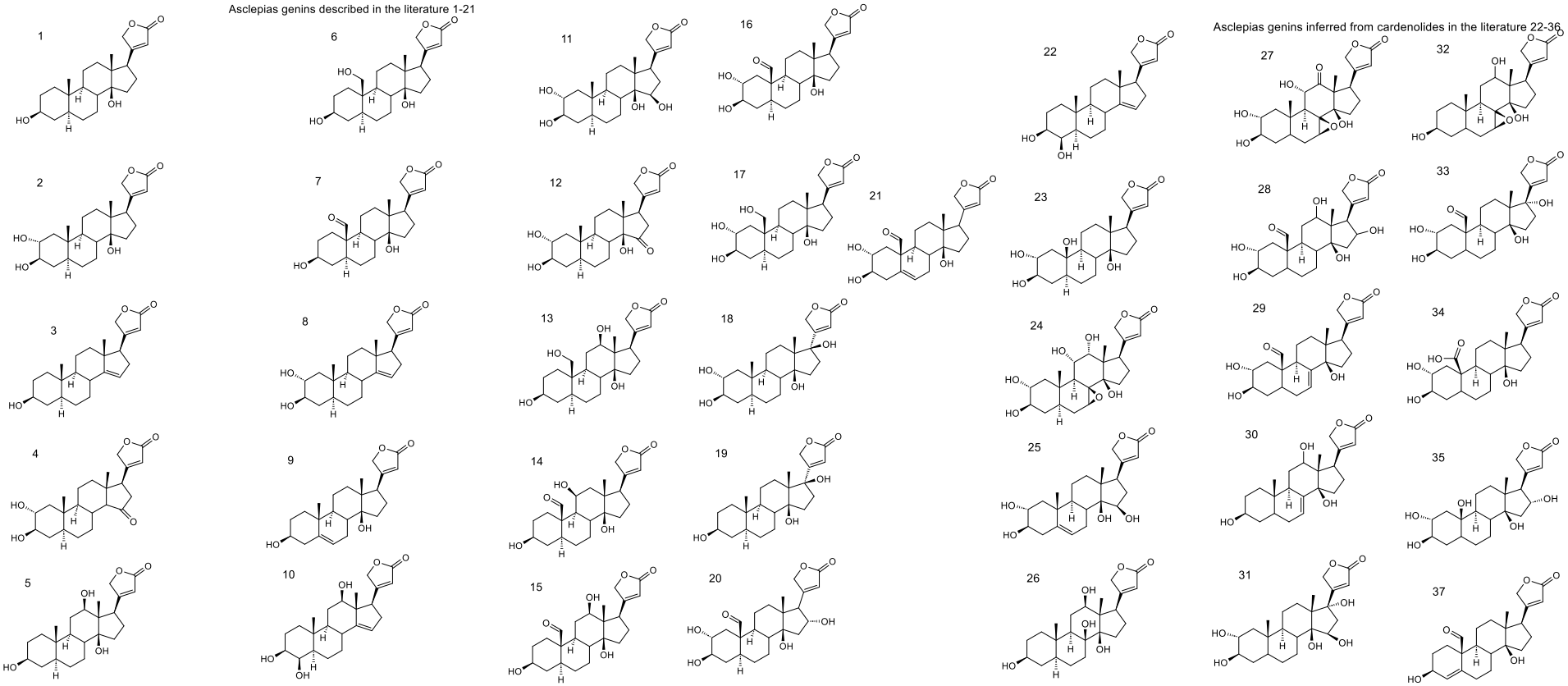
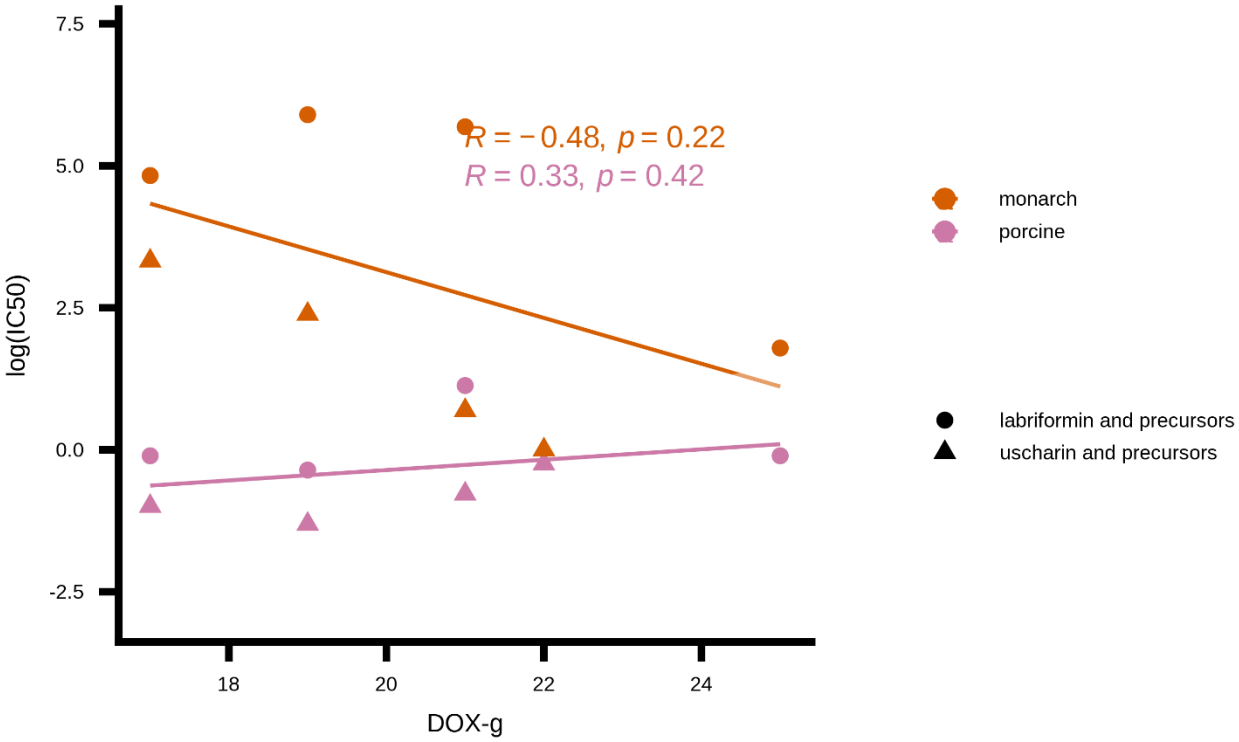


Fig. S4 Isolated genins and theoretical genins based on reported cardenolides from milkweed.



	Std.Error	DF	t-value	p-value
(Intercept)	4.522573	6	2.47	0.05
dox_g	0.222980	6	-1.80	0.12
Enzyme	4.684487	6	-2.85	0.03*
dox_g:enzyme	0.230963	6	2.14	0.07

Fig. S5 Pearson's correlation between DOX-g metric and the toxicity data for the N,S-cardenolides and biosythetical precursors. Data is parametric after log correction (KS test: n.s). Inset table shows linear mixed-effects model for DOX-g and log (IC_{50}) with enzyme as group, and compound as random intercept

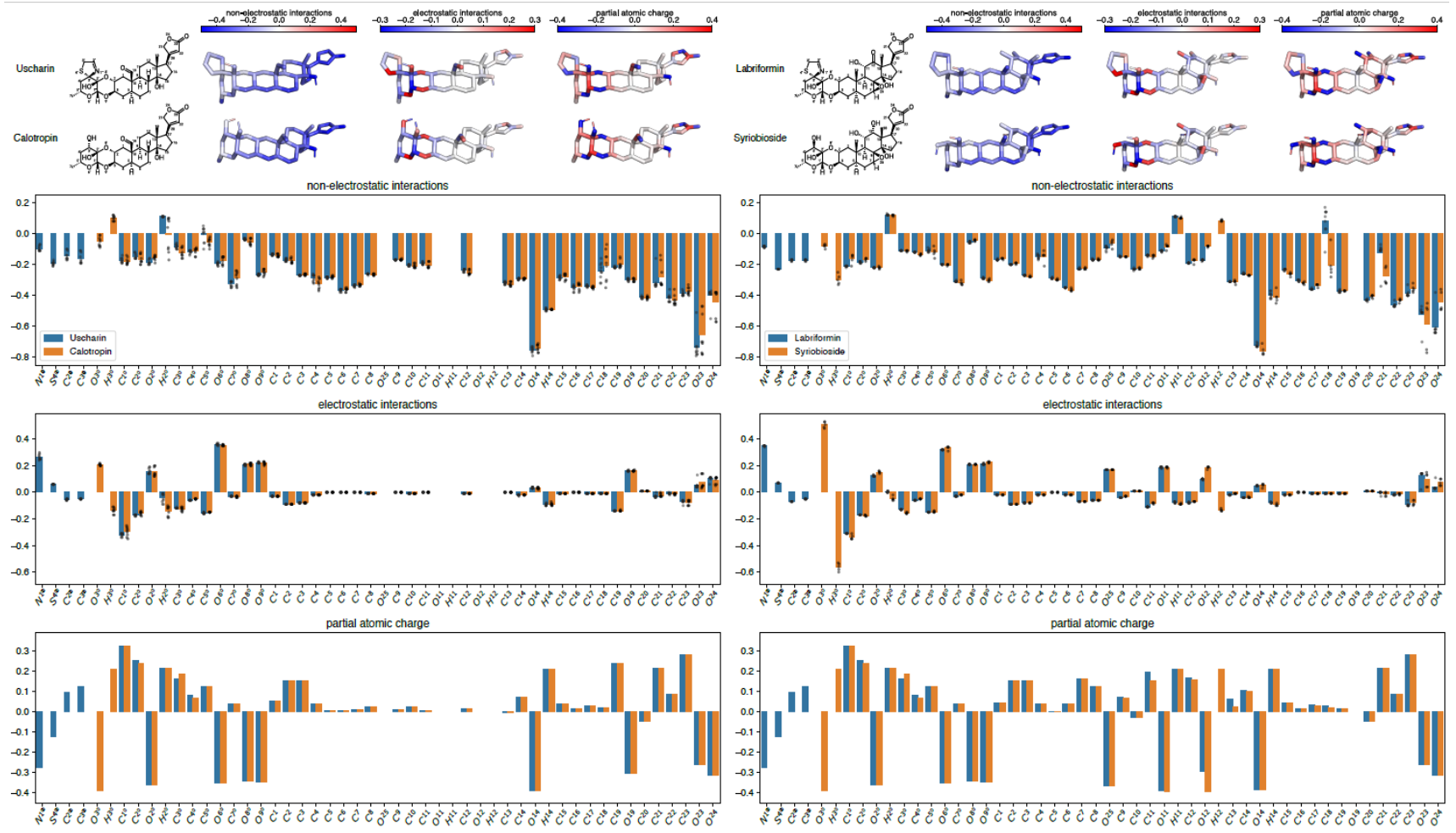


Fig. S6 Per-atom distributions of interaction terms and partial atomic charges. Based on AutoDock4 outputs, the distributions of non-electrostatic and electrostatic interaction terms, together with partial atomic charges, are plotted for each ligand. The conformations with the lowest predicted binding free energy among the ten docking simulations are shown above the plots as the representatives, together with the corresponding color-coded per-atom values. Atom labels on the x-axis match the numbered atoms in the structure shown at the top.

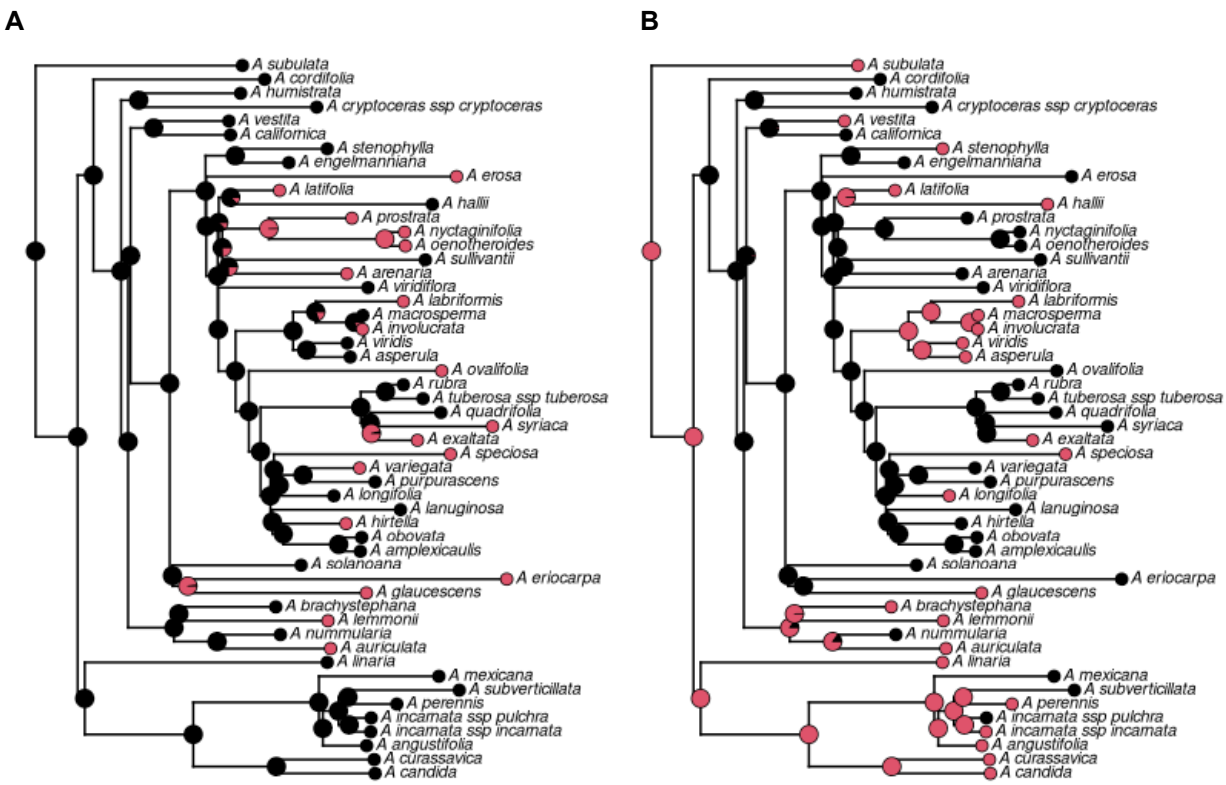


Fig. S7. Stochastic character mapping (SIMMAP) for the evolution of reduced labriformin (**A**) and voruscharin (**B**) the latex of *Asclepias* species. Observed states are shown at the tips; pie charts at internal nodes represent posterior probabilities of each state (black=absent, red=present), estimated using an equal-rates (ER) model and Bayesian MCMC sampling.

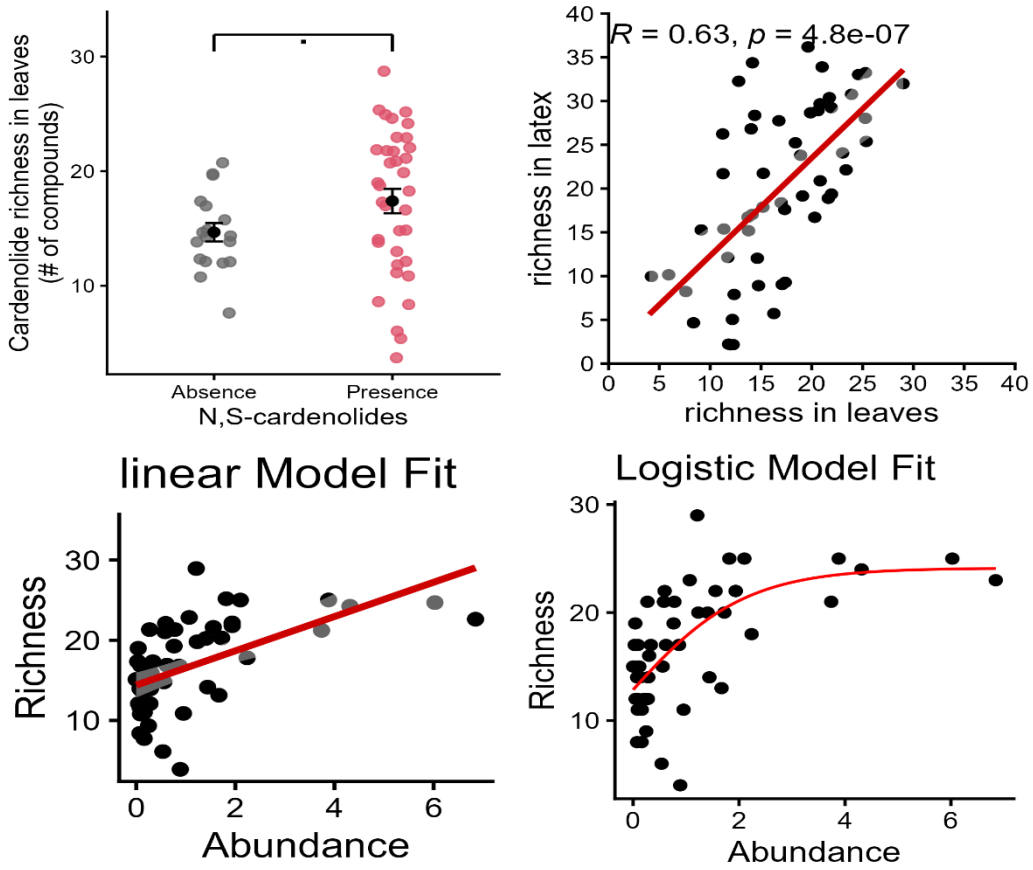
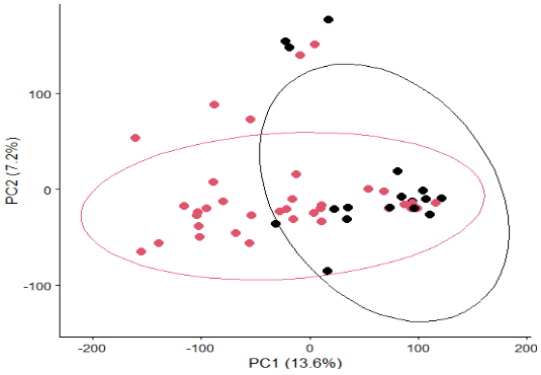
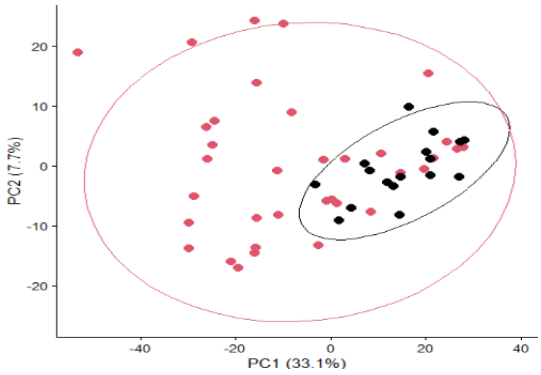


Fig. S8 Analysis of several metrics of cardenolide diversity. (**A**) Difference cardenolide richness in leaves between species that produce or not N,S-cardenolides (p-value: 0.09,r^2=0.03). (**B**) Correlations between cardenolide richness in leaves and latex: Levene's test no homogeneity ($F_{1,104}$ = 20.3, p < 0.0001***) Spearman correlation: p-value: < 0.0001 ***, r^2=0.67. F test to assess difference and direction of variance (alternative = "less"): $Ftest_{1,52}$ = 0.32, p< 0.0001***. (**C**) Correlation between leaves' cardenolide richness and abundance of milkweed leaves, linear model fit (p-value < 0.0001 ***, r^2 of 0.3. AIC: 315.36, BIC: 321.27). (**D**) Logistic model fit (phi1 φ1 estimate = 24.13, p < 0.001**, φ2 estimate = 0.13, p = 0.569, φ3 estimate = 0.91, p = 0.015*. AIC: 310.03, BIC: 317.91). ANOVA of the linear vs logistic model: p-value < 0.0001 ***.

PCA-All features



PCA-1200 features



PERMANOVA results

Ordination	Distance	R2	p_value
PCA (All Features)	Euclidean	0.04	0.001 ***
PCA (Top 1200 Features)	Euclidean	0.22	0.001***
Phylogenetic PCA Top 1200 Features	Euclidean	0.22	0.001***

Dispersion of Phylogenetic PCA Top 1200 Features of species with (P) and without (A) N,S cardenolides

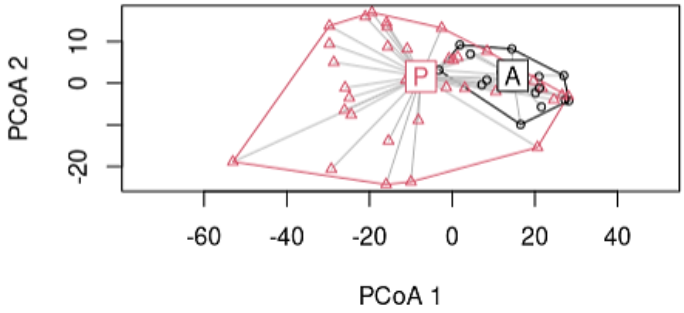


Fig. S9 Multivariate analysis of the latex metabolome of milkweed species that produce or not N,S-cardenolides These panels accompany Fig. 4c.

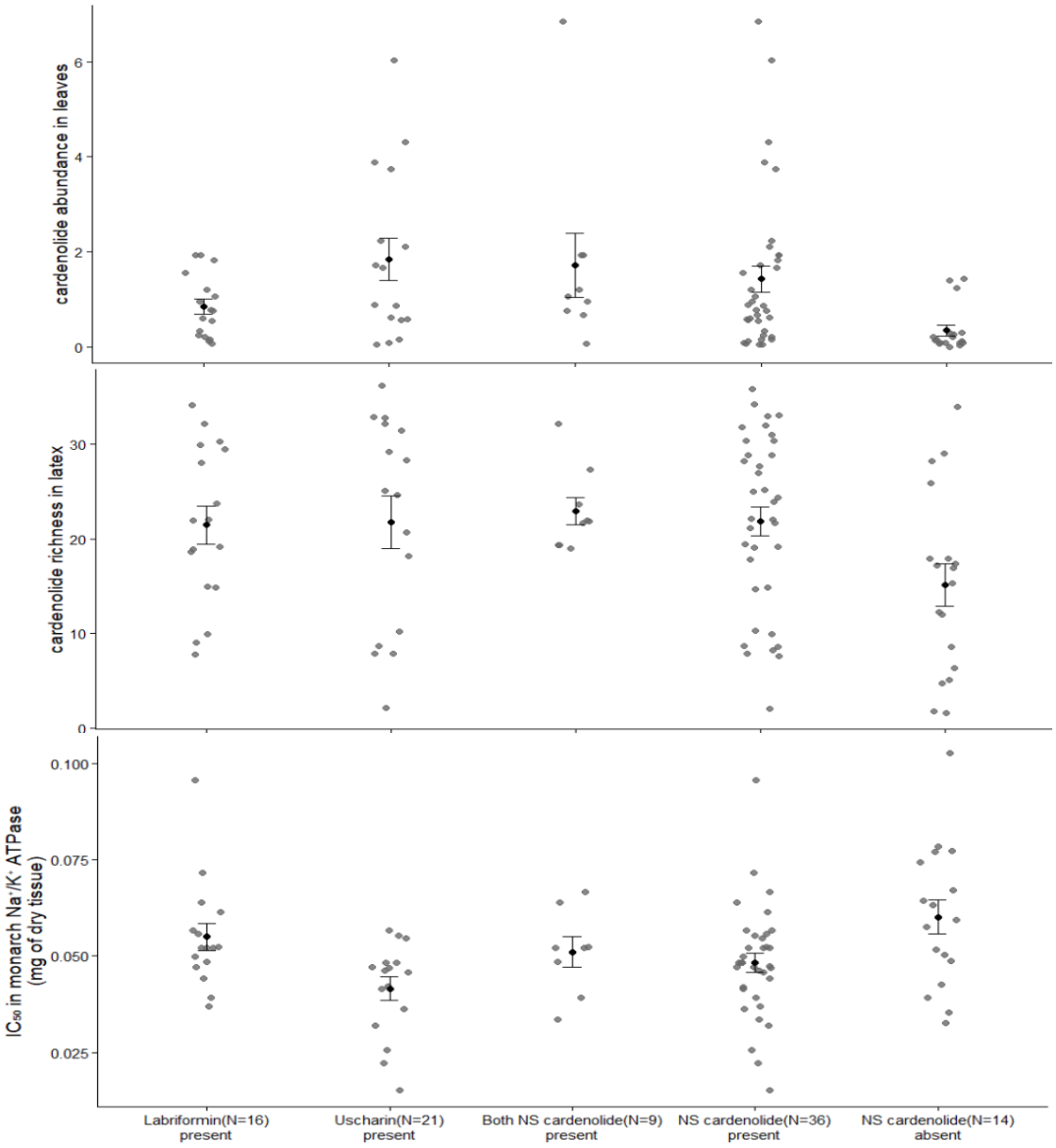


Fig. S10. Breakdown analysis of comparisons between species by presence or absence of specific N,S-cardenolides in fig 4 analysis comparison.

Table S1. Measures of structure complexity.

	MEASURE OF STRUCTURE COMPLEXITY	FORMULA/DEFINITION	DATA NEEDED*	INTERPRETATION
MW	Molecular weight	Molecular weight	Mass data	↑ number ↑complex
Fsp3	Fraction of sp3 carbons	#sp3 carbons/#total Carbons. Sp2 carbons indicate greater chemical reactivity.	2d structure	↓Lower number ↑complex
HBD	Hydrogen bond donor	Molecule or functional group that provides the hydrogen atom involved in the hydrogen bond e.j. –OH, NH	2d structure	↑number ↑ complex
HBA	Hydrogen bond acceptor	Molecule or functional group that accepts the hydrogen atom involved in the hydrogen bond e.j. –O, N	2d structure	↑ number ↑ complex
TPSA	Topological polar surface area	Summation of tabulated surface contributions of polar fragments (i.e. Atoms regarding also their bonding pattern)	2d structure	↑ number ↑ complex
DU	Degree of unsaturation	$\boldsymbol{DU = \frac{1}{2}(2 + 2(\#O_2) - (\#H_2) + (\#N_2) - (\#X)}$	Chemical formula	↑ number ↑ complex
p	Number of heteroatoms	# of heteroatoms (O,N,X,S)	Chemical formula	↑ number ↑ complex
DOX	Degree of oxidation	DU+p	Chemical formula	↑ number ↑ complex

*All the metrics can be obtained with SMILES—which can be generated for structures where planar structure is available (https://swissadme.ch/)

Table S2 PCA loadings for the different metrics associated with chemical complexity on the three main principal components (96% cumulative variation explained) with the values from 184 reported milkweed cardenolides.

	PC1	**PC2**	**PC3**
exact_mass	0.452	0.078	-0.006
fsp3	0.092	-0.627	-0.765
hbd	0.324	-0.468	0.392
hba	0.464	-0.017	0.012
tpsa	0.455	-0.127	0.151
dox	0.443	0.231	-0.017
dox_adjusted (DOX-g)	0.247	0.558	-0.488
Standard deviation	2.133	1.305	0.704
Proportion of Variance	0.650	0.243	0.071
Cumulative Proportion	0.650	0.894	0.964

Table S3 Presence of N,S cardenolides and their combinations by latex sample injected in LC-MS. Only *Asclepias* species with at least one N,S-cardenolide.

	Labriformin	Red Labriformin	Uscharin	Voruscharin	Labriformin Uscharin	Red Labriformin Voruscharin	Labriformin Voruscharin	Red Labriformin Uscharin	Voruscharin Uscharin	Red Labriformin Labriformin
viridis		89	154, 72, 89	154, 72, 89		89		89	154, 72, 89	
vestita			50, 6, 99	50, 6, 99					50, 6, 99	
variegata	116	116								116
syriaca	131, 160, 74	131, 160, 74								131, 160, 74
subulata			114, 155	114, 155, 85					114, 155	
stenophylla			59	59					59	
speciosa	159, 18, 42	159, 18, 42		159, 18		159, 18	159, 18			159, 18, 42
pulchra	144	144		104						144
prostrata	101	101								101
perennis			148, 53, 93	148, 53, 93					148, 53, 93	
ovalifolia	126, 167, 79	126, 167, 79								126, 167, 79
oenotheroides	9	44, 9								9
nyctaginofolia	19	19								19
macrosperma			68	68					68	
longifolia			83	83					83	
linaria			157, 62	157, 62					157, 62	
lemmonii		127, 143	127, 143, 78	127, 143, 78		127, 143		127, 143	127, 143, 78	
latifolia	111, 124, 73	111, 124, 73		124, 73		124, 73	124, 73			111, 124, 73
labriformis	122, 22	122, 22	22	122, 22	22	122, 22	122, 22	22	22	122, 22
involucrata		115, 2	2	115, 2		115, 2		2	2	
incarnata				137						
hirtella	75	75								75
hallii			12, 158, 55, 97	12, 158, 55, 97					12, 158, 55, 97	
glaucescens	21, 54	21, 54		21		21	21			21, 54
exaltata	170, 25	170, 25	125, 170, 25	125, 170, 25	170, 25	170, 25	170, 25	170, 25	125, 170, 25	170, 25
erosa	117, 169, 88	117, 169, 88		169, 88		169, 88	169, 88			117, 169, 88
eriocarpa	140, 46	140, 46								140, 46
curassavica			119, 28, 63	119, 28, 63					119, 28, 63	
cordifolia			171, 31							
candida		23	23, 96	23, 96		23		23	23, 96	
brachystephana			109, 37, 57	109, 37, 57					109, 37, 57	
auriculata	102, 161, 32, 33, 84	102, 161, 32, 33, 84	102, 161, 32, 33, 84	102, 161, 32, 33, 84	102, 161, 32, 33, 84	102, 161, 32, 33, 84	102, 161, 32, 33, 84	102, 161, 32, 33, 84	102, 161, 32, 33, 84	102, 161, 32, 33, 84
asperula			152, 40, 98	152, 40, 98					152, 40, 98	
arenaria	1, 100, 60	1, 100, 60								1, 100, 60
angustifolia			10, 166, 90	10, 166, 90					10, 166, 90	
amplexicaulis			11, 136							

Table S3 continued. After manual analysis for species where N,S cardenolides were detected in one replicate:

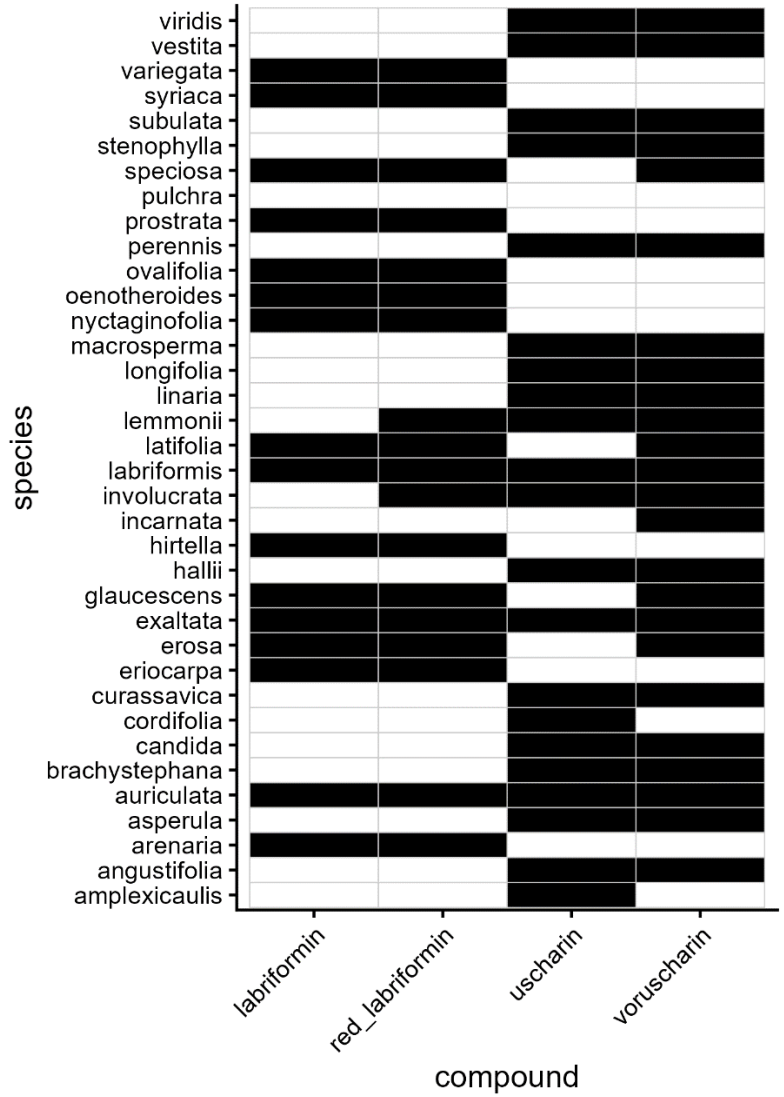


Table S4. Phylogenetic signal of individual N,S-cardenolides (δ value for discrete traits). Changes, gains, and losses were calculated based on ancestral state reconstructions from 1000 stochastic mapping iterations (mean and ± standard error) (See Fig. 2).

	Phylogenetic signal		**Ancestral state reconstruction (equal rates model, 1000 simulations)**		
	δ	p-value	Changes	Gains	Losses
N,S-cardenolides (all)	0.39	0.95	15.01±0.09	0.07±0.01	14.93±0.01
Labriformin	0.88	0.11	12.05±0.23	11.68±0.02	0.37±0.02
Reduced labriformin	0.58	0.69	14.06±0.24	13.49±0.02	0.56±0.02
Uscharin	0.60	0.02	13.01±0.07	8.63±0.04	4.37±0.04
Voruscharin	0.72	0.12	14.01±0.09	9.14±0.01	4.86±0.01

SI References

1. L. Rodríguez-Hahn, G. Fonseca, The cardenolide content of Asclepias linaria. *Phytochemistry* **30**, 3941–3942 (1991).
2. G. H. Robinson, G. E. Burrows, E. M. Holt, R. J. Tyrl, A. D. Jones, Investigation of the neurotoxic compounds in Asclepias subverticillata, western-whorled milkweed in *Proceedings of the International Symposium on Poisonous Plants*, (1998), pp. 435–439.
3. M. C. Roy, F. R. Chang, H. C. Huang, M. Y. N. Chiang, Y. C. Wu, Cytotoxic principles from the Formosan milkweed, Asclepias curassavica. *J. Nat. Prod.* **68**, 1494–1499 (2005).