New Phytologist

Research

Diversity and divergence: evolution of secondary metabolism in the tropical tree genus *Inga*

Dale L. Forrister[1], María-José Endara[1,2], Abrianna J. Soule[1], Gordon C. Younkin[3,4], Anthony G. Mills[1], John Lokvam[1], Kyle G. Dexter[5], R. Toby Pennington[6], Catherine A. Kidner[7,8], James A. Nicholls[9], Oriane Loiseau[5], Thomas A. Kursar[1] and Phyllis D. Coley[1]

[1]School of Biological Sciences, University of Utah, Aline W. Skaggs Biology Building, 257 S 1400 E, Salt Lake City, UT 84112-0840, USA; [2]Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud-BIOMAS – Universidad de las Américas, 170513 Quito, Ecuador; [3]Boyce Thompson Institute, Ithaca, NY 14853, USA; [4]Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA; [5]School of Geosciences, University of Edinburgh, Old College, South Bridge, Edinburgh, EH8 9YL, UK; [6]Department of Geography, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4QE, UK; [7]School of Biological Sciences, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh, EH9 3JW, UK; [8]Royal Botanic Gardens Edinburgh, 20a Inverleith Row, Edinburgh, EH3 5LR, UK; [9]The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australian National Insect Collection (ANIC), Building 101, Clunies Ross Street, Black Mountain, ACT 2601, Australia

Author for correspondence:
Dale L. Forrister
Email: dale.forrister@utah.edu

Received: *17 August 2022*
Accepted: *29 September 2022*

New Phytologist (2023) **237:** 631–642
doi: 10.1111/nph.18554

Key words: chemical defense, evolution, *Inga*, metabolomics, phytochemical diversity, plant specialized metabolism, secondary metabolism, UPLC-MS.

Summary

- Plants are widely recognized as chemical factories, with each species producing dozens to hundreds of unique secondary metabolites. These compounds shape the interactions between plants and their natural enemies. We explore the evolutionary patterns and processes by which plants generate chemical diversity, from evolving novel compounds to unique chemical profiles.
- We characterized the chemical profile of one-third of the species of tropical rainforest trees in the genus *Inga* (*c.* 100, Fabaceae) using ultraperformance liquid chromatography-mass spectrometry-based metabolomics and applied phylogenetic comparative methods to understand the mode of chemical evolution.
- We show: each *Inga* species contain structurally unrelated compounds and high levels of phytochemical diversity; closely related species have divergent chemical profiles, with individual compounds, compound classes, and chemical profiles showing little-to-no phylogenetic signal; at the evolutionary time scale, a species' chemical profile shows a signature of divergent adaptation. At the ecological time scale, sympatric species were the most divergent, implying it is also advantageous to maintain a unique chemical profile from community members; finally, we integrate these patterns with a model for how chemical diversity evolves.
- Taken together, these results show that phytochemical diversity and divergence are fundamental to the ecology and evolution of plants.

Introduction

For sessile organisms such as plants, secondary metabolism plays a fundamental role in mediating biotic interactions ranging from mutualisms (e.g. pollination) to antagonisms (e.g. competition and defense). Plant secondary metabolites, sometimes referred to as specialized metabolites, which are classically considered nonessential for basic cellular function, are exceedingly diverse, with nearly 1000 000 predicted to exist across the plant kingdom (Afendi *et al.*, 2012). It has long been thought that this incredible diversity strongly influences the ecology and evolution of interactions between plants and their pests and pathogens (Ehrlich & Raven, 1964; Endara *et al.*, 2017, 2018b, 2021). Plant secondary metabolites are also essential for plants' ability to survive in harsh abiotic environments by offering protection from UV damage and desiccation (Weng, 2014). The evolution of novel compounds or unique combinations of compounds (hereafter, chemical profile) can be highly adaptive, increase plant fitness, and facilitate species coexistence (Salazar *et al.*, 2016; Vleminckx *et al.*, 2018; Forrister *et al.*, 2019). Thus, understanding the origin and maintenance of chemical diversity is central to both the evolution and the ecology of plants.

Much of the theoretical and empirical literature supports the idea that selection has placed a premium on chemical diversity in plants (Jones *et al.*, 1991; Berenbaum & Zangerl, 1996; Richards *et al.*, 2016; Kessler & Kalske, 2018; Salazar *et al.*, 2018; Wetzel & Whitehead, 2019). A species' chemical profile is thought to arise from a diverse set of selective pressures ranging from abiotic factors, such as water loss and solar radiation, as well as selection exerted by a multitude of herbivores, pathogens, and mutualists (Weng, 2014; Endara *et al.*, 2017; Salazar *et al.*, 2018). For example, increased phytochemical diversity in tropical forests is

negatively correlated with both the number of herbivore species associated with a given host (Salazar *et al.*, 2018; Endara *et al.*, 2022) and herbivory (Richards *et al.*, 2015). In addition to producing a diverse set of compounds, recent studies have highlighted the importance of a given species to maintain a unique chemical profile relative to other species in its community (Kursar *et al.*, 2009; Forrister *et al.*, 2019; Endara *et al.*, 2021). While there is a clear consensus on the value of phytochemical diversity, the underlying evolutionary processes that generate chemical diversity in plant lineages remain widely debated (Wetzel & Whitehead, 2019).

Here we ask how plants generate chemical diversity and what evolutionary processes lead to novel compounds and unique chemical profiles. To address this question, we build on the classic 'escape and radiate' theoretical framework, first suggested a half-century ago by the work of Dethier (1954), Fraenkel (1959), and Ehrlich & Raven (1964). In this model, random mutations in biosynthetic genes lead to the production of novel defense compounds, often through the gradual embellishment of core structures into more complex and derived compounds (Berenbaum & Feeny, 1981; Berenbaum, 1983; Coley *et al.*, 2019). If these derived compounds have stronger deterrent properties or are effective against different enemies, selection acts to promote the novel genotype. In this study, we test the prediction put forth by the 'escape and radiate model' that chemical evolution proceeds in a gradual stepwise manner through the modification of core structures (Ehrlich & Raven, 1964; Berenbaum, 1983). To test this, we combine untargeted metabolomics and comparative phylogenetic methods to characterize the chemical profiles for nearly 100 species of tropical trees in the genus *Inga Mill* (Fabaceae). By focusing on a recently radiated monophyletic genus of trees, we attempt to understand how chemistry evolves at the tips of the phylogenetic tree over a relatively short period of evolutionary history. This offers a different perspective to studies of chemical evolution focused on deeper phylogenetic scales such as divergence among families (e.g. Wink, 2003).

Inga is a useful case study for exploring how secondary metabolism evolves over short phylogenetic distances. *Inga* is a speciose genus with *c.* 300 tree species in tropical moist forests throughout the New World. At any given site, it usually constitutes one of the most abundant and speciose genera, with up to 40 coexisting species (Valencia *et al.*, 2004). Multiple lines of evidence have implicated the importance of chemistry in the ecological and evolutionary processes that have shaped the genus (Kursar *et al.*, 2009; Endara *et al.*, 2017; Coley *et al.*, 2018). Moreover, *Inga* and other speciose tropical genera such as *Bursera*, *Psychotria*, *Piper*, and *Protium* are among the most phytochemically diverse plant lineages that have been documented, often having more compounds in a single genus than entire plant communities in temperate ecosystems (Sedio *et al.*, 2018). Thus, *Inga* is an illustrative model for the generation of phytochemical diversity as a whole. The results presented in this study build on previous work in *Inga*, which focused on a few specific metabolites (Coley *et al.*, 2019) or broad compound classes (Kursar *et al.*, 2009). Here we increase the phylogenetic coverage and leverage metabolomics to greatly expand our exploration of the relationship between evolutionary history and chemical similarity.

We use untargeted metabolomics to quantify intraspecific phytochemical diversity, examine how chemical similarity between congeners changes over evolutionary time and geographic distance, and finally quantify the phylogenetic signal of individual compounds and larger chemical classes. In doing so, we aim to address the following questions and hypotheses:

(1) Do species invest in phytochemical diversity by producing structurally unrelated compounds? Investment in structurally diverse defensive compounds is adaptive for protection against a broad suite of pests and pathogens (Salazar *et al.*, 2018; Wetzel & Whitehead, 2019; Endara *et al.*, 2022), yet investment in chemical defense comes at a cost (known as the 'growth-defense tradeoff') (Strauss *et al.*, 2002; Monson *et al.*, 2021; Panda *et al.*, 2021). Investment in chemical defense is expensive both in terms of the carbon and nitrogen used as inputs for the biosynthetic products and in terms of transcribing and regulating enzymes involved in secondary metabolism (Gershenzon, 1994). It is unclear whether biosynthetic constraints and pleiotropy of biosynthetic enzymes limit phytochemical diversity or lead to evolutionary trade-offs between chemical classes (Koricheva *et al.*, 2004; Agrawal *et al.*, 2009; Gershenzon *et al.*, 2012). Because phytochemical diversity is potentially adaptive (Richards *et al.*, 2015; Salazar *et al.*, 2018; Endara *et al.*, 2022), we hypothesize that selection will favor investment in a diverse suite of compounds rather than structurally related ones.

(2) Does the entire chemical profile diverge between closely related species and does it evolve under divergent selection? The 'escape and radiate' model predicts that closely related species would have similar defensive profiles (Ehrlich & Raven, 1964; Berenbaum & Feeny, 1981; Berenbaum, 1983; Coley *et al.*, 2019). However, it has also been posited that diffuse coevolution between plants and their natural enemies would result in divergent adaptation in defense traits (Endara *et al.*, 2015; Maron *et al.*, 2019). The latter argues that it is advantageous for a species to not only have a diversity of compound classes but also to be different from other species in their community in order to not share pests and pathogens (Kursar *et al.*, 2009; Bagchi *et al.*, 2014; Salazar *et al.*, 2018; Forrister *et al.*, 2019). Here we ask whether species' chemical profiles show phylogenetic signal, or whether they have diverged sufficiently to erase the effect of shared evolutionary history. We also incorporate biogeography, asking whether sympatric species are more or less divergent in their chemical profile than species occurring in parapatry. Biogeography is an important factor because at the population (within species) level, selection pressures may differ at different sites. Additionally, because sympatric species should be divergent in ecologically relevant traits to coexist (Chesson, 2000), we hypothesize that sympatric relatives will be more divergent in their chemical profile than parapatric ones. Finally, we use a novel modeling framework (Anderson & Weir, 2020) to formally test the hypothesis that chemical profiles are evolving under divergent adaptation.

(3) Are individual compounds phylogenetically conserved? The evolution of novel chemistry is assumed to be the result of stepwise changes in chemical structures resulting in more derived chemical defenses over evolutionary time (Berenbaum & Feeny, 1981; Coley *et al.*, 2019). This process should lead to a pattern of phylogenetic conservatism of metabolites and biosynthetic pathways (Ehrlich & Raven, 1964; Salazar *et al.*, 2018). To test this prediction, we mapped all individual compounds present in *Inga* onto the phylogeny and estimated their phylogenetic signal. We then used ancestral state reconstruction to estimate the number of times each compound had transitioned on the phylogenetic tree (Courtois *et al.*, 2016). In contrast to the 'escape and radiate' model, we hypothesize that in order for species to invest in structurally diverse compounds and diverge from close relatives, the mode of chemical evolution would not proceed in a stepwise manner. Rather, rapid changes based on transcriptional regulation would result in low phylogenetic signal of individual compounds.

(4) Is there evidence of metabolic integration or apparent trade-offs between biosynthetic pathways? Comparative phylogenetic analyses of defense traits have revealed both trade-offs (negative correlations) (Kursar & Coley, 2003; Agrawal & Fishbein, 2006; Agrawal *et al.*, 2009; Coley *et al.*, 2018; Monson *et al.*, 2021) and positive correlations (Agrawal & Fishbein, 2006), providing evidence for evolutionary integration and defense syndromes. For example, trade-offs between compound classes that share the same biosynthetic precursor are well-supported in the literature (Keinänen *et al.*, 1999; Nyman & Julkunen-Tiitto, 2005; Agrawal *et al.*, 2009). Nevertheless, other studies have found little evidence for these trade-offs based on meta-analysis (Koricheva *et al.*, 2004). Here we ask whether biosynthetic constraints lead to trade-offs that persist over evolutionary time scales or whether each branch of the biosynthetic pathway evolves independently.

Materials and Methods

Study sites and species sampling

We studied *Inga* between 2005 and 2014 at five lowland tropical rainforest sites across the Amazon basin and in Panama (Supporting Information Table S1), where we extensively surveyed understory saplings, a prolonged and key vulnerable stage in the life cycle of tropical forest trees (Coley *et al.*, 2018). We sampled *Inga* across the full distributional range of the genus. We spent *c.* 16 people-months per site collecting data in the field. Specifically, we exhaustively searched each site for all *Inga* species, taking measurements on morphological and defense traits for a total of 97 species and one species from its sister genus, *Zygia*. Species delimitation was based on the combination of morphology, phylogenetic reconstruction (Nicholls *et al.*, 2015), and in some cases for morphologically difficult to identify individuals, we relied on chemocoding to confirm species identifications (Endara *et al.*, 2018a). Young leaves at *c.* 50% full expansion were collected in the understory from 5 to 10 spatially separated individuals (with very few exceptions for rare species where we included three individuals). We focused on expanding leaves, as they receive >70% of the lifetime damage of a leaf (Coley & Aide, 1991), and their chemical profiles are an important factor for host associations of insect herbivores (Endara *et al.*, 2017, 2018b, 2021). In general, we found the chemical profile of each species to be highly canalized, and previous work has shown that five individuals are sufficient to capture *c.* 75% of compounds encountered in up to 15 individuals (Endara *et al.*, 2021). Samples were dried in the field at ambient temperature *in silica* immediately following collection, and then stored at −20°C.

Characterization of *Inga* chemistry

Soluble secondary metabolites Metabolites were extracted from dried leaf samples in the Coley/Kursar laboratory at the University of Utah using a solution of (60 : 40, v/v) (ammonium acetate buffered water), pH 4.8: acetonitrile, producing 2 ml of retained supernatant from 100 mg (±2.5 mg) of sample for chromatographic analysis following the UPLC-MS methods developed in Wiggins *et al.* (2016). Extraction weight (percent dry weight (DW)) was measured gravimetrically by subtracting dry marc from the mass of pre-extraction plant material. Small molecules (detector range of 50–2000 Da) from the extraction supernatant were analyzed using ultraperformance liquid chromatography (UPLC-MS) (Waters Acquity I-Class, 2.1 × 150 mm BEH C18 and 2.1 × 100 mm BEH Amide columns) and mass spectrometry (Waters Xevo G2 QToF, Waters, Milford, MA, USA) in negative ionization mode. A 45-min reverse-phase gradient was used for the C18 column with water (0.1% formic acid) as the mobile phase and acetonitrile (0.1% formic acid) as the stationary phase, flow rate was 0.5 ml min^{-1}, and column temperature was 40°C (46). For the amide column, we used regular phase chromatography starting with 95% acetonitrile (+0.1% formic acid) and 5% water (+0.1% formic acid). We used a linear gradient over 12 min ending with 30% acetonitrile (+0.1% formic acid). MS/MS spectra were acquired by running DDA, whereby MS/MS data were collected for all metabolites that ionized above a set threshold (5000 TIC).

L-Tyrosine Some *Inga* species invest in the overexpression of the essential amino acid L-tyrosine as an effective chemical defense (Coley *et al.*, 2019). Tyrosine is insoluble in our extraction buffer, so a different protocol was used to determine the percentage of leaf DW. Extractable nitrogenous metabolites were extracted from a 5 mg subsample of each leaf using 1 ml of aqueous acetic acid (pH 3) for 1 h at 85°C (Coley *et al.*, 2019). Fifteen microliters of the supernatant was injected into a 4.6 × 250 mm amino-propyl high-performance liquid chromatography (HPLC) column (Microsorb 5u; Varian, Palo Alto, CA, USA). Metabolites were chromatographed using a linear gradient (17–23%) of aqueous acetic acid (pH 3.0) in acetonitrile over 25 min. The mass of solutes in each injection was measured using an evaporative light scattering detector (ELSD) (Sedere SA, Alfortville, France). ELSD temperature was 75°C with 2.2 bars of compressed N_2, and instrument gain was set to 6. Tyrosine concentrations were determined by reference to a four-point standard curve (0.2–3.0 mg tyrosine ml^{-1}, $r^2 = 0.99$) prepared from pure tyrosine.

Compound separation, annotation, and assignment to species Following HPLC and UPLC-MS data acquisition, metabolites were quantified and assigned available structural information in all samples using an untargeted metabolomics pipeline developed by our research group (see Endara *et al.*, 2021 for details). In this pipeline, spectral features are extracted from raw MS data, and related features are grouped into compounds based on shared retention time and correlated abundance between scans using CAMERA (Kuhl *et al.*, 2012). We employed a variety of techniques in order to assign individual compounds into classes including NMR structural characterization, MS/MS-based spectral library searches using GNPS (Wang *et al.*, 2016), *in silico* compound annotation, and machine learning prediction. As a result, MS/MS data for each compound were uploaded to GNPS for annotation of putative structures and compound classes. These analyses generate: a species-by-compound abundance (MS-1 peak intensity measured by total ion current) matrix; a compound-by-compound MS/MS spectral cosine similarity matrix, which are then combined into a pairwise species similarity matrix, which accounts for both shared compounds between species and the MS/MS structural similarity of unshared compounds; a classification table is created with the assignment for all annotated compounds based on CLASSYFIRE (Djoumbou Feunang *et al.*, 2016). All code for this pipeline is deposited in a Git repository (https://gitlab.chpc.utah.edu/01327245/evolution_of_inga_chemistry).

Indices for chemical similarity and phytochemical diversity To test for phylogenetic signal of the entire chemical profile and quantify divergence between species, we developed a method for quantifying overall chemical similarity between two species (Endara *et al.*, 2021). This provides a challenge because few compounds are shared between species, making classic distance metrics such as Bray–Curtis uninformative (Sedio *et al.*, 2017; Endara *et al.*, 2021). Our method, which is similar to the method developed by Sedio *et al.* (2017), accounts for the fact that two species may have different compounds that are structurally similar (Endara *et al.*, 2018b, 2021). Specifically, we leverage MS/MS spectra as a proxy for the structural similarity between compounds (Wang *et al.*, 2016). In this method, total chemical similarity between species is a function of the normalized abundance of shared compounds plus the normalized abundance of unshared compounds weighted by their structural similarity in the molecular network (see (18) for details).

We quantified investment in phytochemical diversity for each focal species using its chemical profile and the MS/MS molecular network to calculate the functional Hill number (Chao *et al.*, 2014). This diversity measure accounts for both variations in compound abundance and structural similarity in the molecular network. In short, it calculates the effective number of equally abundant and structurally distinct compounds produced by a given species (Chao *et al.*, 2014). We compared this diversity index with a null model where we assembled compounds into chemical profiles through a bifurcating process from root to tip on the *Inga* phylogenetic tree. This null model is rooted in the null models often employed in community ecology but is expanded to incorporate phylogenetic relatedness. The null model represents the chemical profiles randomly drawn from the entire pool of compounds found in our study samples while controlling for evolutionary history, compound frequency, and abundance (see Methods S1 for a detailed explanation of the null model). To make a representative null model, we matched the number of compounds produced by a given species and the number of compounds shared between any two closely related species with the values observed in the actual data while randomizing the structural relatedness of shared compounds. We normalized phytochemical diversity values of each species relative to our null model.

Phylogenetic reconstruction of *Inga*

A phylogenetic tree containing 165 *Inga* accessions, including taxa sampled at multiple sites, was reconstructed using a newly generated targeted enrichment (HybSeq) dataset of 810 genes. These 810 loci include those presented by Nicholls *et al.* (2015), supplemented with a subset of the loci from work by Koenen *et al.* (2020). DNA library preparation, sequencing, and the informatics leading to final sequence alignments follow protocols in Nicholls *et al.* (2015). For the phylogenetic inference, we accounted for the putative effect of incomplete lineage sorting by constraining the maximum likelihood phylogeny with the topology obtained from a coalescent-based method. First, we inferred gene trees for 810 loci using IQTREE 2 (Minh *et al.*, 2020). The best substitution model was estimated for each loci using the MODELFINDER (Kalyaanamoorthy *et al.*, 2017) module implemented in IQTREE 2. For each gene tree, we performed 1000 bootstrap replicates with the ultrafast bootstrap approximation (Hoang *et al.*, 2018). The resulting gene trees were subsequently used as the input for ASTRAL-III to estimate a phylogeny in a summary coalescent framework (Zhang *et al.*, 2018), after contracting branches with bootstrap support < 10. We then used the topology obtained with ASTRAL to perform a constrained maximum likelihood tree search in IQTREE 2. We performed a partitioned analysis (Chernomor *et al.*, 2016) after inferring the best-partition scheme for the 810 genes and the best substitution model for each partition using MODELFINDER. Branch support was estimated with ultrafast bootstrap approximation (1000 replicates). The phylogenetic tree was subsequently time-calibrated using penalized likelihood implemented in the program TREEPL (Smith & O'Meara, 2012). We used cross-validation to estimate the best value of the smoothing parameter and implemented secondary calibration points on the crown and node ages of *Inga* with an interval of 9.2–11.9 and 13.4–16.6 Myr, respectively. Finally, the complete phylogeny was pruned to include only the 98 species for which chemistry data were available.

Phylogenetic comparative methods and ancestral state reconstruction

For phylogenetic signal of continuous traits, we calculated Blomberg's K (Blomberg *et al.*, 2003) using function *phylosignal* in the R package PICANTE v.1.8.2 (Kembel *et al.*, 2010). K is close to zero for traits lacking phylogenetic signal, and higher than 1 when close relatives are more similar than expected under the classic

Brownian motion evolutionary model. For the presence and absence of individual compounds, we calculated the *D*-statistic (Fritz & Purvis, 2010) using the CAPER package (Orme, 2013).

We took a stochastic character mapping approach for the ancestral state reconstruction of compound presence/absence on the *Inga* phylogeny. Specifically, we used the function *make.simmap* (Bollback, 2006) from R package PHYTOOLS v.0.7-47 (Revell, 2012) to estimate the state of each internal node on the phylogeny using 100 simulated trees. Based on the ancestral state reconstruction of each compound, we created an index of evolutionary lability, calculated as the number of times a given compound transitioned between present and absent divided by the number of species where a compound is present. Low values for this index indicate strong phylogenetic conservatism, where a compound likely evolved few times and was retained within a given lineage. Values near or above 1 indicate that a compound is evolutionarily labile, having been gained or lost as many times as the compound was present.

To model how the complete chemical profile changes over time, we used a modeling framework developed by Anderson & Weir (2020), which uses simulated trait values based on either Brownian motion or Ornstein–Uhlenbeck. This framework also tests for divergent adaptation by adding a term for the interactions between lineages during simulated trait evolution.

Results

Our untargeted metabolomics pipeline (Endara *et al.*, 2021) allowed us to characterize thousands of individual compounds and determine the similarity of chemical profiles across species. In total, we observed 9105 unique compounds across 808 samples. *Inga* species invest substantial resources in soluble secondary metabolites, averaging 194 ± 103 (mean $\pm$ SD) unique compounds per species, and comprising $37 \pm 11\%$ (mean $\pm$ SD) of the expanding leaf's DW (Fig. S1). We were able to classify 42.5% of compounds, a substantial improvement from the 2.9% achieved from library matches alone (Fig. 1). Although our extraction and detection methods did not explicitly exclude primary metabolites, the vast majority of annotated compounds were assigned to secondary metabolites, specifically chemical classes that have been classically implicated in plant defense against pathogens and herbivores, including flavonoids and saponins. Similarly, given the scale of this study, it should be noted that a small fraction of the chemical compounds analyzed in the study are not likely to be found in planta, as they could be adducts, chemical artifacts, and decomposition products. The inclusion of said artifacts should not influence the general conclusions of this study because they are relatively rare.

Individual species invest in structurally diverse compounds

We asked whether biosynthetic trade-offs constrain a plant's ability to invest in structurally unrelated compounds (i.e. the cost of maintaining enzymes in multiple metabolic pathways) or whether selection promotes investment in chemical diversity. To answer this question, we quantified investment in phytochemical diversity using functional Hill numbers and compared these findings with a null model. For the majority (94%) of species, phytochemical diversity was within the range of values expected by our null model. The rest of the species exceeded that range (4%) or were underdispersed (2%) (Fig. 2). The rarity of species with lower phytochemical diversity than the null model indicates that all species invest in structurally diverse compounds.

Chemical profiles evolve under divergent adaptation

To test for phylogenetic signal of the entire chemical profile and quantify divergence between species, we developed a method for

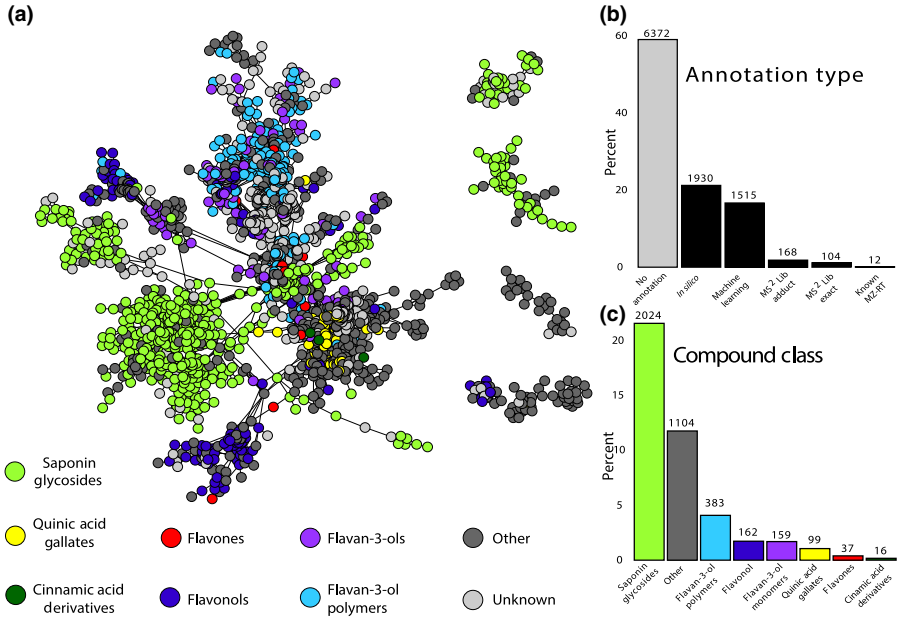

Fig. 1 Compound-based molecular network. (a) Subset of molecular network (see Supporting Information Fig. S2 for the full network) containing all compounds observed across 98 studied *Inga* species. Nodes represent individual compounds identified in the metabolomics pipeline, and connections between compounds (edges) are based on the MS/MS cosine similarity score from GNPS (https://gnps.ucsd.edu). (b) Percent of compounds that were annotated using different methods – *in silico* fragmentation, machine learning, MS/MS library exact matches and adducts, and comparison to authentic standards on our UPLC-MS system based on mass-charge ratio (*m*/*z*) and retention time (RT). (c) Percent of compounds with annotations represented by each compound class. For (b, c), the total numbers of compounds are reported at the top of bars.

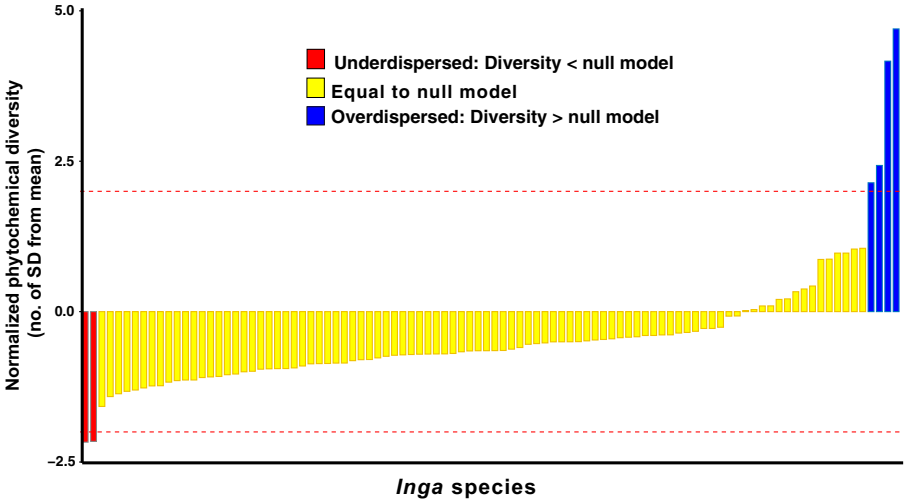

Fig. 2 Normalized phytochemical diversity in each *Inga* species. Bars represent individual *Inga* species ordered by increasing phytochemical diversity measured as the functional Hill number. Values represent the number of SDs above or below the mean calculated in the null model, with dashed red lines indicating two SDs above and below the null mean. Values < 0 represent species that are less chemically diverse than a random draw (under-dispersed in the MS/MS network) and values above zero represent species that are more diverse (over-dispersed in the MS/MS network). Hill numbers are calculated with $Q = 0$.

quantifying overall chemical similarity between two species (Endara *et al.*, 2021). We compared these calculations with estimates of chemical similarity expected from a null model (Methods S1). We found that chemical similarity was highest for intraspecific comparisons, but quickly decreased to the point where two species were as dissimilar as expected under our null model based on all interspecific comparisons (Figs 3, S3). Within a species, chemical similarity was highest between individuals at a single site but rapidly decreased between individuals of the same species at different sites (Fig. 3). We also found that interspecific chemical similarity was highly divergent even between sister species and that the majority (83%) of pairwise comparisons between species fell within the range of our null model (Fig. 3; Fig. S3). Sister species at different sites (parapatric) were divergent, and sympatric sister species were more divergent than parapatric sister species. Interspecific chemical similarity of the entire chemical profile showed no phylogenetic signal (Mantel test: $r = -0.03$, $P = 0.68$; Fig. S3).

To formally test the hypothesis that a species' chemical profile is evolving under divergent selection, we used recently developed phylogenetic comparative methods to model different modes of trait evolution and select the best-fitting model. We found strong support for the divergent adaptation model over models that assume all lineages evolve independently of others on a tree (i.e. Divergent vs Brownian motion and the Ornstein–Uhlenbeck process) (Table S2). Our results show that each species evolves to have a unique chemical profile compared with close relatives. Unlike a species' chemical profile, we found that traits related to the amount of chemical investment (number of compounds, gravimetric chemical investment, and phytochemical diversity; Fig. S1) were best explained by an Ornstein–Uhlenbeck process model, indicating that these traits are evolving toward an optimal trait value (Table S2) rather than diverging.

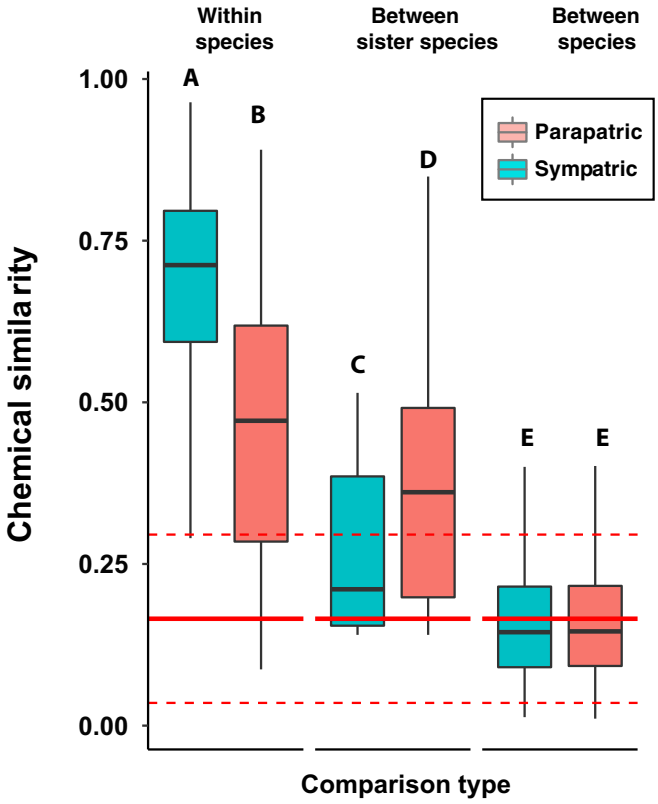

Fig. 3 Comparison of entire chemical profiles between *Inga* species. Boxplot comparison of chemical similarity scores for *Inga* within a species, between sister species, and between all other species. Boxplots indicate the mean value (bold line) and upper and lower quartile. Comparisons between and within sites are indicated by red and blue boxes, respectively. Significantly different groups are denoted by A, B, C, D, and E (ANOVA). The solid red line indicates the mean chemical similarity score observed in the null model, which simulates the expected chemical similarity between two randomly assembled chemical profiles. The dashed red lines represent two SDs above and below the null mean.

Many compounds showed no phylogenetic signal and were evolutionarily labile

The majority of compounds are detected in only a few species (median = 4), and roughly half (53%) of the compounds showed no phylogenetic signal (Fig. 4a). Although some compounds are clustered in specific clades, many compounds are found dispersed across the phylogeny (Fig. 4b). We found that the majority of

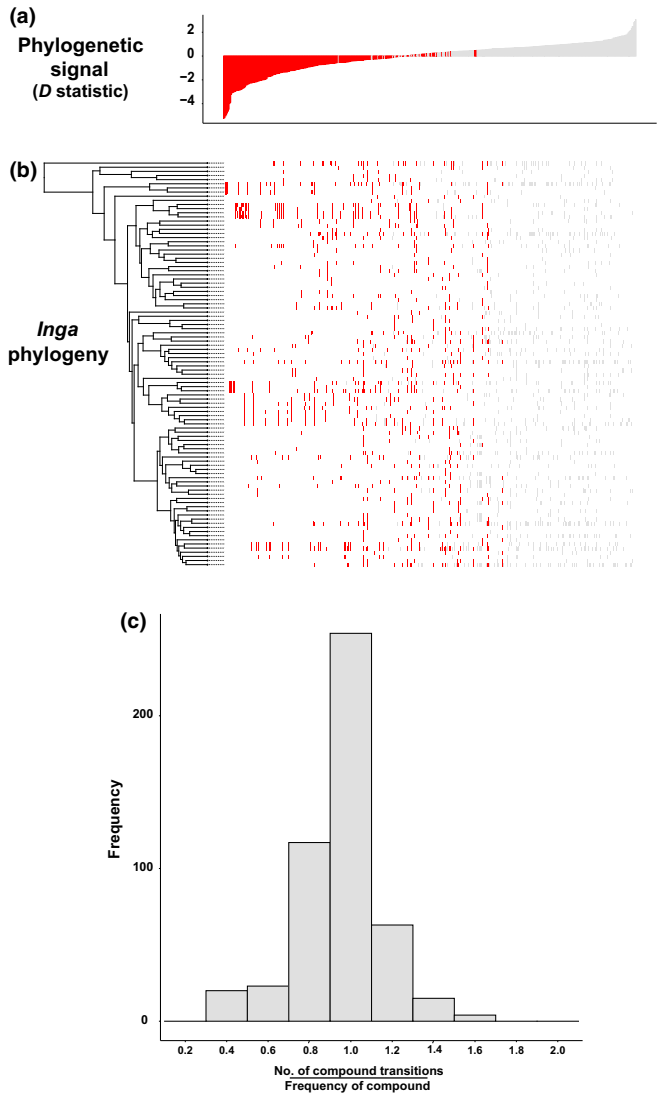

Fig. 4 Expression patterns of individual compounds mapped on to the *Inga* phylogeny. (a) Phylogenetic signal for 500 randomly sampled compounds ordered from most-to-least phylogenetically conserved using the *D*-statistic. For visualization purposes, we display 500 randomly chosen compounds. Red bars indicate compounds with significant phylogenetic signal ($P < 0.05$). (b) Heat map demonstrating expression of individual compounds on the *Inga* phylogeny. Red (significant phylogenetic signal) and gray (nonsignificant) bars indicate where a compound is present in a given species. (c) Histogram for the compound lability index for all compounds present in > 2 species.

compounds (58%; lability ≥ 1.0) were labile having evolved as many or more times than they were present (Fig. 4c).

Evidence for phylogenetic signal at larger chemical scales

The chemical profiles of *Inga* species are dominated by two classes of compounds that can be broadly categorized as phenolics and saponins. Phenolic chemistry arises from the flavonoid pathway (Fig. S4 contains a summary of *Inga* phenolics). *Inga* phenolic chemistry is based on flavone and mono/polymeric flavan backbones that are extensively modified. *Inga* saponins are glycosylated triterpenoids that have their origin in the mevalonic acid pathway and as such are biosynthetically distinct from phenolic compounds. We mapped investment in each of these classes onto the phylogeny (Fig. 5) and then tested for phylogenetic signal of each subclass of these compounds. We found that quinic acid gallates ($K = 0.68$; $P = 0.02$), tyrosine and related depsides ($K = 0.73$; $P = 0.03$) as well as saponin glycosides ($K = 1.02$; $P = 0.007$) showed significant phylogenetic signal. By contrast, none of the flavonoid subclasses showed phylogenetic signal (Fig. 5).

We used phylogenetic structural equation modeling (SEM) to determine whether chemical classes were correlated with each other (Fig. S5). We applied this approach because it controls for the phylogenetic nonindependence of species and the biosynthetic nonindependence of predictor variables. Our SEM model revealed several trade-offs between compound classes, suggesting that there may be switch points between major branches of the biosynthetic pathway: saponin glycosides were negatively correlated with the left and right branch of the flavonoid pathway; quinic acid gallates were negatively correlated with the right side of the flavonoid pathway; and the right branch of the flavonoid pathway was negatively correlated with the left branch (Fig. S5).

Discussion

In this manuscript, we set out to thoroughly characterize the profile of plant secondary metabolites produced in nearly 100 species of *Inga* from across their geographic range. We combine untargeted metabolomics and phylogenetic comparative methods to answer questions about how chemical profiles evolve. Our analysis uncovered nearly 10 000 unique metabolites produced across the genus. Based on compound annotations, most of these compounds were flavonoids and saponin glycosides (Fig. 1), both prominent secondary metabolite classes in plants. These profiles largely exclude primary metabolites because they are generally observed in much lower concentrations than secondary metabolites and therefore are not readily detected in our UPLC-MS pipeline. Moreover, when these chemical extracts were incorporated at only 0.5–2% DW into artificial diets, they were highly detrimental to larval growth and survival, suggesting that they are toxic and contain defensive compounds (reviewed in Coley *et al.*, 2018). Although many of the compounds observed in this study may play a role in defense, determining the function of compounds is very challenging in metabolomics studies. To that end, in this study, we characterize the chemical profile as a whole, which contains a diversity of compounds likely selected for a variety of functions.

Diversity and divergence

Based on our analytical models, we found that each *Inga* species produces compounds that are more phytochemically diverse than would be expected by chance. This result underscores the strong selective pressure to generate and maintain chemical diversity that plants and other sessile organisms face from both harsh abiotic conditions and from a multitude of herbivores, pathogens, and

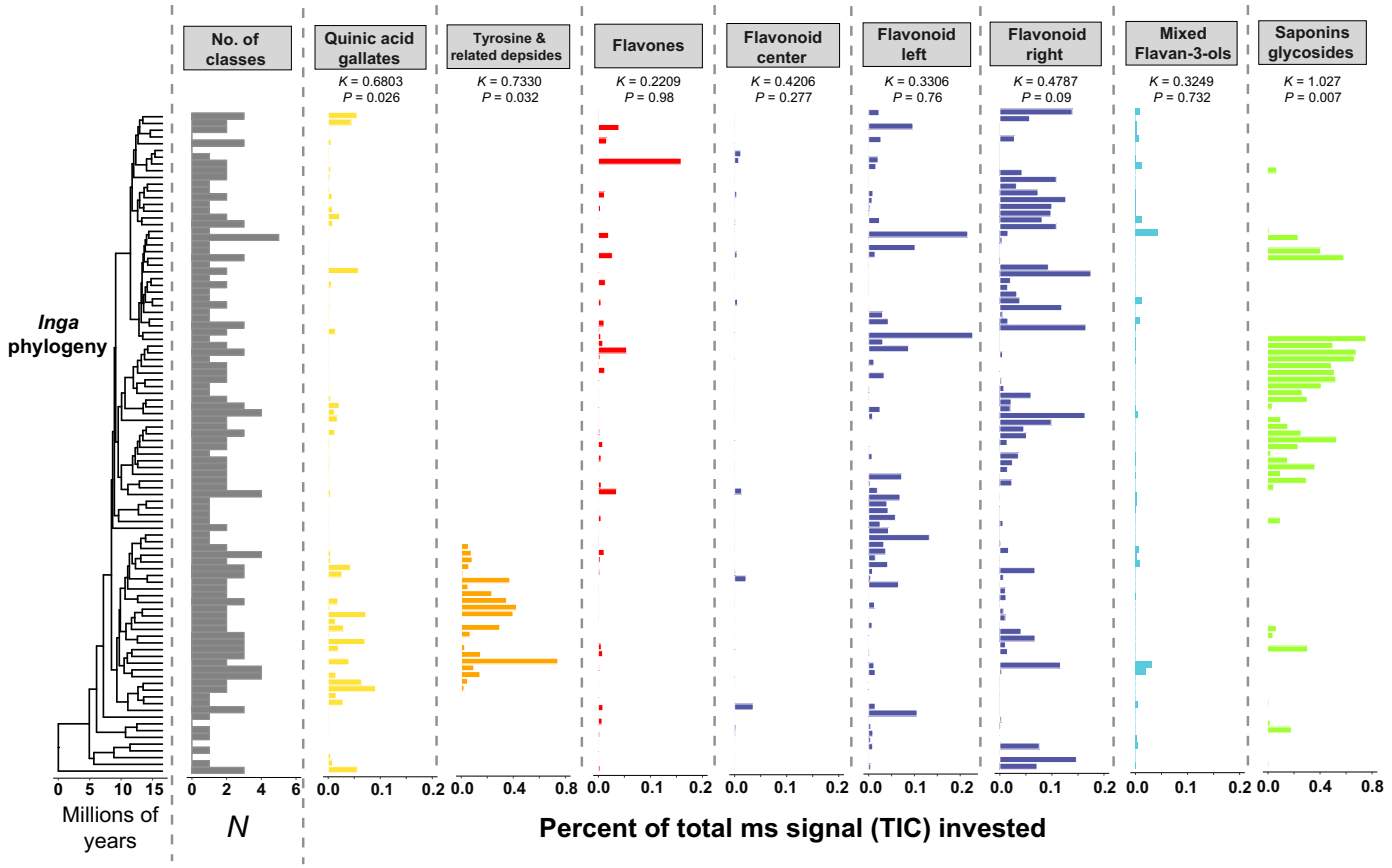

Fig. 5 Expression of defensive compound classes mapped on the *Inga* phylogeny. Total number of compound classes expressed per species, followed by expression per species of distinct classes including quinic acid gallates, tyrosine and related depsides, flavones, flavonoids, flavan-3-ols, and saponins. The expression of individual compound classes is measured as a percentage of the total MS-level-1 ion current (TIC, total ion current; metric of abundance) constituted by each class. The phylogenetic signal of each compound class and its significance are represented by Blomberg's *K* and corresponding *P*-values.

mutualists (Weng, 2014; Salazar *et al.*, 2018; Wetzel & Whitehead, 2019). Our results rely on a null model framework and the use of functional Hill numbers, which are a unifying and flexible approach to diversity measures (Chao *et al.*, 2014). They consider functional relatedness (cosine-based structural similarity between compounds) and compound abundance. We chose to exclude abundance measures in our measure ($Q = 0$), which results in a cosine-weighted structural similarity score.

We found strong evidence that a species' chemical profile evolved rapidly with little phylogenetic signal in chemical similarity (Figs 3, S3). These results confirm previous findings that defense strategy has little phylogenetic signal in *Inga* and other plant lineages (Becerra, 2007; Kursar *et al.*, 2009; Endara *et al.*, 2017; Salazar *et al.*, 2018; Volf *et al.*, 2018). We also found evidence for population-level divergence across sites in a species' chemical profile (Fig. 3a). This occurred despite the fact that there is essentially no limitation on the dispersal of *Inga* species across the Amazon, such that the metacommunity for any site is the entire Amazon basin (Dexter *et al.*, 2017; Endara *et al.*, 2021). Instead, site differences in abiotic and biotic conditions may drive intraspecific population-level differences in chemical profiles, including variation in soil types and precipitation patterns or the potentially complete turnover of herbivore communities (P. D. Coley *et al.*, unpublished). The fact that we observed divergent chemical profiles between close relatives in parapatry (Fig. 3) is unsurprising given many differences across sites in abiotic and biotic selection pressures (Thompson, 2005). However, the fact that sister species in sympatry (where all individuals are exposed to a similar community of pests and abiotic conditions) displayed much higher niche divergence (Fig. 3) is consistent with natural selection to not share pests and pathogens (Bagchi *et al.*, 2014; Forrister *et al.*, 2019). These results also highlight the importance of chemistry as an important niche axis facilitating species' coexistence (Chesson, 2000; Endara *et al.*, 2021).

Our modeling framework selected divergent adaptation as the best model to explain how interspecific differences in chemical profiles are evolving (Table S2). This divergent adaptation model shows that ecological interactions among coexisting species shape the evolutionary trajectory of a trait. A pattern of divergent adaptation also requires a divergent selective force, such as the one imposed by specialists pests and pathogens (Ehrlich & Raven, 1964). By contrast, if a species' chemical profile was evolving in response to an abiotic stressor, such as solar radiation, we would expect chemistry to converge among coexisting species. We posit that defenses, including a species' chemical profile, are

one of the first traits to diverge during or after the speciation process, especially compared with nondefensive traits such as those used for resource acquisition (Endara *et al.*, 2015).

What is the mode of chemical evolution in *Inga*?

Increasingly, evidence is supporting the adaptive value of chemical diversity both within and among plant species (Richards *et al.*, 2015; Salazar *et al.*, 2018; Wetzel & Whitehead, 2019; Whitehead *et al.*, 2021). But how are novel structures generated and what is the mode of chemical evolution? In the 'escape and radiate' model for defense evolution, novel structures evolve through the gradual embellishment of core structures into more complex and derived compounds (Berenbaum & Feeny, 1981; Berenbaum, 1983; Coley *et al.*, 2019). However, the results presented in this study do not support a model of chemical evolution underpinned by stepwise gradual embellishments. Instead, we found that each *Inga* maximizes phytochemical diversity by producing structurally unrelated compounds (Fig. 2); chemical similarity decreases rapidly over short phylogenetic distances (Fig. 3); and chemical profiles are evolving under divergent adaptation (Table S2). This high divergence between closely related species is supported by the fact that most compounds are highly labile (Fig. 4), and many compound classes show low phylogenetic signal (Fig. 5). Taken together, these patterns point toward regulation of gene expression as the more likely mechanism facilitating the rapid evolution of species' chemical profiles and for generating unique combinations of compounds that are divergent from neighbors within a community and from close relatives.

Regulatory changes facilitate divergence We propose that changes in gene regulation are a parsimonious explanation for the pattern of phylogenetically dispersed expression of individual compounds. Although compounds spread throughout the phylogeny could have evolved independently by convergent evolution, the scale of how frequently they are apparently gained and lost is more consistent with the up- and downregulation of key enzymes via transcriptional regulation (Moore *et al.*, 2014; Courtois *et al.*, 2016).

The role of regulation also applies at the compound class level where we find low phylogenetic signal and moderate trade-offs across biosynthetic pathways (Figs 5, S5). Consistent with our findings that *Inga* species invest in phytochemical diversity (Fig. 2), many species of *Inga* produce compounds from multiple biosynthetically distinct classes (Fig. 5). The ability of some species to produce compounds from up to five different compound classes coupled with the fact that one class did not completely exclude the production of other classes indicates that these trade-offs may not be driven by hard physiological constraints. For example, saponin production was negatively correlated with investment in flavan-3-ols, yet there were nine species that invested in both pathways simultaneously. The lack of strong physiological constraints likely facilitates the evolution of novel chemical profiles and divergence between closely related species.

Changes in gene expression would allow an evolutionary fluidity not possible via changes in genes coding for biosynthetic enzymes (structural genes). Regulatory changes in existing biosynthetic genes permit distantly related species to express the same compound and closely related species to express divergent compounds (Courtois *et al.*, 2016). For example, one sister species could make saponins and its close relative could make phenolics, presenting very different detoxification challenges for pests and pathogens. Thus, the evolutionary fluidity of defensive chemistry may be a major factor allowing long-lived trees to effectively persist in the arms race with insect herbivores and plant pathogens.

Regulation as a model for chemical evolution would imply that species maintain a complete set of biosynthetic enzymes within their genome that are up- or downregulated in different species and that 'unused' genes would have to remain functional over evolutionary time scales. Preliminary results from two *Inga* genomes indicate that the core biosynthetic genes involved in flavonoid and saponin biosynthesis are in fact present in all species even when they do not produce these compound classes (C. A. Kidner, 2021, pers. comm.). The maintenance of these supposedly unused enzymes may be required by deep homology and pleiotropy for core biosynthetic enzymes (Moore *et al.*, 2014; Moghe & Last, 2015). We offer several possibilities for how viable genes are maintained. First, many compounds, including pathway intermediates, do not accumulate to physiologically significant levels. However, because they are essential for the synthesis of downstream compounds, the enzymes responsible for them must be transcribed and maintained. This is the case for the phenylpropanoid compounds that link the shikimic acid pathway with the flavonoid pathway (Fig. S4). Second, it is possible that many compounds that are absent in leaves could be present in other tissues (Van Dam *et al.*, 2009; Schneider *et al.*, 2021).

'Lego-chemistry' as a mechanism for novel structures While regulatory changes may explain novel combinations of metabolites, regulation alone cannot generate novel structures. The classic 'escape and radiate' model proposes gradual embellishments to a compound's core structure. Instead, in *Inga,* we more commonly see the addition of larger structures, such as phenolic acids and carbohydrates, which are precursors and intermediates in secondary metabolism pathways (Figs 5, S5). The addition of these side groups in a combinatorial manner referred to as 'Lego-chemistry' has been shown to generate an impressively diverse array of larger structures from a small group of building blocks (Menzella *et al.*, 2005; Sherman, 2005).

Lego-chemistry could be particularly important for the generation of novel structures in the phenolic biosynthetic pathway, which produces the most diverse class of compounds in *Inga* (Fig. S4). *Inga* produces several subclasses of flavonoids that are further modified by the addition of divergent combinations of R-groups to key linkage sites on the basic scaffold molecule (flavonoid aglycones). For example, (epi)catechin (Fig. S4; comp 27), one of the most common compounds in *Inga*, is modified into at least four divergent structures (illustrated in Fig. S6), which upon polymerization lead to the generation of at least a dozen unique polymers (Fig. S4; comp 34).

The idea that combinatory Lego-chemistry may generate structural diversity in plants is in line with the growing body of literature on the underlying genetic and biochemical mechanisms for the evolution of plant secondary metabolism (Schwab, 2003; Gershenzon *et al.*, 2012; Kreis & Munkert, 2019; Monson *et al.*, 2021). There is a wide consensus that secondary metabolites originate from a small group of precursor compounds derived from primary metabolism with gene duplication and subsequent neofunctionalization driving novel metabolites (Moore *et al.*, 2014; Weng, 2014). Finally, because there are many more secondary metabolites than enzymes that produce them, it has been argued that a core set of enzymes with low substrate specificity is capable of producing a broad set of chemical structures (Schwab, 2003; Gershenzon *et al.*, 2012). This concept has proven to be important for generating novel structures via Lego-chemistry (Schwab, 2003; Gershenzon *et al.*, 2012; Kreis & Munkert, 2019).

Taken together, we hypothesize that the mode of chemical evolution for *Inga* is the combination of Lego-chemistry to generate novel structures along with changes in the regulation of gene expression to generate unique chemical profiles in each species. We put forth this model of chemical evolution to integrate the patterns we observed in our study of *Inga* metabolomes, with their underlying genetic, biochemical, and regulatory mechanisms. Future studies using multiomic approaches (Monson *et al.*, 2021) that integrate genomics, transcriptomics, and metabolomics are needed to further test and refine this working model.

Conclusions

In this paper, we integrate untargeted metabolomics and phylogenetic comparative methods to characterize the chemical profile of nearly 100 species of tropical trees from the genus *Inga*. We set out to address the fundamental questions of how phytochemical diversity evolves and what is the mode of chemical evolution. We show that each species maximizes phytochemical diversity by investing in structurally unrelated compounds. We also show that chemistry evolves rapidly, under a model of divergent adaptation. We find that sympatric sister species are more divergent than parapatric sister species implying an advantage to be distinct from other species in a community. Finally, we integrate these patterns into a hypothesized model of chemical evolution in which novel structures are generated through 'Lego-chemistry' and divergent profiles arise through transcriptional regulation. Understanding the evolution of plant chemistry is of fundamental importance because chemistry underpins a plant's ability to survive stressful abiotic conditions, as well as their ecological interaction such as interactions with pests, pathogens, and pollinators.

Acknowledgements

This work was greatly facilitated by research permits from the governments of Panama, Peru, Brazil and Ecuador and French Guiana. Thanks to Zachary Benavidez, Carine Emer, Julio Grandez, Wilder Hidalgo, Emily Kearny, Mayra Ninazunta, Wilmer Rosendo, Joe Sixto Saldaña, Yamara Serrano, Georgia Sinimbu, Allison Thompson, and Marjorie Weber for their valuable field assistance. This work was supported by the following grants: NSF GRFP to DLF; Nouragues Travel Grants Program, CNRS, France, and the National Science Foundation (DEB-0640630 and DIMENSIONS of Biodiversity DEB-1135733) to PDC; and the Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador (SENESCYT) to M-JE.

Author contributions

TAK, M-JE, DLF and PDC designed and conducted the research. DLF designed and performed the data analysis. TAK, DLF, AJS, GCY and AGM contributed to the metabolomic analysis. TAK and JL provided initial characterization of *Inga* chemistry via NMR. JAN, RTP, KGD, CAK and OL contributed to the phylogeny of *Inga*. DLF, M-JE, AJS and PDC wrote the manuscript. All authors provided feedback and edited the manuscript.

ORCID

Kyle G. Dexter https://orcid.org/0000-0001-9232-5221
María-José Endara https://orcid.org/0000-0002-8805-1456
Dale L. Forrister https://orcid.org/0000-0001-8170-7187
Catherine A. Kidner https://orcid.org/0000-0001-6426-3000
Oriane Loiseau https://orcid.org/0000-0002-9852-857X
James A. Nicholls https://orcid.org/0000-0002-9325-563X
R. Toby Pennington https://orcid.org/0000-0002-8196-288X
Abrianna J. Soule https://orcid.org/0000-0001-9164-7317
Gordon C. Younkin https://orcid.org/0000-0002-3735-3534

Data availability

Chemical data and scripts to estimate chemical similarity, downstream analysis, and figure generation, are deposited in a git repository (https://github.com/dlforrister/Forrister_et_al_2022_evolution_of_inga_chemistry.git).

References

Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, Ikeda S, Takahashi H, Altaf-Ul-Amin M, Darusman LK *et al.* 2012. KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant & Cell Physiology* **53**: 1–12.

Agrawal AA, Fishbein M. 2006. Plant defense syndromes. *Ecology* **87**: 132–149.

Agrawal AA, Salminen JP, Fishbein M. 2009. Phylogenetic trends in phenolic metabolism of milkweeds (*Asclepias*): evidence for escalation. *Evolution* **63**: 663–673.

Anderson SAS, Weir JT. 2020. A comparative test for divergent adaptation: inferring speciation drivers from functional trait divergence. *The American Naturalist* **196**: 429–442.

Bagchi R, Gallery RE, Gripenberg S, Gurr SJ, Narayan L, Addis CE, Freckleton RP, Lewis OT. 2014. Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature* **506**: 85–88.

Becerra JX. 2007. The impact of herbivore–plant coevolution on plant community structure. *Proceedings of the National Academy of Sciences, USA* **104**: 7483–7488.

Berenbaum MR. 1983. Coumarins and caterpillars: a case for coevolution. *Evolution* **37**: 163–179.

Berenbaum MR, Feeny P. 1981. Toxicity of angular furanocoumarins to swallowtail butterflies: escalation in a coevolutionary arms race? *Manajemen Pengembangan Bakat Minat Siswa di MtsAlWathoniyyah Pedurungan Semarang* **212**: 927–929.

Berenbaum MR, Zangerl AR. 1996. Phytochemical diversity: adaptation or random variation? *Recent Advances in Phytochemistry* **30**: 1–24.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**: 717–745.

Bollback JP. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**: 1–7.

Chao A, Chiu C-H, Jost L. 2014. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annual Review of Ecology, Evolution, and Systematics* **45**: 297–324.

Chernomor O, Von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology* **65**: 997–1008.

Chesson P. 2000. Mechanisms of maintenance of species diversity. *Annual Review of Ecology, Evolution, and Systematics* **31**: 343–366.

Coley PD, Aide TM. 1991. Comparison of plant defenses in temperate and tropical broad-leaved forests. In: Price PW, Lewinsohn TM, Fernandes GW, Benson WW, eds. *Plant–animal interactions: evolutionary ecology in tropical and temperate regions*. New York, NY, USA: Wiley, 25–49.

Coley PD, Endara MJ, Ghabash G, Kidner CA, Nicholls JA, Pennington RT, Mills AG, Soule AJ, Lemes MR, Stone GN *et al.* 2019. Macroevolutionary patterns in overexpression of tyrosine: an anti-herbivore defence in a speciose tropical tree genus, *Inga* (Fabaceae). *Journal of Ecology* **107**: 1620–1632.

Coley PD, Endara MJ, Kursar TA. 2018. Consequences of interspecific variation in defenses and herbivore host choice for the ecology and evolution of *Inga*, a speciose rainforest tree. *Oecologia* **187**: 361–376.

Courtois EA, Dexter KG, Paine C, Timothy E, Stien D, Engel J, Baraloto C, Chave J. 2016. Evolutionary patterns of volatile terpene emissions across 202 tropical tree species. *Ecology and Evolution* **6**: 2854–2864.

Dethier VG. 1954. Evolution of feeding preferences in phytophagous insects. *Evolution* **8**: 33–54.

Dexter KG, Lavin M, Torke BM, Twyford AD, Kursar TA, Coley PD, Drake C, Hollands R, Pennington RT. 2017. Dispersal assembly of rain forest tree communities across the Amazon basin. *Proceedings of the National Academy of Sciences, USA* **114**: 2645–2650.

Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, Fahy E, Steinbeck C, Subramanian S, Bolton E *et al.* 2016. CLASSYFIRE: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics* **8**: 61.

Ehrlich PR, Raven PH. 1964. Butterflies and plants: a study in coevolution. *Evolution* **18**: 586–608.

Endara MJ, Coley PD, Ghabash G, Nicholls JA, Dexter KG, Donoso DA, Stone GN, Pennington RT, Kursar TA. 2017. Coevolutionary arms race versus host defense chase in a tropical herbivore–plant system. *Proceedings of the National Academy of Sciences, USA* **114**: E7499–E7505.

Endara MJ, Coley PD, Wiggins NL, Forrister DL, Younkin GC, Nicholls JA, Pennington RT, Dexter KG, Kidner CA, Stone GN *et al.* 2018a. Chemocoding as an identification tool where morphological- and DNA-based methods fall short: *Inga* as a case study. *The New Phytologist* **218**: 847–858.

Endara MJ, Forrister DL, Nicholls JA, Stone GN, Kursar TA, Coley PD. 2022. Impacts of plant defenses on host choice by Lepidoptera in Neotropical rainforests. In: Marquis RJ, Kopter S, eds. *Caterpillars in the middle: tritropic interactions in a changing world*. Cham, Switzerland: Springer, 93–114.

Endara MJ, Nicholls JA, Coley PD, Forrister DL, Younkin GC, Dexter KG, Kidner CA, Pennington RT, Stone GN, Kursar TA. 2018b. Tracking of host defenses and phylogeny during the radiation of neotropical *Inga*-feeding sawflies (Hymenoptera; Argidae). *Frontiers in Plant Science* **9**: 1237.

Endara MJ, Soule AJ, Forrister DL, Dexter KG, Pennington RT, Nicholls JA, Loiseau O, Kursar TA, Coley PD. 2021. The role of plant secondary metabolites in shaping regional and local plant community assembly. *Journal of Ecology* **110**: 34–45.

Endara MJ, Weinhold A, Cox JE, Wiggins NL, Coley PD, Kursar TA. 2015. Divergent evolution in antiherbivore defences within species complexes at a single Amazonian site. *Journal of Ecology* **103**: 1107–1118.

Forrister DL, Endara MJ, Younkin GC, Coley PD, Kursar TA. 2019. Herbivores as drivers of negative density dependence in tropical forest saplings. *Science* **363**: 1213–1216.

Fraenkel GS. 1959. The raison d'Être of secondary plant substances. *Science* **129**: 1466–1470.

Fritz SA, Purvis A. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* **24**: 1042–1051.

Gershenzon J. 1994. Metabolic costs of terpenoid accumulation in higher plants. *Journal of Chemical Ecology* **20**: 1281–1328.

Gershenzon J, Fontana A, Burow M, Wittstock U, Degenhardt J. 2012. Mixtures of plant secondary metabolites: metabolic origins and ecological benefits. In: *The ecology of plant secondary metabolites from genes to global processes*. Cambridge, UK: Cambridge University Press, 335.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBOOT2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**: 518–522.

Jones CG, Firn RD. 1991. On the evolution of plant secondary chemical diversity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **333**: 273–280.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. MODELFINDER: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Keinänen M, Julkunen-Tiitto R, Mutikainen P, Walls M, Ovaska J, Vapaavuori E. 1999. Trade-offs in phenolic metabolism of silver birch: effects of fertilization, defoliation, and genotype. *Ecology* **80**: 1970–1986.

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. 2010. PICANTE: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464.

Kessler A, Kalske A. 2018. Plant secondary metabolite diversity and species interactions. *Annual Review of Ecology, Evolution, and Systematics* **49**: 115–138.

Koenen EJM, Ojeda DI, Steeves R, Migliore J, Bakker FT, Wieringa JJ, Kidner C, Hardy OJ, Pennington RT, Bruneau A *et al.* 2020. Large-scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a near-simultaneous evolutionary origin of all six subfamilies. *New Phytologist* **225**: 1355–1369.

Koricheva J, Nykänen H, Gianoli E. 2004. Meta-analysis of trade-offs among plant antiherbivore defenses: are plants jacks-of-all-trades, masters of all? *The American Naturalist* **163**: E64–E75.

Kreis W, Munkert J. 2019. Exploiting enzyme promiscuity to shape plant specialized metabolism. *Journal of Experimental Botany* **70**: 1435–1445.

Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. 2012. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry* **84**: 283–289.

Kursar TA, Coley PD. 2003. Convergence in defense syndromes of young leaves in tropical rainforests. *Biochemical Systematics and Ecology* **31**: 929–949.

Kursar TA, Dexter KG, Lokvam J, Pennington RT, Richardson JE, Weber MG, Murakami ET, Drake C, McGregor R, Coley PD. 2009. The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proceedings of the National Academy of Sciences, USA* **106**: 18073–18078.

Maron JL, Agrawal AA, Schemske DW. 2019. Plant–herbivore coevolution and plant speciation. *Ecology* **100**: e02704.

Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, Patel KG, Hopwood DA, Santi DV. 2005. Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nature Biotechnology* **23**: 1171–1176.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.

Moghe GD, Last RL. 2015. Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiology* **169**: 1512–1523.

Monson RK, Trowbridge AM, Lindroth RL, Lerdau MT. 2021. Coordinated resource allocation to plant growth–defense tradeoffs. *New Phytologist* **233**: 1051–1066.

Moore BD, Andrew RL, Külheim C, Foley WJ. 2014. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytologist* **201**: 733–750.

Nicholls JA, Pennington RT, Koenen EJ, Hughes CE, Hearn J, Bunnefeld L, Dexter KG, Stone GN, Kidner CA. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science* **6**: 1–20.

Nyman T, Julkunen-Tiitto R. 2005. Chemical variation within and among six northern willow species. *Phytochemistry* **66**: 2836–2843.

Orme D. 2013. *The CAPER package: comparative analyses in phylogenetics and evolution in R.* R package v.5.2, 1–36.

Panda S, Kazachkova Y, Aharoni A. 2021. Catch-22 in specialized metabolism: balancing defense and growth. *Journal of Experimental Botany* **72**: 6027–6041.

Revell LJ. 2012. PHYTOOLS: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

Richards LA, Dyer LA, Forister ML, Smilanich AM, Dodson CD, Leonard MD, Jeffrey CS. 2015. Phytochemical diversity drives plant–insect community diversity. *Proceedings of the National Academy of Sciences, USA* **112**: 10973–10978.

Richards LA, Glassmire AE, Ochsenrider KM, Smilanich AM, Dodson CD, Jeffrey CS, Dyer LA. 2016. Phytochemical diversity and synergistic effects on herbivores. *Phytochemistry Reviews* **15**: 1153–1166.

Salazar D, Jaramillo A, Marquis RJ. 2016. The impact of plant chemical diversity on plant–herbivore interactions at the community level. *Oecologia* **181**: 1199–1208.

Salazar D, Lokvam J, Mesones I, Pilco MV, Zuñiga JMA, De Valpine P, Fine PVA. 2018. Origin and maintenance of chemical diversity in a species-rich tropical tree lineage. *Nature Ecology & Evolution* **2**: 983–990.

Schneider GF, Salazar D, Hildreth SB, Helm RF, Whitehead SR. 2021. Comparative metabolomics of fruits and leaves in a hyperdiverse lineage suggests fruits are a key incubator of phytochemical diversification. *Frontiers in Plant Science* **12**: 693739.

Schwab W. 2003. Metabolome diversity: too few genes, too many metabolites? *Phytochemistry* **62**: 837–849.

Sedio BE, Parker JD, McMahon SM, Wright SJ. 2018. Comparative foliar metabolomics of a tropical and a temperate forest community. *Ecology* **99**: 2647–2653.

Sedio BE, Rojas Echeverri JC, Boya CA, Wright SJ. 2017. Sources of variation in foliar secondary chemistry in a tropical forest tree community. *Ecology* **98**: 616–623.

Sherman DH. 2005. The lego-ization of polyketide biosynthesis. *Nature Biotechnology* **23**: 1083–1084.

Smith SA, O'Meara BC. 2012. TREEPL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**: 2689–2690.

Strauss SY, Rudgers JA, Lau JA, Irwin RE. 2002. Direct and ecological costs of resistance to herbivory. *Trends in Ecology & Evolution* **17**: 278–285.

Thompson JN. 2005. *The geographic mosaic of coevolution (interspecific interactions).* Chicago, IL, USA: University of Chicago Press.

Valencia R, Condit R, Foster RB, Romoleroux K, Munoz GV, Svenning J-C, Magård E, Bass M, Losos EC, Balslev H. 2004. Yasuní forest dynamics plot, Ecuador. In: Losos EC, Leigh JG, eds. *Forest diversity and dynamism: findings from a large-scale plot network.* Chicago, IL, USA: University of Chicago Press, 609–620.

Van Dam NM, Tytgat TOG, Kirkegaard JA. 2009. Root and shoot glucosinolates: a comparison of their diversity, function and interactions in natural and managed ecosystems. *Phytochemistry Reviews* **8**: 171–186.

Vleminckx J, Salazar D, Fortunel C, Mesones I, Dávila N, Lokvam J, Beckley K, Baraloto C, Fine PVA. 2018. Divergent secondary metabolites and habitat filtering both contribute to tree species coexistence in the Peruvian Amazon. *Frontiers in Plant Science* **9**: 836.

Volf M, Segar ST, Miller SE, Isua B, Sisol M, Aubona G, Šimek P, Moos M, Laitila J, Kim J *et al.* 2018. Community structure of insect herbivores is driven by conservatism, escalation and divergence of defensive traits in *Ficus*. *Ecology Letters* **21**: 83–92.

Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T *et al.* 2016. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nature Biotechnology* **34**: 828–837.

Weng JK. 2014. The evolutionary paths towards complexity: a metabolic perspective. *New Phytologist* **201**: 1141–1149.

Wetzel WC, Whitehead SR. 2019. The many dimensions of phytochemical diversity: linking theory to practice. *Ecology Letters* **23**: 16–32.

Whitehead SR, Bass E, Corrigan A, Kessler A, Poveda K. 2021. Interaction diversity explains the maintenance of phytochemical diversity. *Ecology Letters* **24**: 1205–1214.

Wiggins NL, Forrister DL, Endara MJ, Coley PD, Kursar TA. 2016. Quantitative and qualitative shifts in defensive metabolites define chemical defense investment during leaf development in *Inga*, a genus of tropical trees. *Ecology and Evolution* **6**: 478–492.

Wink M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**: 3–19.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 15–30.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Defense investment traits mapped on to the *Inga* phylogeny.

Fig. S2 Compound-based molecular network containing all compounds observed in 98 study species

Fig. S3 Correlation between chemical similarity and phylogenetic distance (Myr) for all interspecific comparisons.

Fig. S4 Biosynthetic context of phenolic compounds in *Inga*.

Fig. S5 Structural equation modeling-based correlation among major defensive compound classes mapped on the *Inga* phylogeny.

Fig. S6 Illustration of Lego-chemistry concept.

Methods S1 Code and description of null model for phytochemical diversity and chemical similarity.

Table S1 Site and sampling information for all 98 study species.

Table S2 Maximum-likelihood estimates for different evolutionary models of trait evolution.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.