Diversity and Divergence: Evolution of secondary metabolism in the tropical tree genus Inga

Tittle: Diversity and Divergence: Evolution of secondary metabolism in the tropical tree genus *Inga*

Authors: Dale L. Forrister[a], María-José Endara[a,b], Abrianna J. Soule[a], Gordon C. Younkin[c,d], Anthony G. Mills[a], John Lokvam[a], Kyle G. Dexter[e], R. Toby Pennington[f], Catherine A. Kidner[gh], James A. Nicholls[i], Oriane Loiseau[e], Thomas A. Kursar[a], Phyllis D. Coley[a]

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

Corresponding Author: Dale L. Forrister
Address: School of Biological Sciences, University of Utah, Aline W. Skaggs Biology Building, 257 S 1400 E, Salt Lake City, UT 84112-0840, USA.
Email: dale.forrister@utah.edu

This PDF file includes:

Introduction: 1,539
Materials and Methods: 1,870
Results: 930
Discussion: 2,213

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. Here 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* (~ 100, Fabaceae) using UPLC-MS based metabolomics and applied phylogenetic comparative methods to understand the mode of chemical evolution.

- We show that: 1) Each *Inga* species contain structurally unrelated compounds and exceptionally high levels of phytochemical diversity. 2) Closely related species have divergent defense profiles, with individual compounds, major compound classes and complete chemical profiles showing little to no phylogenetic signal. 3) 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. 4) Finally, we integrate these patterns of chemical diversity with a model for how chemical diversity evolves. We posit that the combinatorial "Lego-chemistry" and rapid changes in regulatory mechanisms may explain the observed large shifts in chemical profiles between closely related taxa.

Keywords: chemical defense, secondary metabolism, evolution, metabolomics, phytochemical diversity, plant secondary metabolism, UPLC-MS, *Inga*

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 1,000,000 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 and Raven 1964; Endara et al. 2017; Endara et al. 2018). 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 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 1991; Berenbaum and Zangerl 1996; Richards et al. 2016; Kessler and Kalske 2018; Salazar et al. 2018; Wetzel and Whitehead 2020). 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. 2021) and herbivory (Richards et al. 2015). In addition to producing a diverse set of compounds, recent studies have highlighted the importance for 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 and Whitehead 2020).

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 off the classic 'escape and radiate' theoretical frame, first suggested a half-century ago by the work of Dethier (1954), Fraenkel (1959), and Ehrlich and 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 1983, Berenbaum and Feeny 1981; 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 step-wise mannor through the modification of core structures (Ehrlich and Raven 1964, Berenbaum 1983). To test this, we combine untargeted metabolomic and comparative phylogenetic methods to characterize the chemical profiles for nearly 100 species of tropical trees in the genus *Inga* (Fabaceae). By focusing on a recently radiated monophyletic genus of trees, we attempt to understand how chemistry evolves at tips of the phylogenetic tree over a relatively short period of evolutionary history. This offers a different perspective than 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 ~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 off of 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 as well as 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 and Whitehead 2020; Endara et al. 2021), yet investment in chemical defense comes at a cost (known as the 'growth-defense trade-off') (Strauss et al. 2002; Panda et al. 2021; Monson et al. 2022). Investment in chemical defense is expensive both in terms of the carbon and nitrogen used as inputs for the biosynthetic products, as well as 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. 2021), 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 and Raven 1964; Berenbaum and 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 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 if species' chemical profiles show phylogenetic signal, or if they have diverged sufficiently to erase the effect of shared evolutionary history. We also incorporate biogeography asking if 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 and 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 to chemical structures resulting in more derived chemical defenses over evolutionary time (Berenbaum and Feeny 1981; Coley et al. 2019). This process should lead to a pattern of phylogenetic conservatism of metabolites and biosynthetic pathways (Ehrlich and 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 and Coley 2003; Agrawal and Fishbein 2006; Agrawal et al. 2009; Coley et al. 2018; Monson et al. 2022) and positive correlations (Agrawal and 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 and 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 timescales or if 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 (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 approximately 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 as well as 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. 2018). Young leaves at approximately 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 3 individuals). We focused on expanding leaves, as they receive more than 70% of the lifetime damage of a leaf (Coley and Aide 1991), and their chemical profiles are an important factor for host associations of insect herbivores (Endara et al. 2017; Endara et al. 2018). In general, we found the chemical profile of each species to be highly canalized and previous work has shown that 5 individuals is sufficient to capture ~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^{o} C.

Characterization of *Inga* Chemistry:

a) Soluble secondary metabolites:

Metabolites were extracted from dried leaf samples in the Coley/Kursar lab at the University of Utah using a solution of (60:40, v/v) ammonium acetate buffered water, pH 4.8: acetonitrile, producing 2mL of retained supernatant from 100mg (+/- 2.5 mg) of sample for chromatographic analysis following the UPLC-MS methods developed in Wiggins et al. (2016). Extraction weight (percent dry weight) 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 (Waters Acquity I-Class, 2.1 x 150mm BEH C18 and 2.1 x 100 mm BEH Amide columns) and mass spectrometry (Waters Xevo G2 QToF) (UPLC-MS) in negative ionization mode. A 45 minute 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 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 minutes 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).

b) *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 dry weight. 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^{o}C (Coley et al. 2019). . Fifteen microliters of the supernatant were injected on a 4.6 x 250 mm amino-propyl HPLC column (Microsorb 5u, Varian). Metabolites were chromatographed using a linear gradient (17–23%) of aqueous acetic acid (pH 3.0) in acetonitrile over 25 min. Mass of solutes in each injection were measured using an evaporative light scattering detector (SEDERE S.A., 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, r^2=0.99) prepared from pure tyrosine.

a) *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 1) a species by compound

abundance (MS-1 peak intensity measured by total ion current) matrix and 2) 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. 3) 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 (Forrister & Soule, 2020; https://gitlab.chpc.utah.edu/01327245/evolution_of_inga_chemistry).

a) 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 (Endara et al. 2021; Sedio et al. 2017). 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. 2018; Endara et al. 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 variation 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 Appendix 1 for 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 in Nicholls et al. (Nicholls et al. 2015), supplemented with a subset of the loci from work by Koenen et al. (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 1,000 bootstrap replicates with the ultrafast bootstrap approximation (Hoang et al. 2017). The resulting gene trees were subsequently used as the input for ASTRAL-III to

estimate a phylogeny in a summary coalescent framework (Chernomor et al. 2016), 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 (1,000 replicates). The phylogenetic tree was subsequently time-calibrated using penalized likelihood implemented in the program treePL (Smith and 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 My and 13.4-16.6 My, 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 and Purvis 2010) using the *caper* package (Orme 2012).

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 and Weir (2020) which uses simulated trait values based on either Brownian motion or Ornstein–Uhlenbeck. This framework also test 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 9,105 unique compounds across 808 samples. *Inga* species invest substantial resources in soluble secondary metabolites, averaging 194 ± 103 (mean ± s.d.) unique compounds per species, and comprising 37 ± 11% (mean ± s.d.) of the expanding leaf's dry weight (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.

1. *Individual species invest in structurally diverse compounds.*

We asked whether biosynthetic tradeoffs 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 to 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). The absence of species with lower phytochemical diversity than the null model indicates that all species invest in structurally diverse compounds.

2. *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 quantifying overall chemical similarity between two species (Endara et al. 2021) . We compared these calculations to estimates of chemical similarity expected from a null model (Appendix 1). 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 (Fig. 3; Fig. 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 to 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 towards an optimal trait value (Table S2) rather than diverging.

3. *Many compounds showed no phylogenetic signal and were evolutionary labile.*

The majority of compounds are detected in only a few species (median = 4), and roughly half (53%) of 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 compounds (58%; lability >= 1.0) were labile having evolved as many or more times than they were present (Fig. 4C).

4. *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. S5 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. In contrast, all flavonoid subclasses showed no phylogenetic signal (Fig. 5).

We used phylogenetic structural equation modeling (SEM) to determine if chemical classes were correlated with each other (Fig. S4). We applied this approach because it controls for the phylogenetic non-independence of species as well as the biosynthetic non-independence 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: 1) saponin glycosides were negatively correlated with the left and right branch of the flavonoid pathway, 2) quinic acid gallates were negatively correlated with the right side of the flavonoid pathway and 3) the right branch of the flavonoid pathway was negatively correlated with the left branch (Fig. S4).

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 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 mutualists (Weng 2014; Salazar et al. 2018; Wetzel and Whitehead 2020). 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) as well as 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 (Fig. 3, Fig. 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 (our unpublished data). 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 one imposed by specialists pests and pathogens (Ehrlich and Raven 1964). In 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 non-defensive traits such as those used for resource acquisition (Endara et al. 2015).

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. S4). The ability for some species to produce compounds from up to five different classes coupled with the fact that one class did not completely exclude the production of other classes indicate that physiological constraints may not impose large biosynthetic trade-offs among compound classes. For example, saponin production was negatively correlated with investment in flavan-3-ols, yet there were nine species that invested in both of these pathways simultaneously. The lack of strong physiological constraints likely facilitates the evolution of novel chemical profiles and divergence between closely related species.

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 and Whitehead 2020; 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 and 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 and produces 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. S2). Taken together, these patterns point towards 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 is 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 down-regulation 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 (Fig. 5, Fig S4). 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. S4). The ability for 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 to genes coding for biosynthetic enzymes (structural genes). Regulatory changes of 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 down-regulated in different species and that "unused" genes would have to remain functional over evolutionary timescales. 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 (pers. comm. C.A. Kidner, 2021). The maintenance of these supposedly unused enzymes may be required by deep homology and pleiotropy for core biosynthetic enzymes (Moore et al. 2014; Moghe and 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. S5). 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 (Fig. 5, Fig S4). 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. S5). *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. S5, 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. S5, 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 and Munkert 2019; Monson et al. 2022). 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 and 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 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. 2022) 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 advantageous 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.

Acknowledgments:
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 D.L.F.; Nouragues Travel Grants Program, CNRS, France, and the National Science Foundation (DEB-0640630 and DIMENSIONS of Biodiversity DEB-1135733) to P.D.C.; and the Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador (SENESCYT) to M-J.E.

Author Contributions:

T.A.K., M-J.E., D.L.F., and P.D.C designed and conducted the research. D.L.F. designed and performed the data analysis. T.A.K, D.L.F., A.J.S., G.C.Y., and A.G.M, contributed to the metabolomic analysis. T.A.K. and J.L. provided initial characterization of *Inga* chemistry via NMR. J.A.N., R.T.P., K.G.D., C.A.K., and O.L. contributed the phylogeny of *Inga*. D.L.F., M-J.E., A.J.S. and P.D.C. wrote the manuscript. All authors provided feedback and edited the manuscript.

Data availability:

Chemical data and scripts to estimate chemical similarity are deposited in a git repository (Forrister & Soule, 2020; https://gitlab.chpc.utah.edu/01327245/evolution_of_inga_chemistry). All scripts for downstream data analysis and figure generation can be found at (Forrister 2021; https://github.com/dlforrister/Evolution_Of_Inga_Chemistry.git)

References:

557 Afendi, F.M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K.,
558 *et al.* (2012). KNApSAcK family databases: Integrated metabolite-plant species
559 databases for multifaceted plant research. *Plant Cell Physiol.*, 53, 1–12.

560 Agrawal, A.A. & Fishbein, M. (2006). Plant defense syndromes. *Ecology*, 87, 132–149.

561 Agrawal, A.A., Salminen, J.P. & Fishbein, M. (2009). Phylogenetic trends in phenolic
562 metabolism of milkweeds (Asclepias): Evidence for escalation. *Evolution (N. Y).*,
563 63, 663–673.

564 Anderson, S.A.S. & Weir, J.T. (2020). A comparative test for divergent adaptation:
565 Inferring speciation drivers from functional trait divergence. *Am. Nat.*, 196, 429–
566 442.

567 Bagchi, R., Gallery, R.E., Gripenberg, S., Gurr, S.J., Narayan, L., Addis, C.E., *et al.*
568 (2014). Pathogens and insect herbivores drive rainforest plant diversity and
569 composition. *Nature*, 506, 85–8.

570 Becerra, J.X. (2007). The impact of herbivore-plant coevolution on plant community
571 structure. *Proc. Natl. Acad. Sci. U. S. A.*, 104, 7483–7488.

572 Berenbaum, M.R. (1983). Coumarins and caterpillars: A case for coevolution. *Evolution*
573 *(N. Y).*, 37, 163.

574 Berenbaum, M.R. & Feeny, P. (1981). Toxicity of angular furanocoumarins to
575 swallowtail butterflies: Escalation in a coevolutionary arms race? *Manaj. Pengemb.*
576 *Bakat Minat Siswa Di Mts Al-Wathoniyyah Pedurungan Semarang*, 212, 927–929.

577 Berenbaum, M.R. & Zangerl, A.R. (1996). Phytochemical diversity: Adaptation or
578 random variation? *Recent Adv. Phytochem.*, 30, 1–24.

579 Blomberg, S.P., Garland, T. & Ives, A.R. (2003). Testing for phylogenetic signal in
580 comparative data: Behavioral traits are more labile. *Evolution (N. Y).*, 57, 717–745.

581 Bollback, J.P. (2006). SIMMAP: Stochastic character mapping of discrete traits on
582 phylogenies. *BMC Bioinformatics*, 7, 1–7.

583 Chao, A., Chiu, C.-H. & Jost, L. (2014). Unifying species diversity, phylogenetic
584 diversity, functional diversity, and related similarity and differentiation measures
585 through hill numbers. *Annu. Rev. Ecol. Evol. Syst.*, 45, 297–324.

586 Chernomor, O., Von Haeseler, A. & Minh, B.Q. (2016). Terrace aware data structure for
587 phylogenomic inference from supermatrices. *Syst. Biol.*, 65, 997–1008.

588 Coley, P.D. & Aide, T.M. (1991). Comparison of herbivory and plant defences in
589 temperate and tropical broad-leaved forests. In: *Plant–animal interactions:*
590 *Evolutionary ecology in tropical and temperate regions*. Wiley, New York, pp. 25–
591 49.

592 Coley, P.D., Endara, M.J., Ghabash, G., Kidner, C.A., Nicholls, J.A., Pennington, R.T., *et*
593 *al.* (2019). Macroevolutionary patterns in overexpression of tyrosine: An anti-
594 herbivore defence in a speciose tropical tree genus, Inga (Fabaceae). *J. Ecol.*, 107,
595 1620–1632.

596 Coley, P.D., Endara, M.J. & Kursar, T.A. (2018). Consequences of interspecific variation
597 in defenses and herbivore host choice for the ecology and evolution of Inga, a
598 speciose rainforest tree. *Oecologia*, 187, 361–376.

599 Courtois, E.A., Dexter, K.G., Paine, C., Timothy, E., Stien, D., Engel, J., *et al.* (2016).
600 Evolutionary patterns of volatile terpene emissions across 202 tropical tree species.
601 *Ecol. Evol.*, 6, 2854–2864.

602 Van Dam, N.M., Tytgat, T.O.G. & Kirkegaard, J.A. (2009). Root and shoot
603 glucosinolates: A comparison of their diversity, function and interactions in natural
604 and managed ecosystems. *Phytochem. Rev.*, 8, 171–186.

605 Dethier, V.G. (1954). Evolution of feeding preferences in phytophagous insects.
606 *Evolution (N. Y).*, 8, 33–54.

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

611 Dexter, K.G., Lavin, M., Torke, B.M., Twyford, A.D., Kursar, T.A., Coley, P.D., *et al.*
612 (2017). Dispersal assembly of rain forest tree communities across the Amazon basin.
613 *Proc. Natl. Acad. Sci. U. S. A.*, 114, 2645–2650.

614 Ehrlich, P.R. & Raven, P.H. (1964). Butterflies and plants: A study in coevolution.
615 *Evolution (N. Y).*, 18, 586–608.

616 Endara, M.J., Coley, P.D., Ghabash, G., Nicholls, J.A., Dexter, K.G., Donoso, D.A., *et al.*
617 (2017). Coevolutionary arms race versus host defense chase in a tropical herbivore–
618 plant system. *Proc. Natl. Acad. Sci. U. S. A.*, 114, E7499–E7505.

619 Endara, M.J., Coley, P.D., Wiggins, N.L., Forrister, D.L., Younkin, G.C., Nicholls, J.A.,
620 *et al.* (2018a). Chemocoding as an identification tool where morphological- and
621 DNA-based methods fall short: Inga as a case study. *New Phytol.*, 218, 847–858.

622 Endara, M.J., Forrister, D.L., Nicholls, J.A., Stone, G.N., Kursar, T.A. & Coley, P.D.
623 (2021a). Impacts of plant defenses on host choice by Lepidoptera in Neotropical
624 rainforests. *In: Caterpillars in the middle: Tritropic interactions in a changing*
625 *world.* (eds. Marquis, R.J. & Kopter, S.). Springer, Cham. 93-114.

626 Endara, M.J., Nicholls, J.A., Coley, P.D., Forrister, D.L., Younkin, G.C., Dexter, K.G., *et*
627 *al.* (2018b). Tracking of host defenses and phylogeny during the radiation of
628 neotropical Inga-feeding sawflies (Hymenoptera; Argidae). *Front. Plant Sci.*, 9.

629 Endara, M.J., Soule, A.J., Forrister, D.L., Dexter, K.G., Pennington, R.T., Nicholls, J.A.,

630 *et al.* (2021b). The role of plant secondary metabolites in shaping regional and local
631 plant community assembly. *Journal of Ecology*, *110*, 34–45

632 Endara, M.J., Weinhold, A., Cox, J.E., Wiggins, N.L., Coley, P.D. & Kursar, T.A.
633 (2015). Divergent evolution in antiherbivore defences within species complexes at a
634 single Amazonian site. *J. Ecol.*, 103, 1107–1118.

635 Forrister, D.L., Endara, M.J., Younkin, G.C., Coley, P.D. & Kursar, T.A. (2019).
636 Herbivores as drivers of negative density dependence in tropical forest saplings.
637 *Science (80-.).*, 363, 1213–1216.

638 Fraenkel, G.S. (1959). The raison d'Être of secondary plant substances. *Science (80-.).*,
639 129, 1466–1470.

640 Fritz, S.A. & Purvis, A. (2010). Selectivity in mammalian extinction risk and threat
641 types: A new measure of phylogenetic signal strength in binary traits. *Conserv.*
642 *Biol.*, 24, 1042–1051.

643 Gershenzon, J. (1994). Metabolic costs of terpenoid accumulation in higher plants. *J.*
644 *Chem. Ecol.*, 20, 1281–1328.

645 Gershenzon, J., Fontana, A., Burow, M., Wittstock, U. & Degenhardt, J. (2012). Mixtures
646 of plant secondary metabolites: metabolic origins and ecological benefits. *Ecol.*
647 *Plant Second. Metab. From Genes to Glob. Process.*, 335.

648 Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018).
649 UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*, 35,
650 518–522.

651 Jones, C.G., Firn, R.D. & B, M.S. (1991). On the evolution of plant secondary chemical
652 diversity. *Philos. Trans. - R. Soc. London, B*, 333, 273–280.

653 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A. & Jermiin, L.S.
654 (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat.*
655 *Methods*, 14, 587–589.

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

659 Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D.,
660 *et al.* (2010). Picante: R tools for integrating phylogenies and ecology.
661 *Bioinformatics*, 26, 1463–1464.

662 Kessler, A. & Kalske, A. (2018). Plant secondary metabolite diversity and species
663 interactions. *Annu. Rev. Ecol. Evol. Syst.*, 115-138.

664 Koenen, E.J.M., Ojeda, D.I., Steeves, R., Migliore, J., Bakker, F.T., Wieringa, J.J., *et al.*
665 (2020). Large-scale genomic sequence data resolve the deepest divergences in the
666 legume phylogeny and support a near-simultaneous evolutionary origin of all six

667 subfamilies. *New Phytol.*, 225, 1355–1369.

668 Koricheva, J., Nykänen, H. & Gianoli, E. (2004). Meta-analysis of trade-offs among plant
669 antiherbivore defenses: Are plants jacks-of-all-trades, masters of all? *Am. Nat.*, 163.

670 Kreis, W. & Munkert, J. (2019). Exploiting enzyme promiscuity to shape plant
671 specialized metabolism. *J. Exp. Bot.*, 70, 1435–1445.

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

675 Kursar, T.A. & Coley, P.D. (2003). Convergence in defense syndromes of young leaves
676 in tropical rainforests. *Biochem. Syst. Ecol.*, 31, 929–949.

677 Kursar, T.A., Dexter, K.G., Lokvam, J., Pennington, R.T., Richardson, J.E., Weber,
678 M.G., *et al.* (2009). The evolution of antiherbivore defenses and their contribution to
679 species coexistence in the tropical tree genus Inga. *Proc. Natl. Acad. Sci. U. S. A.*,
680 106, 18073–18078.

681 Maron, J.L., AGRAWAL, A.A. & Schemske, D.W. (2019). Plant–herbivore coevolution
682 and plant speciation. *Ecology*, 100, e02704.

683 Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., *et al.*
684 (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement
685 of modular polyketide synthase genes. *Nat. Biotechnol.*, 23, 1171–1176.

686 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von
687 Haeseler, A., *et al.* (2020). IQ-TREE 2: New models and efficient methods for
688 phylogenetic inference in the genomic era. *Mol. Biol. Evol.*, 37, 1530–1534.

689 Moghe, G.D. & Last, R.L. (2015). Something old, something new: Conserved enzymes
690 and the evolution of novelty in plant specialized metabolism. *Plant Physiol.*, 169,
691 1512–1523.

692 Monson, R.K., Trowbridge, A.M., Lindroth, R.L. & Lerdau, M.T. (2021). Coordinated
693 resource allocation to plant growth–defense tradeoffs. *New Phytol.*, 233, 1051–1066.

694 Moore, B.D., Andrew, R.L., Külheim, C. & Foley, W.J. (2014). Explaining intraspecific
695 diversity in plant secondary metabolites in an ecological context. *New Phytol.*, 201,
696 733–750.

697 Nicholls, J.A., Pennington, R.T., Koenen, E.J., Hughes, C.E., Hearn, J., Bunnefeld, L., *et*
698 *al.* (2015). Using targeted enrichment of nuclear genes to increase phylogenetic
699 resolution in the neotropical rain forest genus Inga (Leguminosae: Mimosoideae).
700 *Front. Plant Sci.*, 6, 1–20.

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

703 Orme, D. (2013). The caper package: Comparative analyses in phylogenetics and

evolution in R. *R package version 5.2,* 1–36.

Panda, S., Kazachkova, Y. & Aharoni, A. (2021). Catch-22 in specialized metabolism: Balancing defense and growth. *J. Exp. Bot.*, 72, 6027–6041.

Revell, L.J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.*, 3, 217–223.

Richards, L.A., Dyer, L.A., Forister, M.L., Smilanich, A.M., Dodson, C.D., Leonard, M.D., *et al.* (2015). Phytochemical diversity drives plant-insect community diversity. *Proc. Natl. Acad. Sci. U. S. A.*, 112, 10973–10978.

Richards, L.A., Glassmire, A.E., Ochsenrider, K.M., Smilanich, A.M., Dodson, C.D., Jeffrey, C.S., *et al.* (2016). Phytochemical diversity and synergistic effects on herbivores. *Phytochem. Rev.*, 15, 1153–1166.

Salazar, D., Jaramillo, A. & Marquis, R.J. (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, M.V., Zuñiga, J.M.A., De Valpine, P., *et al.* (2018). Origin and maintenance of chemical diversity in a species-rich tropical tree lineage. *Nat. Ecol. Evol.*, 2, 983–990.

Schneider, G.F., Salazar, D., Hildreth, S.B., Helm, R.F. & Whitehead, S.R. (2021). Comparative Metabolomics of Fruits and Leaves in a Hyperdiverse Lineage Suggests Fruits Are a Key Incubator of Phytochemical Diversification. *Front. Plant Sci.*, 12.

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

Sedio, B.E., Parker, J.D., McMahon, S.M. & Wright, S.J. (2018). Comparative foliar metabolomics of a tropical and a temperate forest community. *Ecology*, 99, 2647–2653.

Sedio, B.E., Rojas Echeverri, J.C., Boya, C.A. & Wright, S.J. (2017). Sources of variation in foliar secondary chemistry in a tropical forest tree community. *Ecology*, 98, 616–623.

Sherman, D.H. (2005). The Lego-ization of polyketide biosynthesis. *Nat. Biotechnol.*, 23, 1083–1084.

Smith, S.A. & O'Meara, B.C. (2012). TreePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, 28, 2689–2690.

Strauss, S.Y., Rudgers, J.A., Lau, J.A. & Irwin, R.E. (2002). Direct and ecological costs of resistance to herbivory. *Trends Ecol. Evol.*, 17, 278–285.

Thompson, J.N. (2005). *The geographic mosaic of coevolution (interspecific interactions)*. University of Chicago Press, Chicago.

Valencia, R., Condit, R., Foster, R.B., Romoleroux, K., Munoz, G. V, Svenning, J.-C., *et al.* (2004). Yasuní forest dynamics plot, Ecuador. In: *Forest Diversity and dynamism: Findings from a large-scale plot network* (eds. Losos, E.C. & Leigh, J.G.). University of Chicago Press, Chicago, pp. 609–620.

Vleminckx, J., Salazar, D., Fortunel, C., Mesones, I., Dávila, N., Lokvam, J., *et al.* (2018). Divergent secondary metabolites and habitat filtering both contribute to tree species coexistence in the Peruvian Amazon. *Front. Plant Sci.*, 9, 836.

Volf, M., Segar, S.T., Miller, S.E., Isua, B., Sisol, M., Aubona, G., *et al.* (2018). Community structure of insect herbivores is driven by conservatism, escalation and divergence of defensive traits in Ficus. *Ecol. Lett.*, 21, 83–92.

Wang, M., Carver, J.J., Phelan, V. V, Sanchez, L.M., Garg, N., Peng, Y., *et al.* (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.*, 34, 828–837.

Wetzel, W.C. & Whitehead, S.R. (2019). The many dimensions of phytochemical diversity: Linking theory to practice. *Ecol. Lett.23.1 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.

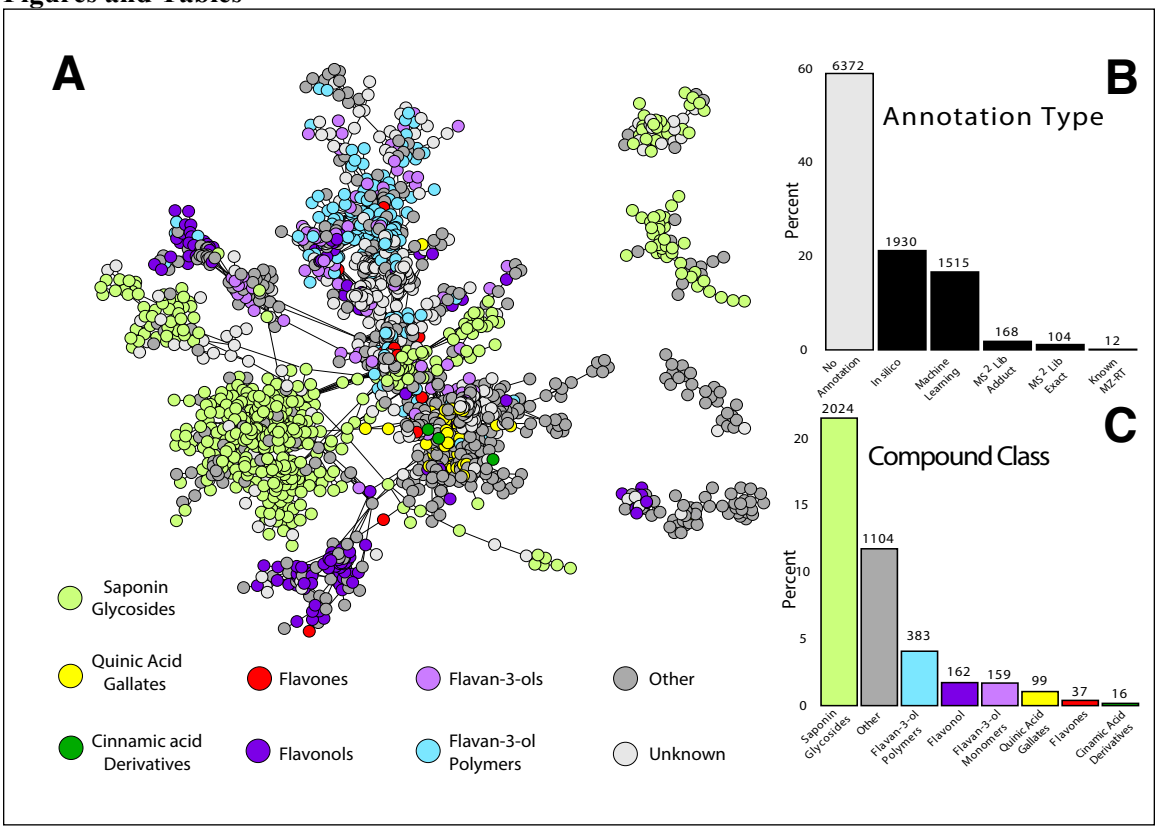
Wiggins, N.L., Forrister, D.L., Endara, M.J., Coley, P.D. & Kursar, T.A. (2016). Quantitative and qualitative shifts in defensive metabolites define chemical defense investment during leaf development in Inga, a genus of tropical trees. *Ecol. Evol.*, 6, 478–492.

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

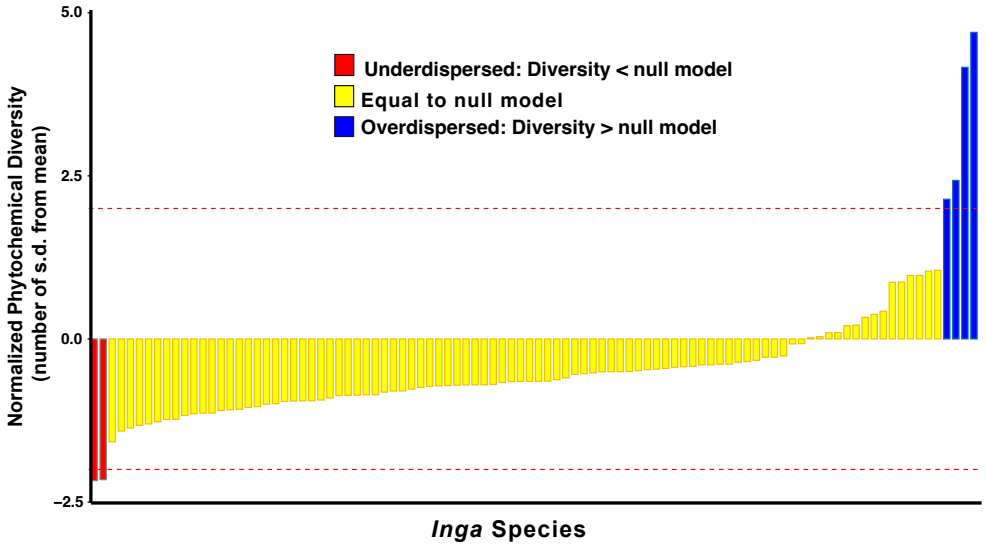
Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19, 15–30.

768

Figures and Tables



769 **Figure 1:** Compound based molecular network: (A) Subset of molecular network (see Fig. S2. for
770 the full network) containing all compounds observed across 98 study species. Nodes represent
771 individual compounds identified in the metabolomics pipeline, and connections between
772 compounds (edges) are based on the MS/MS cosine similarity score from GNPS
773 (https://gnps.ucsd.edu). (B) Percent of compounds that were annotated using different methods -
774 *in silico* fragmentation, machine learning, MS/MS library exact matches and adducts, and
775 comparison to authentic standards on our UPLC-MS system based on mass-charge ratio (*m/z*) and
776 retention time (RT). (C) Percent of compounds with annotations represented by each compound
777 class. For B and C, total number of compounds are reported at top of bars.



778 **Figure 2:** Normalized phytochemical diversity in each *Inga* species: Bars represent individual
779 *Inga* species ordered by increasing phytochemical diversity measured as the functional Hill
780 number. Values represent the number of standard deviations above or below the mean calculated
781 in the null model, with dashed red lines indicating 2 standard deviations above and below the null
782 mean. Values less than zero represent species that are less chemically diverse than a random draw
783 (under-dispersed in the MS/MS network) and values above zero represent species that are more
784 diverse (over-dispersed in the MS/MS network). Hill numbers are calculated with Q=0.

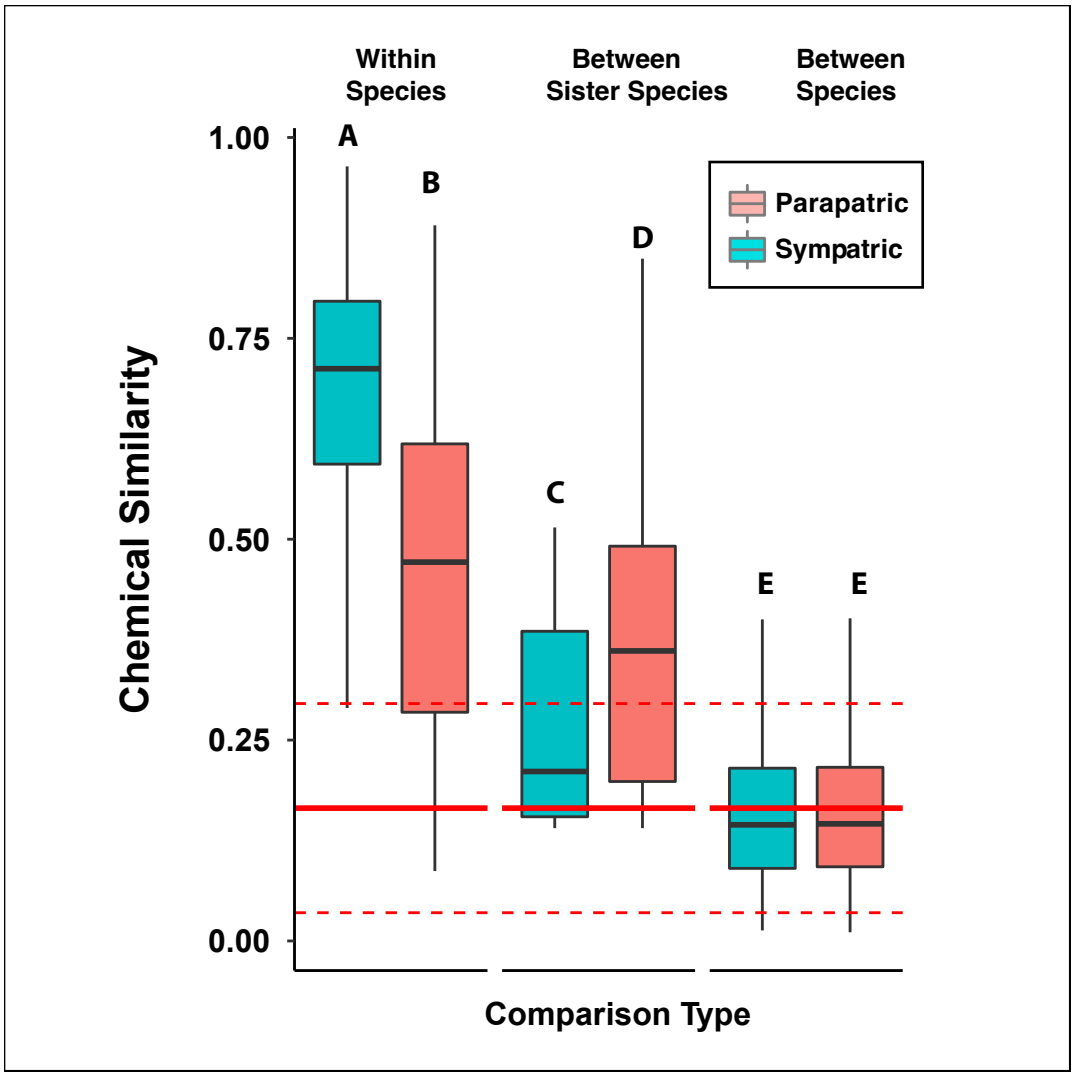


Figure 3: Comparison of entire chemical profiles between *Inga* Species: A) Boxplot comparison of chemical similarity scores for *Inga* within a species, between sister species, and between all other species. Comparisons between and within sites are indicated by red and blue boxes, respectively. Significantly different groups are denoted by A, B, C, D, and E. 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 2 standard deviations above and below the null mean.

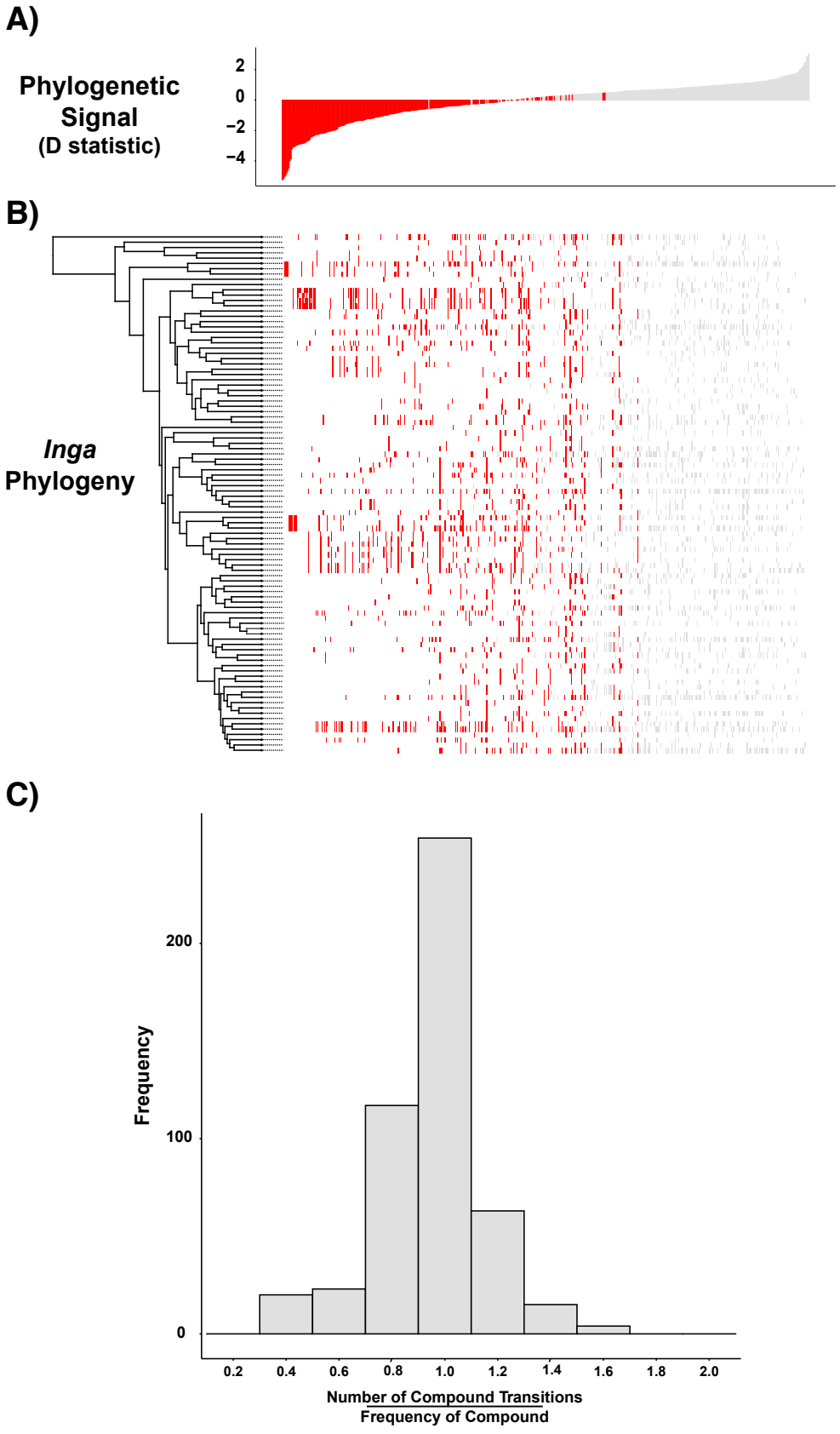

A)
Phylogenetic Signal
(D statistic)
2
0
−2
−4
B)
Inga
Phylogeny
C)
Frequency
200
100
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Number of Compound Transitions
Frequency of Compound

Figure 4. Expression patterns of individual compounds mapped onto the *Inga* phylogeny: (A) Phylogenetic signal of 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 phylogeny. Red (significant phylogenetic signal) and grey (non-significant) 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.

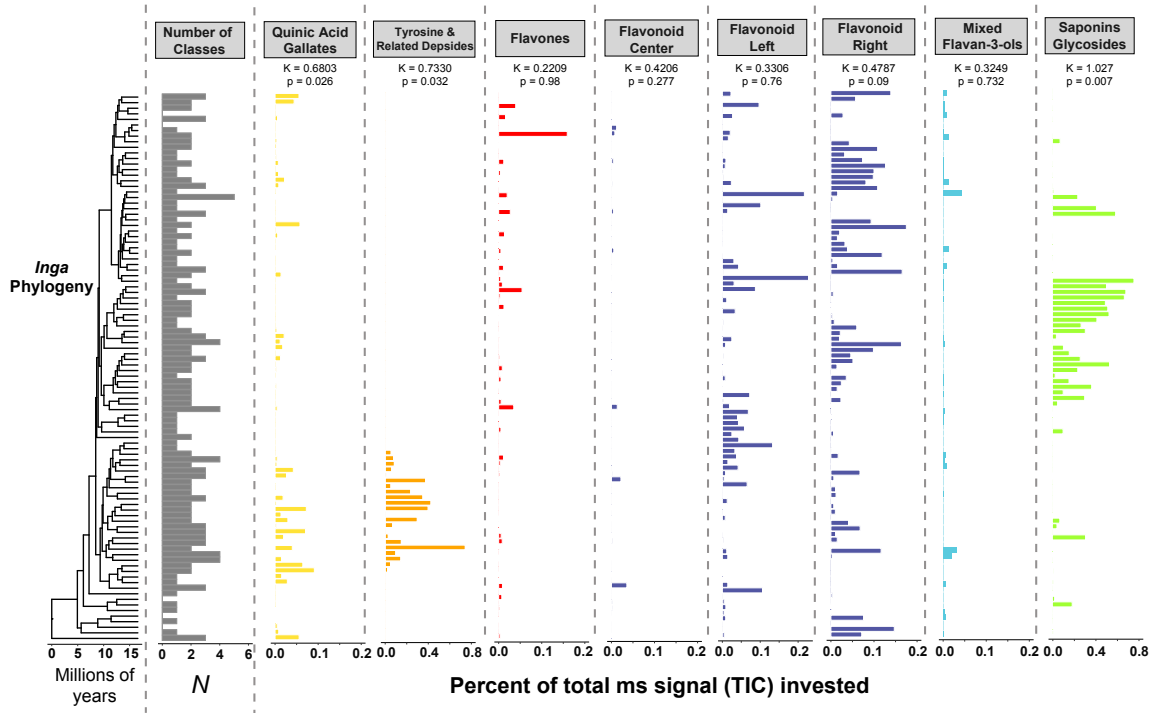


Fig. 5 Expression of defensive compound classes mapped on the *Inga* phylogeny: Total number of compound classes expressed per species, followed by expression per species of distinct classes including quinic acid gallates, tyrosine and related depsides, flavones, flavonoids, flavan-3-ols, and saponins. The expression of individual compound classes is measured as a percentage of the total MS-level-1 ion current (TIC; 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.

New Phytologist

***New Phytologist* Supporting Information**

Article title: Diversity and Divergence: Evolution of defense chemistry in the tropical tree genus *Inga*

Authors: Dale L. Forrister, María-José Endara, Abrianna J. Soule, Gordon C. Younkin, Anthony G. Mills, John Lokvam Kyle G. Dexter, R. Toby Pennington, Catherine A. Kidner, James A. Nicholls, Oriane Loiseau, Thomas A. Kursar, Phyllis D. Coley
Article acceptance date: Click here to enter a date.

The following Supporting Information is available for this article:

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

Fig. S2 Expression of major defensive compound classes mapped on the *Inga* phylogeny and SEM based correlation among them.

Fig. S3 Compound based molecular network containing all compounds observed in 98 study species

Fig. S4 Correlation between chemical similarity and phylogenetic distance (My) for all interspecific comparisons

Fig. S5 Biosynthetic context of phenolic compounds in *Inga*

Fig. S6 Comparison of Phytochemical diversity with different Q values.

Fig. S7 Illustration of Lego-chemistry concept

Table S1 Site and Sampling information for all 98 study species

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

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

Site	Country	Latitude	Longitude	Annual Rainfall (mm)	*Inga* Species (n)
Barro Colorado Island	Panama	9°S	80°W	2623	14
Nouragues	French Guiana	4°N	53°W	3000	46
Tiputini (Yasuni National Park)	Ecuador	0°N	75°W	3200	41
Los Amigos (Madre de Dios)	Peru	13°S	70°W	2648	39
Manaus	Brazil	2°S	60°W	2100	29

Table S1 Site and Sampling information for all 98 study species

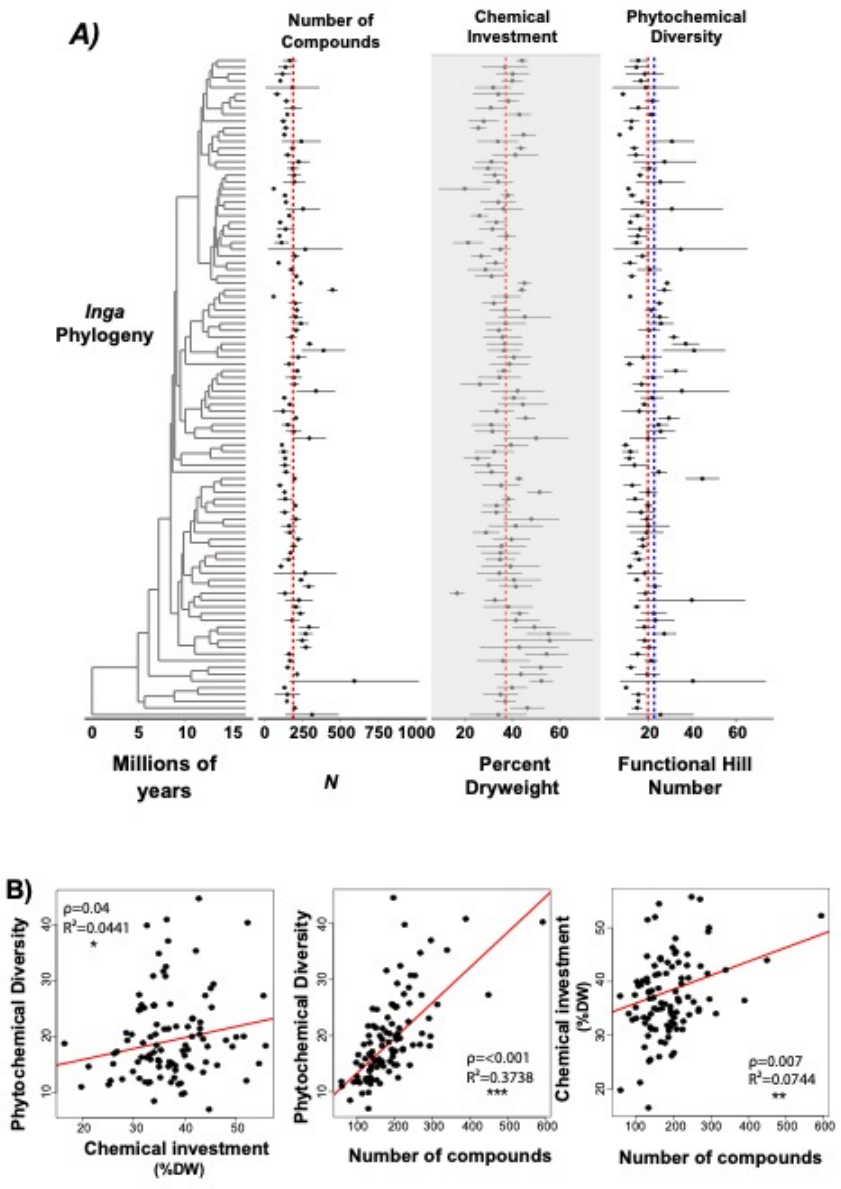


Fig. S1 A) Defense investment traits mapped on to the *Inga* phylogeny . Number of unique compounds per species, percent of leaf dry weight invested in secondary metabolism per species, and the phytochemical diversity (measured as functional Hill numbers, q = 2) of each species profile are represented by points. Horizontal bars indicate one standard deviation. Dotted red lines represent mean trait values across all species and the blue line represents the mean value for

phytochemical diversity estimated in the null model. **B)** Defense investment trait correlations: (1) Phytochemical diversity vs. percent of leaf dry weight invested in secondary metabolism, (2) phytochemical diversity vs. number of compounds, and (3) percent of leaf dry weight invested in secondary metabolism vs. number of compounds. Points represent individual *Inga* species; red lines represent the phylogenetic linear model estimate of best fit (package: phylolm[1]). Pearson's correlation (ρ), and R-squared are reported, and significance of model fit is represented by asterisks ($p < 0.05$ = *; $p < 0.01$ = **; $p < 0.001$ = ***)

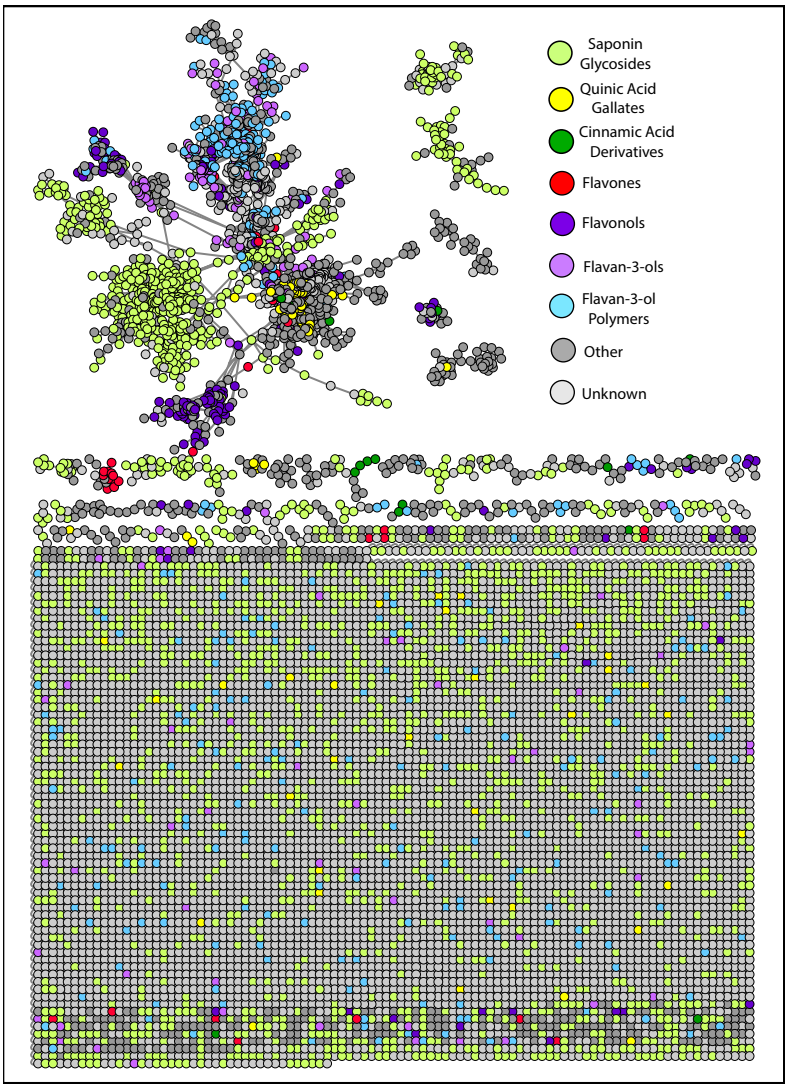


Fig. S2 Compound based molecular network containing all compounds observed in 98 study species. Nodes represent individual compounds identified in the metabolomics pipeline, and connections between compounds (edges) are based on MS/MS cosine similarity score from GNPS (https://gnps.ucsd.edu). Node color represents compound annotations into major

compound classes. Unconnected nodes at the bottom of the network are spectrally unique compounds, that did not match with any other compound in the network, a common feature of ms/ms based metabolomics studies.

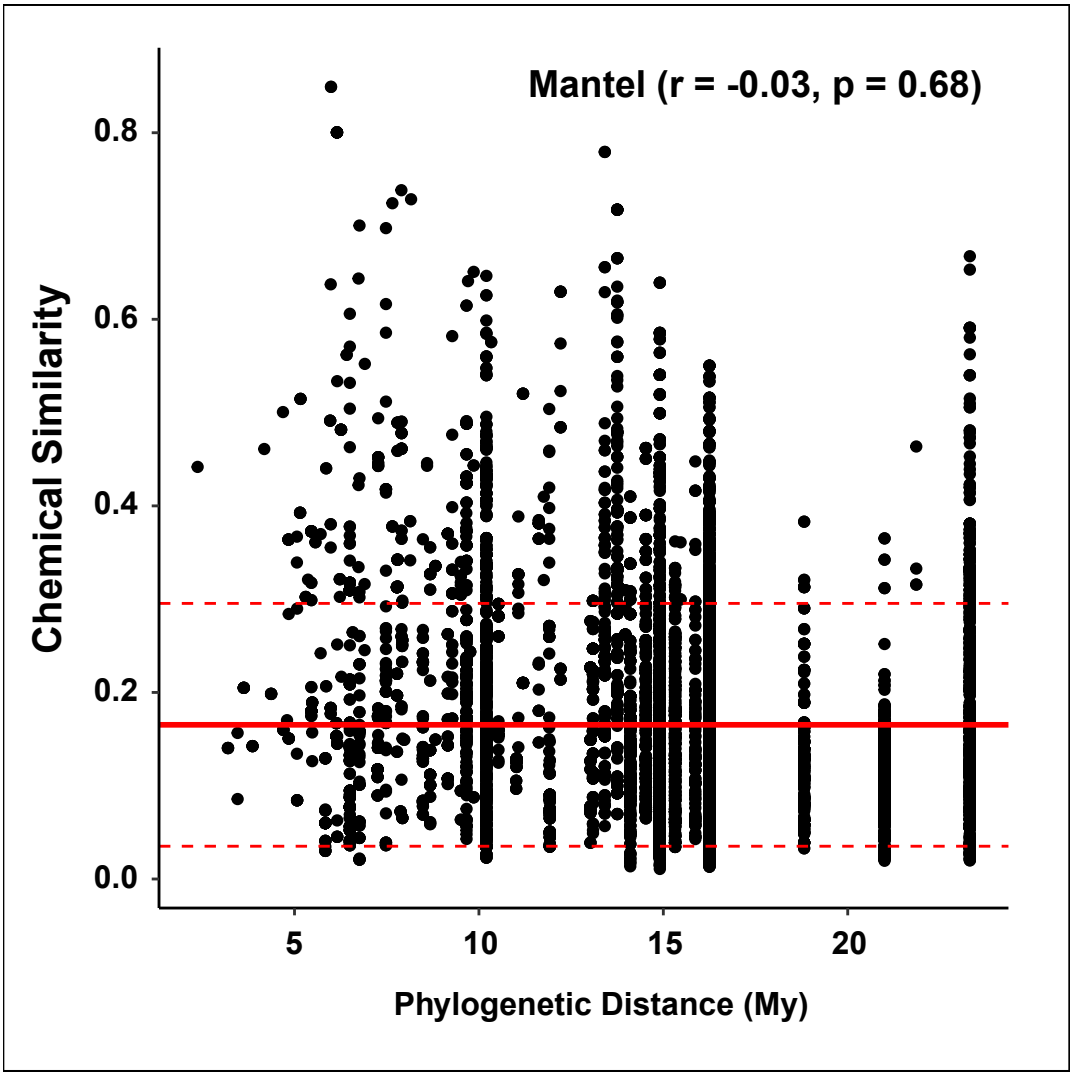


Fig. S3 Correlation between chemical similarity and phylogenetic distance (My) for all interspecific comparisons. The solid red line represents the mean chemical similarity score observed in the null model which simulates the expected chemical similarity between two randomly assembled chemical profiles. The dashed red lines represent 2 standard deviations above and below the null mean.

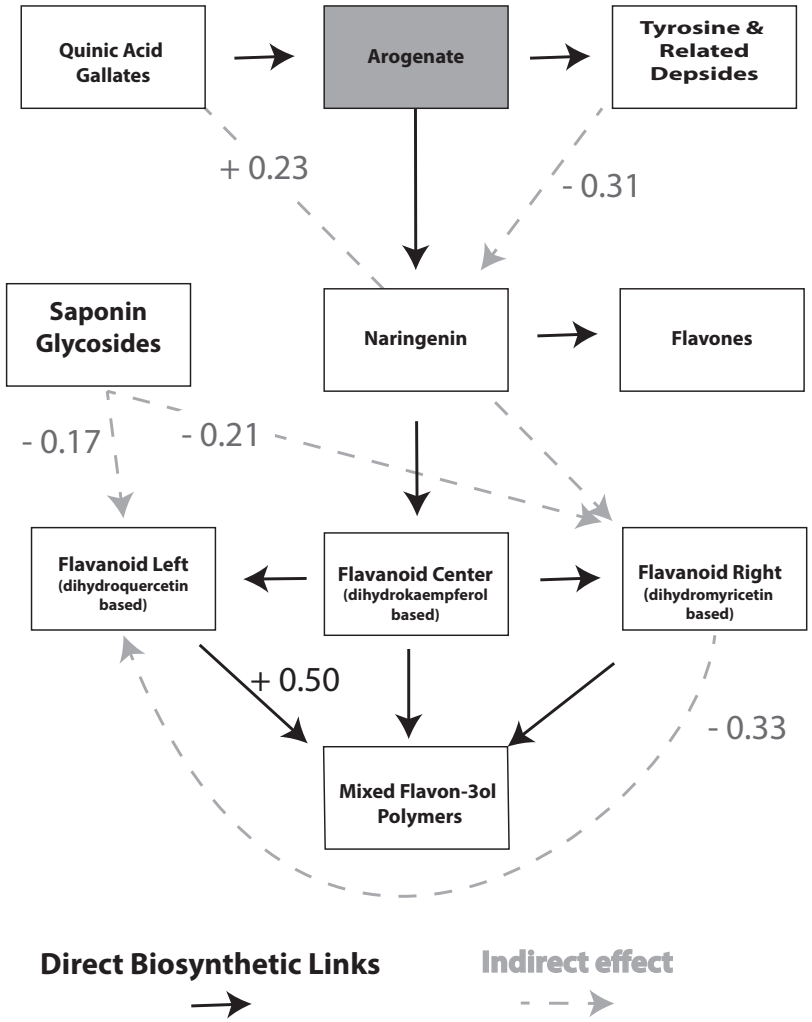


Fig. S4 Structural Equation Model (SEM) showing correlation between investment in major defense compound classes produced by *Inga*. Significant ($p < 0.05$) relationships between compound classes indicated by a correlation value listed next to arrows. Solid black arrows represent direct biosynthetic links, regardless of significance of the correlation in the SEM. Dashed grey arrows represent significant indirect relationships between compound classes.

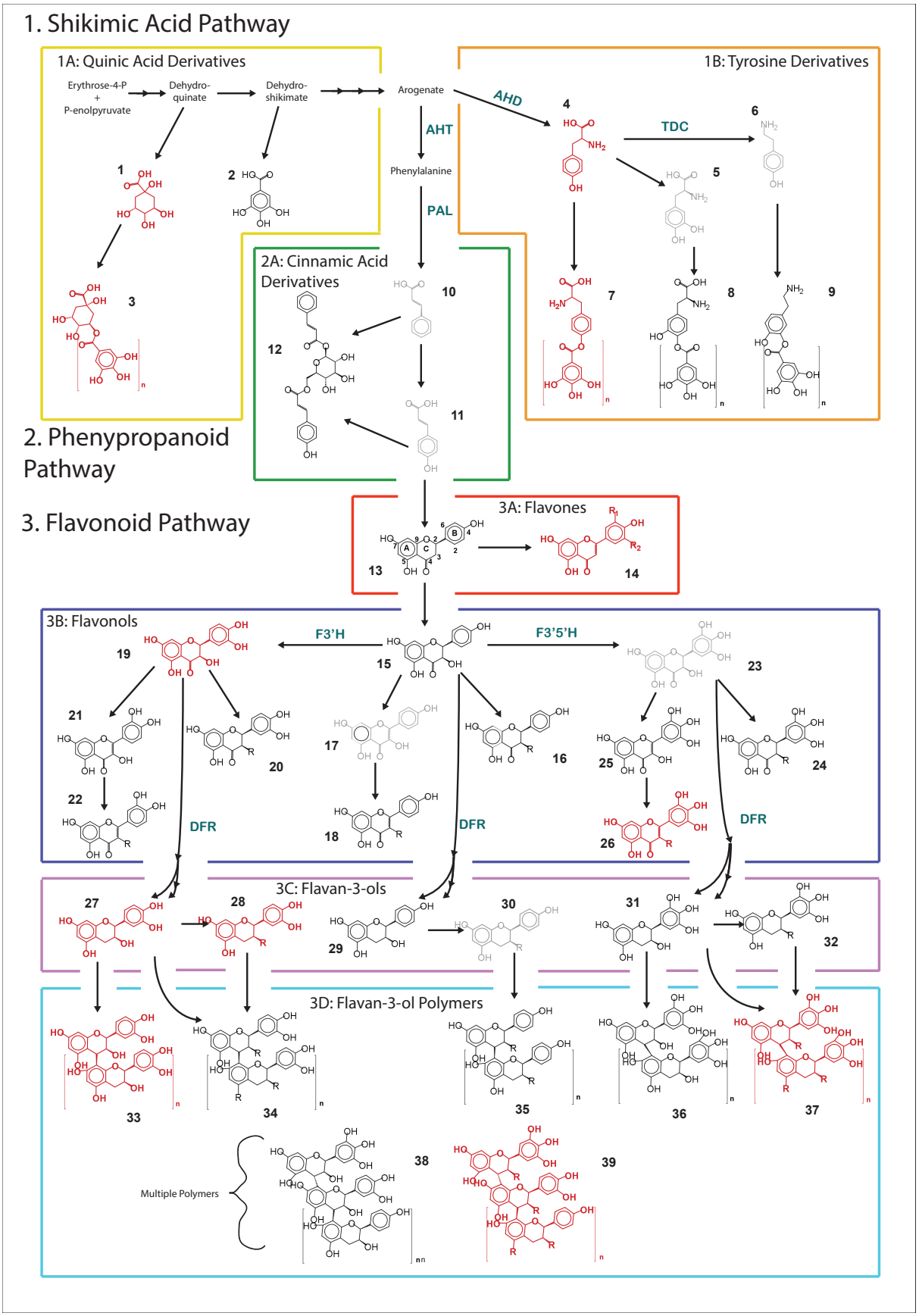

1. Shikimic Acid Pathway
1A: Quinic Acid Derivatives
Erythrose-4-P + P-enolpyruvate
Dehydro-quinate
Dehydro-shikimate
Arogenate
AHD
AHT
PAL
TDC
Phenylalanine
1B: Tyrosine Derivatives
2A: Cinnamic Acid Derivatives
2. Phenypropanoid Pathway
3. Flavonoid Pathway
3A: Flavones
3B: Flavonols
F3'H
F3'5'H
DFR
3C: Flavan-3-ols
3D: Flavan-3-ol Polymers
Multiple Polymers

Fig. S5 Biosynthetic context of phenolic compounds in *Inga*: (A) Structures and substructures of compounds observed in a survey of 98 focal species and their positions in the biosynthetic pathways that produce them. Compounds that accumulate to significant levels are red and bold; low abundance are black and non-accumulating intermediates are light grey. Wavy bonds indicate variable stereochemistry. Compound names for each compound are listed in Table S3. Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, Marvin 20.20.0, ChemAxon (https://www.chemaxon.com)

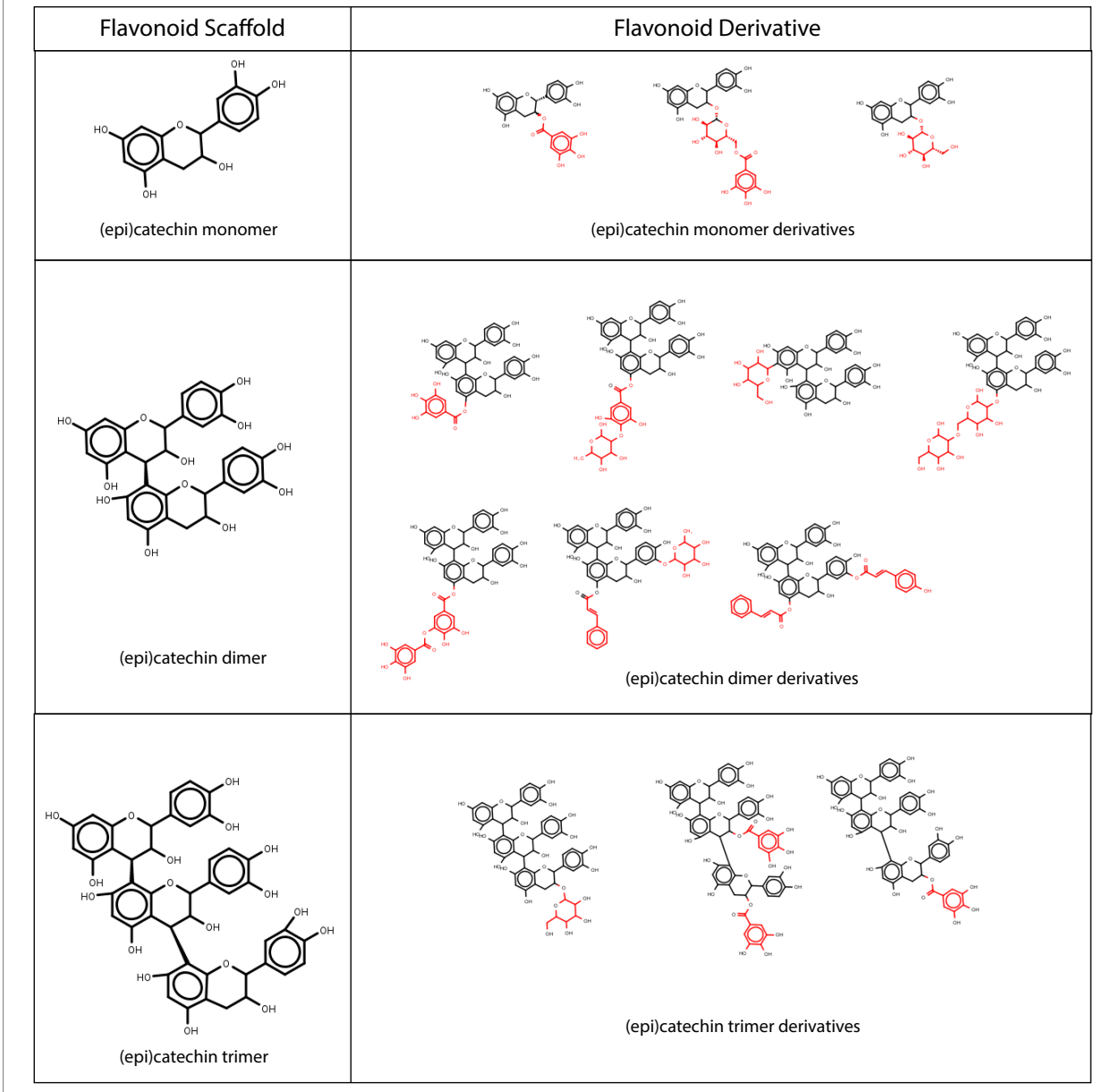


Fig. S6 Illustration of Lego-chemistry concept based on annotation of monomeric and polymeric Flavan-3-ol compounds observed in *Inga* based on NMR structure elucidation and MS/MS annotation. Red substructures represent commonly observed R-Groups, which are added in a combinatorial manner to generate a variety of compounds.

Trait	Phyl. Signal	Evol. Model	MLE	Δ AIC	Akaike Weight	P	Interpretation
Chemical Profile (Chemotype)	No Phyl. Signal Mantel R = -0.03, p = 0.68	DA	σ^2 =14.01, α = 0.24 psi = 0.83	0	1.0	***	Trait lacks phylogenetic signal and is evolving by divergent adapation.
		OU	σ^2 =0.13, α = 0.13	17823	0	***	
		BM	σ^2 =0.02	17989	0		
Number of Compounds	Significant Phyl. Signal K= 0.56, P= 0.04	BM	σ^2 =352.9	0	0.51		Trait shows moderate phylogenetic signal and is evolving under Brownian motion
		OU	σ^2 =372.43, α = 0.004	0.71	0.36	ns	
		DA	σ^2 = 372.43, α = 0.004, psi = 3.054	2.72	0.13	ns	
Chemical Investment (% Dry Weight)	Marg. Signif. Phyl Signal K= 0.51, P= 0.06	OU	σ^2 =5.02, α = 0.03	0	0.73	***	Trait shows moderate phylogenetic signal and is evolving towards an optimal value.
		DA	σ^2 =5.02, α = 0.03, psi = 0.00001	2	0.26	ns	
		BM	σ^2 = 3.19	66	0		
Phytochemical Diversity	No Phyl. Signal K= 0.36, P= 0.58	OU	σ^2 =16.1, α = 0.12	0	0.72	***	Trait lacks phylogenetic signal and is evolving toward an optimal value.
		DA	σ^2 =16.21, α = 0.12 psi = 1.33	1.9	0.27	ns	
		BM	σ^2 =4.34	372	0		

Table S2 Maximum-likelihood estimates for different evolutionary models of trait evolution. For each trait we fit three models of trait evolution: A random walk model characterized by Brownian Motion (BM), The Ornstein-Uhlenbeck (OU) model where a trait evolves under BM with a constraining central tendency, and a divergent adaptation (DA) model where trait values the OU model but different lineages interact such that lineage's mean values diverge. We selected the best model based on AIC; significance of model parameters was evaluated by likelihood ratio (LR) tests to determine if a more complex model was significant. Significance indicated by asterisks ($p < 0.05$ = *; $p < 0.01$ = **; $p < 0.001$ = ***).