nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-018-0552-0

Origin and maintenance of chemical diversity in a species-rich tropical tree lineage

Diego Salazar[1,2*], John Lokvam[2], Italo Mesones[2], Magno Vásquez Pilco[3,4], Jacqueline Milagros Ayarza Zuñiga[5,6], Perry de Valpine[7] and Paul V. A. Fine[2,8,9]

Plant secondary metabolites play important ecological and evolutionary roles, most notably in the deterrence of natural enemies. The classical theory explaining the evolution of plant chemical diversity is that new defences arise through a pairwise co-evolutionary arms race between plants and their specialized natural enemies. However, plant species are bombarded by dozens of different herbivore taxa from disparate phylogenetic lineages that span a wide range of feeding strategies and have distinctive physiological constraints that interact differently with particular plant metabolites. How do plant defence chemicals evolve under such multiple and potentially contrasting selective pressures imposed by diverse herbivore communities? To tackle this question, we exhaustively characterized the chemical diversity and insect herbivore fauna from 31 sympatric species of Amazonian Protieae (Burseraceae) trees. Using a combination of phylogenetic, metabolomic and statistical learning tools, we show that secondary metabolites that were associated with repelling herbivores (1) were more frequent across the Protieae phylogeny and (2) were found in average higher abundance than other compounds. Our findings suggest that generalist herbivores can play an important role in shaping plant chemical diversity and support the hypothesis that chemical diversity can also arise from the cumulative outcome of multiple diffuse interactions.

Tropical plant species contain dozens of diverse secondary metabolites comprising up to half of their dry weight[1]. The current understanding of the mechanisms involved in the origin and maintenance of the vast diversity of plant secondary chemistry is based on the escape and radiate hypothesis (E&R)[2], which links the rise of new plant species to the escape from herbivores via novel plant secondary metabolites. The evolution of highly effective chemical defences is hypothesized to significantly reduce the attack from herbivores, greatly increasing plant fitness, and giving a species the opportunity to enter a new adaptive zone and expand its ecological and geographic distribution[2]. Similarly, counter-adaptations by specialized herbivores would allow them to overcome or circumvent these 'silver bullet' defences. This co-evolutionary arms race is thought to represent the evolutionary machinery responsible for the immense diversity of plant natural products, and spur diversification in both plants and insects[3–5].

However, most studies testing E&R have focused on the interaction of a single class of metabolites across one or a small group of plant species, and a few specialist herbivore species[6,7] (but see recent work by ref. [8]). Yet, plants are confronted with a multitude of natural enemies with diverse ecologies and metabolic constraints, probably selecting for different defences simultaneously, and exerting distinct and potentially contrasting selection pressures on the same defence traits through a process called diffuse co-evolution[9–12]. Furthermore, as particular herbivore species independently overcome[13] or, in some cases, adapt to use to their advantage specific secondary plant metabolites (for example, semiochemical cues or toxin sequestration[14,15]), defensive traits might not only lose effectiveness through time, but actually attract herbivores, thus increasing herbivore pressure, leading to negative selection across the host phylogeny. Therefore, to accurately assess the role of insect herbivores in the evolution of plant chemical diversity, one needs to not only holistically evaluate a plant's secondary metabolite complexity and herbivore diversity, but also assess the impact of each chemical defence across each herbivore species.

To better elucidate the mechanisms that mediate the evolution of plant chemical defences under the selective pressures of equally diverse herbivore communities, we conducted a detailed fine-scale study of the chemical diversity and plant–herbivore interactions of 31 sympatric tree species from the monophyletic clade Protieae (Burseraceae) from the lowland rainforest in the Allpahuayo-Mishana National Reserve (Iquitos, Peru).

Here, we investigate how secondary chemical diversity of Protieae is related to plant–herbivore interactions. Specifically, we tackle this objective using four different approaches. First, we use our field collection data to determine the relationship between the diversity and abundance of secondary metabolites across Protieae species and the diversity and abundance of their insect herbivores. Second, we use phylogenetic approaches to test for the presence of broad co-evolutionary patterns between plants and insect herbivores. Third, we use a variable selection approach to determine which specific Protieae secondary metabolites are associated with a reduction or increase in attack rates from specific herbivore species and across all herbivores. Finally, we evaluate whether plant secondary metabolites associated with an overall reduction of herbivore pressure are more frequent across the Protieae phylogeny, and whether they

[1]Department of Biological Sciences, International Center for Tropical Botany, Florida International University, Miami, FL, USA. [2]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, USA. [3]Facultad de Ciencias Forestales, Universidad Nacional de la Amazonía Peruana, Iquitos, Peru. [4]Programa de Pós-Graduação em Botânica, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas, Brazil. [5]Facultad de Ciencias Biológicas, Universidad Nacional de la Amazonía Peruana, Iquitos, Peru. [6]Programa de Pós-Graduação em Entomologia, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas, Brazil. [7]Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, USA. [8]University and Jepson Herbaria, University of California, Berkeley, Berkeley, CA, USA. [9]Essig Museum of Entomology, University of California, Berkeley, Berkeley, CA, USA. *e-mail: dsalazar@fiu.edu

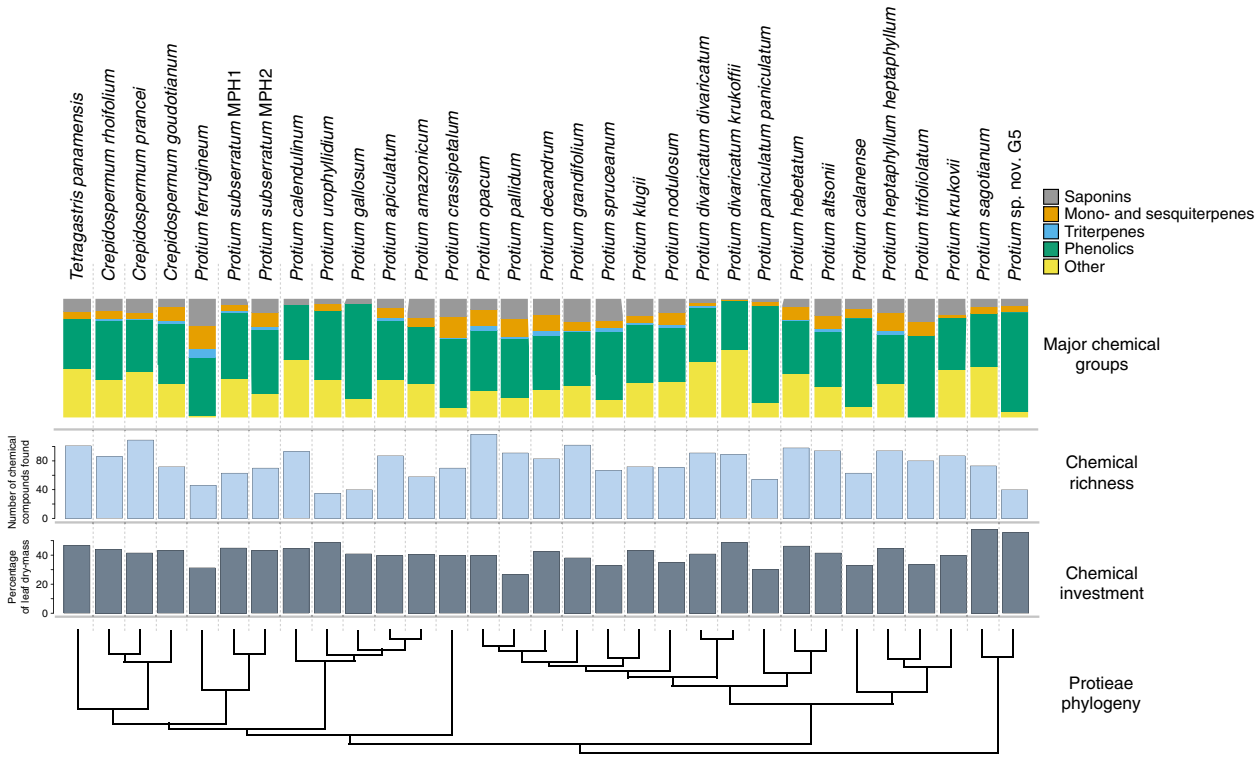

Fig. 1 | Chemical diversity and investment of 31 Protieae species arranged by their phylogenetic relationships. The size of the multicoloured bars at the top represents the proportion of compounds detected for each chemical group. Chemical richness is shown by the number of distinct chemical compounds found. Chemical investment is shown by the percentage of a leaf dry-mass comprising secondary metabolites.

show phylogenetic patterns consistent with strong positive selection across the phylogeny.

Results and discussion

In one of the most comprehensive evaluations of the chemistry of any tree genus, we were able to differentiate and quantify close to 600 putative secondary metabolites among the 31 Protieae species. The Protieae's secondary chemical investment was high, averaging 40% of leaf dry weight (22–58%) and, like other tropical clades, was restricted to a subset of several broad chemical categories (Supplementary Fig. 1)[16–18], including phenolics (procyanidins, flavone glycosides and chlorogenic acids) and terpenoids (saponins, triterpenes and sesquiterpenes; Fig. 1).

We recorded 4,214 feeding records from 231 insect morphospecies (13 families, 5 orders; Supplementary Fig. 2). However, we restricted our analysis to insects that were recorded at least 20 times; these 50 morphospecies represented 87% of all feeding records (Supplementary Fig. 3). On average, herbivore species fed on 11 species of Protieae, and strict monophagy was extremely rare (Fig. 2). Nevertheless, all herbivore species showed strong preferences or avoidance for specific Protieae species (Fig. 2). While the lack of monophagy contrasts with traditional expectations of specialization in tropical systems, it concurs with previous large-scale studies that found most herbivores to be polyphagous yet restricted to a subset of species or a plant clade[19–21].

To assess the effect that the secondary chemistry of Protieae species could have on determining a host-species herbivore community, we examined the relationship between plant species chemical diversity and herbivore diversity. In agreement with long-standing predictions[22], overall chemical diversity and chemical abundance were negatively correlated with total herbivore richness and abundance after controlling for the potential phylogenetic non-independence of plant chemistry via a phylogenetic independent contrast analysis (regression coefficient: $r^2=0.45$, $F=19.67$, $P<0.001$ and $r^2=0.41$, $F=17.13$, $P<0.001$ for richness and abundance, respectively). These results confirm the role of plant chemistry on modulating plant–herbivore interactions (Fig. 3a,b and Supplementary Fig. 4, but see ref. [18]). However, despite the large effect that plant chemistry had on the size and diversity of a host-species herbivore community, overall host-plant chemistry was not associated with Protieae herbivore community assembly, even after controlling for the effect of phylogeny on plant chemistry (partial Mantel test: correlation coefficient: $r=0.06$, $P=0.23$).

To test for co-evolutionary patterns between plants and herbivores, we evaluated the relationship between plant chemical composition and the phylogenies of Protieae and its associated herbivores. For Protieae, overall chemical similarity between species was not associated with phylogeny (Mantel test: $r=0.07$, $P=0.06$). Similarly, insect herbivore evolutionary patterns were also uncorrelated with host chemical similarity (Mantel test: $r=-0.02$, $P=0.65$). Furthermore, although specific secondary compounds appear frequently across Protieae, very few compounds showed significant phylogenetic signal (Supplementary Fig. 5). Instead, the majority of compounds showed varying degrees of phylogenetic divergence (most values of Blomberg's $K<1$; Supplementary Fig. 5), challenging the general expectations from E&R. Nevertheless, the presence of metabolites unrelated to plant–insect herbivore interactions has probably introduced a significant amount of noise in these analyses, obscuring the phylogenetic and statistical signal of metabolites involved in herbivore defence.

Some metabolites represent effective defences against natural enemies, but others could have no effect on herbivores or could even be detrimental to plant fitness (for example, serving as semiochemical cues or granting protection to herbivores via sequestration). Therefore, to evaluate the overall evolutionary and ecological value of a particular plant metabolite on a particular

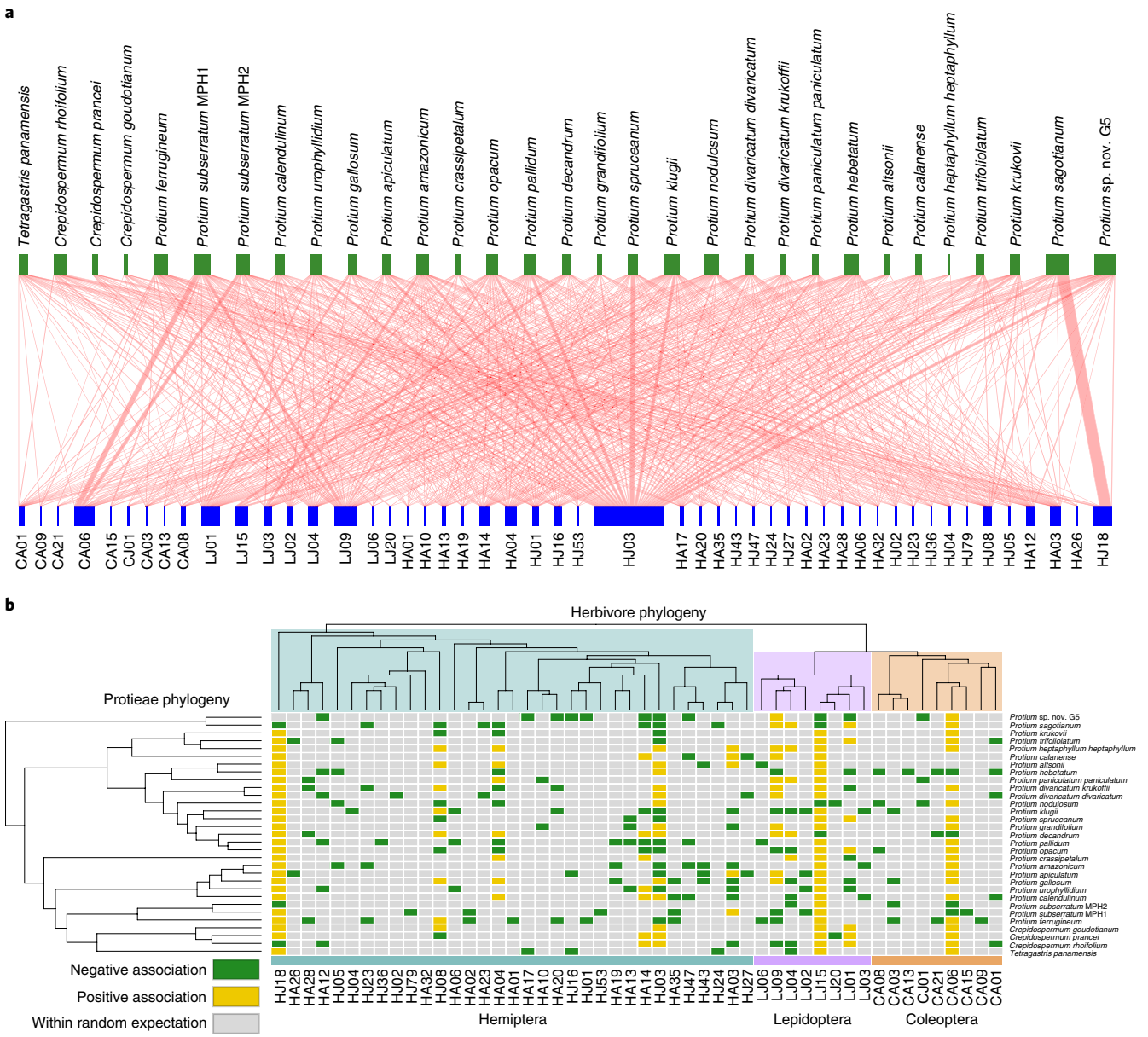

Fig. 2 | Protieae trees and their associated herbivores. a, Plant–herbivore interaction network based on the 50 most abundant Protieae herbivores. The width of the blue lower herbivore nodes represents the abundance of each herbivore species. The width of the green upper plant nodes represents the number of herbivore individuals collected per plant species. The width of the red lines represents the number of feeding events for each plant–herbivore combination. **b**, Herbivore–host-plant associations that occur at higher (green) or lower (yellow) rates than 3 s.d. from a random Poisson model are shown. Herbivore species in **a** and **b** are represented by their morphospecies code, for more information please see Supplementary Table 1.

plant host, one must estimate the effect of that metabolite on each herbivore—a formidable analytical challenge given the many thousands of possible interactions in a dataset comprising so many secondary metabolites, host plants and insect herbivore species. To overcome this challenge and identify specific metabolites associated with the attack rates from particular herbivore species (either positively; henceforth, 'attractant', or negatively; henceforth, 'repellent'), we used Poisson regression with variable selection via the least absolute shrinkage and selection operator (LASSO)—a shrinkage method calibrated by cross-validation commonly used for high-dimensional data[23]. Across the 50 most common herbivores, our approach identified 81 Protieae secondary metabolites (that is, herbivore active metabolites (HAMs)), representing 18% of Protieae's total chemical diversity (Supplementary Fig. 6). These compounds belong to multiple categories of secondary metabolites that are widely distributed across the Protieae clade, such as α- and β-amyrin and many types of procyanidins (Supplementary Materials). Almost all species of herbivores showed an association with at least 1 HAM, but no plant species contained more than 14 repellent compounds.

Given that the relationship between herbivore attack rates and HAMs was not consistent across all herbivore species studied (that is, attracting some herbivore species and repelling others), we calculated each HAM's overall impact on Protieae plant–herbivore interactions (that is, its 'net value') by averaging the estimated effect of each metabolite (that is, 'association strength') across all herbivore species, summing both negative and positive effects (Supplementary Materials). Consequently, Protieae species with metabolites with positive net values have lower total herbivore attack rates and vice versa.

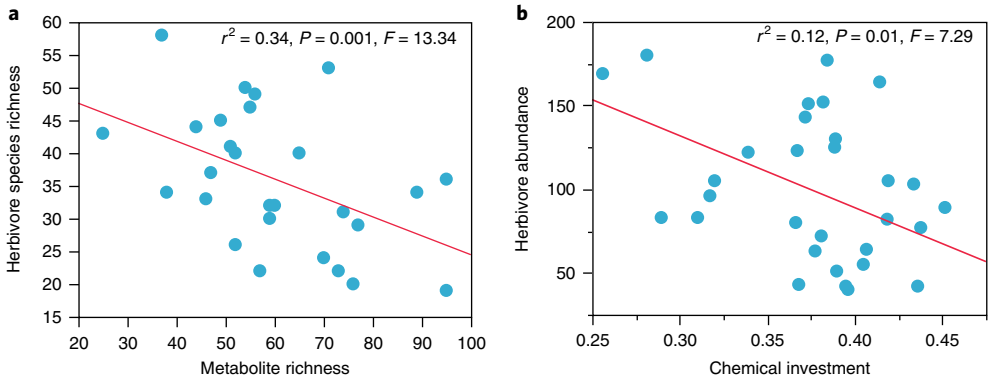

Fig. 3 | Effect of plant secondary chemistry on insect herbivores. Each point represents one Protieae species and the diagonal shows the regression line. **a**, Relationship between plant metabolite richness (number of secondary metabolites per Protieae species) and herbivore species richness (number of herbivore species per Protieae host). **b**, Relationship between chemical investment (proportion of leaf dry-mass comprising secondary metabolites) and herbivore abundance (number of herbivore individuals per Protieae host). The results of linear regressions are embedded.

Finally, we evaluated the role of plant–herbivore interactions on the evolution of HAMs. We found that HAMs with a higher net value (the overall effect of a metabolite across all herbivores) were significantly more frequent across the Protieae phylogeny (Fig. 4a). Furthermore, to determine the relationship between a metabolite net value and its mode and tempo of evolution, we tested how the variation of each HAM across the Protieae phylogeny fitted different macro-evolutionary models of trait evolution (see Methods). Of the 81 HAMs tested, 57 showed some level of directional selection (that is, asymmetrical rates of trait evolution). For these 57 HAMs, we estimated the rates of character gain and loss given the data. We found a very strong positive relationship between evolutionary rates and the metabolite net value, suggesting that HAMs with a strong association with reduced herbivore attacks also showed phylogenetic distribution patterns consistent with strong positive selection across the phylogeny (Supplementary Fig. 7). Additionally, we used a maximum parsimony character reconstruction approach to assess how conserved each HAM was across the Protieae phylogeny. Here, HAMs with a higher net value showed higher degrees of phylogenetic conservatism (Supplementary Fig. 8). Most notably, HAMs associated with higher rates of herbivore attack (low net values) were neither frequent nor highly conserved across the phylogeny. The strong relationship between a metabolite's association across all insect herbivores and its distribution throughout the Protieae phylogeny strongly supports the role of herbivores in the evolution of plant chemical defences as predicted by E&R. Furthermore, while total chemical similarity of host plants did not show any relationship with phylogenetic relatedness (see above), overall HAM composition showed a strong phylogenetic signal (Mantel test: $r = 0.24$, $P = 0.003$). In addition, Protieae with higher HAM richness expressed significantly lower total chemical investment, suggesting that HAMs may be more efficient than other metabolites, potentially allowing these plants to invest more resources in other metabolic processes and therefore further increasing their ecological and evolutionary value (Fig. 4b). Moreover, repellent HAMs were more abundant than

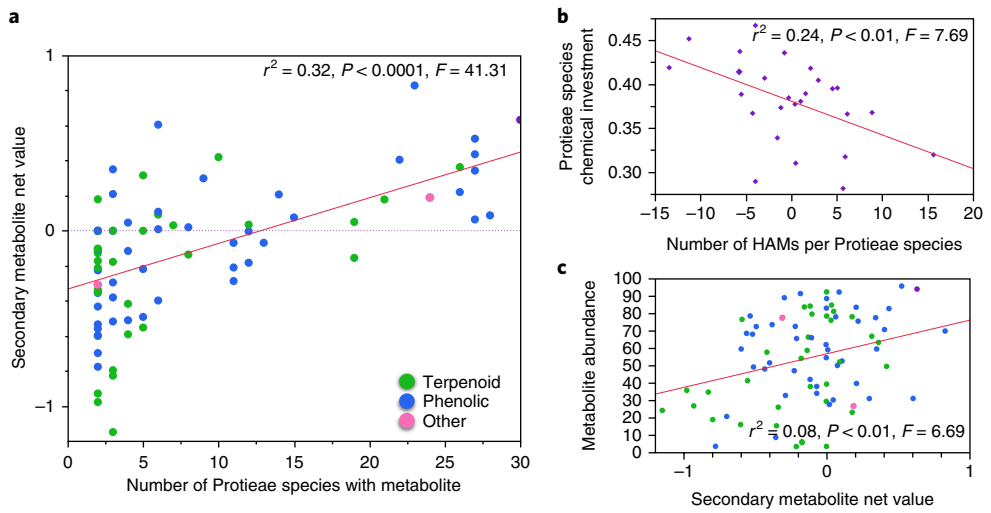

Fig. 4 | Association between plant–herbivore interactions and HAM phylogenetic frequency, plant chemical investment and metabolite abundance. **a**, Relationship between the number of Protieae species with a particular HAM and metabolite net value. Each point represents a HAM, with colours corresponding to different chemical categories (terpenoid, phenolic or other). **b**, Relationship between the number of HAMs per Protieae species (controlled for chemical richness) and their chemical investment (proportion of leaf dry-mass comprising secondary metabolites). Each diamond represents one Protieae species. **c**, Relationship between metabolite net value and average metabolite abundance (mass rank of each metabolite). Each point represents a HAM. The key is the same as in **a**. The results of linear regressions are embedded.

attractive HAMs, supporting the relationship between a compound's 'net value' and its ecological importance (Fig. 4c).

The strong effect of total chemical diversity on insect herbivores is probably linked to the positive association between the total number of chemicals and HAM diversity (Supplementary Fig. 10), as well as the potential synergistic effects between HAMs and other compounds[24]. Nevertheless, we note that the specific approach used here cannot determine whether the effect of HAMs on herbivore attack rates is due to the attraction of herbivore natural enemies (for example, parasitoids), rather than a direct effect on the actual herbivores. It is also important to emphasize that the great majority of the chemical compounds in our dataset are not HAMs and showed no direct correlation with insect herbivore feeding preferences. We speculate that these other compounds probably fall into four categories: (1) defences against fungal pathogens[25], (2) defences effective against herbivores in different parts of the plants' geographic range and maintained by gene flow[26], (3) secondary metabolites associated with abiotic factors and (4) non-effective compounds probably undergoing different degrees of negative selection. As proposed by the screening hypothesis[27], only a very small percentage of compounds derived from a particular metabolic pathway are likely to be effective against enemies at any given time. Accordingly, a plant's best strategy is to produce a great diversity of secondary metabolites in order to have enough effective compounds to significantly reduce total herbivore pressure. Our results agree with the predictions from this hypothesis, as well as many of the predictions of E&R.

Taken together, these findings represent robust and compelling evidence for the role of plant–herbivore interactions on the evolution of plant chemical diversity. Furthermore, our results indicate that the net evolutionary advantage granted by particular secondary metabolites does not necessarily depend on their effect on specific antagonistic species, but on the cumulative effect across all herbivores. Thus, HAMs should not be thought of as 'silver bullets'; instead they represent successful means to reduce the selective pressures imposed by a subset of the entire herbivore community. The rich chemical diversity found in plants probably arises because, although many compounds are active, each one is only effective against a few herbivore species, and plants are confronted by a multitude of different enemy taxa with diverse metabolic traits and evolutionary histories[28]. Furthermore, as herbivores overcome, circumvent or even exploit such metabolites, their net evolutionary value is lost and, consequently, such compounds experience negative selection and become less frequent across the clade.

We did not find a significant association between plant chemical composition and herbivore phylogeny—a pattern probably related to the almost complete absence of strict host-specialized herbivores in our system. Moreover, finding such strong patterns suggestive of selection for effective chemical defences in a system dominated by generalist herbivores strongly suggests that specialization onto a single host-plant species is not a requirement for insects to drive the evolution and maintenance of plant chemical diversity. Future studies on the genetics underlying chemical expression, together with manipulative experiments with active chemicals and important herbivore species, will give us a greater mechanistic understanding of the evolution of chemical diversity in plants.

Methods

Protieae herbivore surveys. We conducted this study in permanent forest transects where multiple plants from 31 species of Protieae were selected by P.V.A.F. and I.M., both of whom have over 15 years of experience identifying Protieae saplings (Fig. 2, main text). Each Protieae individual was georeferenced and marked with a unique and consecutive 'plant identification number'. To have a balanced sampling effort across all Protieae species, 18–25 individuals of each species were marked. A total of 860 Protieae plants were selected for this study. These plants were located within the Allpahuayo-Mishana National Reserve on transects that covered approximately 20 ha of forest near existing trails between km 24 and km 28 of the Iquitos–Nauta highway (see ref. [29] for a detailed map of the study area). Each plant was located at least 3 m (usually 5–10 m) away from any other plant included in the study. Once marked, all experimental individuals were censused for herbivores at least once a week for 64 consecutive weeks (May 2013 to September 2014). Each plant census included two steps: observation and collection. The observation step was implemented with the objective of only recording and collecting herbivores that were observed feeding on the host plant. Herbivores present on the host plant but not observed to feed or show obvious signs of leaf, stem or vascular fluid consumption were not collected or recorded. Herbivores that were observed to be consuming plant material were carefully collected by hand, net and suction traps, and immediately placed in vials. Collected herbivores were fixed in 80% ethanol and later assigned to a morphospecies using a reference collection. The herbivore surveys yielded 231 morphospecies and a total of 4,214 feeding events were recorded. Of all morphospecies in the dataset, 90 were singletons (found only once) and the 50 most abundant species (hereafter, the 'top 50 herbivores') accounted for 87% of total records. Our sampling scheme was designed to accurately assess the quantitative and qualitative herbivore pressure experienced by our 31 focal host-plant species, and not to determine the 'complete' diet breadth of Protieae feeding herbivores.

Insect molecular work. Insect DNA barcoding was conducted at the species level. For each putative insect species (morphospecies), we sequenced at least 12 individuals (more samples were used for recalcitrant species). A small sample from each insect (a leg or a small fraction of the abdomen) was washed with distilled water and dried to remove all traces of ethanol. DNA extraction was performed via proteinase digestion. Polymerase chain reaction amplification was accomplished using standard cytochrome oxidase subunit 1 (COI) primers: LCO1490/HCO2198 (ref. [30]) and LepF1/LepR1 (ref. [31]). Alignment was done using the MUSCLE algorithm[32]. To determine the best substitution model for our data, we used the R package Phangorn via the Akaike information criterion[33]. Phylogenetic reconstruction was carried out using 10 independent runs under a general time reversible model, gamma-distributed and invariable sites (GTR + G + I) model for 3 million generations (four chains, 25% burn-in, sampled every 100 generations) using MrBayes 3.2.6 (ref. [34]) on the public Cipress Science Gateway servers[35,36]. See complete details in Supplementary Materials.

Plant chemistry. *Chemical analysis.* Chemical analysis was conducted at the species level (see Supplementary Materials for details). Between six and nine different individuals per species were analysed (more samples were used for recalcitrant species). Young and mature leaf samples were collected in silica gel from the same marked individuals within the permanent transects and transported to the University of California, Berkeley for analysis. To characterize the maximum number of secondary compounds, we performed separate analyses for high- and low-molecular-weight metabolites. All extraction and chromatographic methods were developed de novo using as a starting point the work of D.S.[37,38] and J.L.[1,39,40].

Low-molecular-weight metabolite chemistry. Sample extraction. For low-molecular-weight metabolites, 100 mg of dry-leaf material was pulverized using a blade mill and passed through a 0.2 mm sieve to standardize the particle size. Then, 75 mg of the sample (±1 mg to the nearest 0.1 mg) was placed in a 0.22 µm spin filter (Corning Costar Spin-X). Chemical compounds were extracted using 150 µl of a 1:4 solution of ethanol:dichloromethane with 0.075 mg l^{-1} of piperine as an internal standard. Spin filters were centrifuged at 14 *g* for 4 min. Finally, flow-through was transvased to volatile organic compound autosampler vials.

Chromatography (gas chromatography mass spectrometry). We injected 2.5 µl of plant extract into a 4.0 mm ID Low Pressure Drop Precision Inlet Liner with glass wool (Restek). The inlet was kept at a constant temperature of 275 °C. We used split injection with a 60:40 ratio. The oven was programmed as follows: 85 °C, hold for 2 min; ramp 1: 10 °C min^{-1}; 155 °C, hold for 1 min; ramp 2: 6 °C min^{-1}; 260 °C, hold for 1 min; ramp 3: 2 °C min^{-1}; 300 °C, hold for 14 min (total run time 60 min). No column guard was used. Mass spectrometry conditions were as follows: electron ionization source with positive ionization, 70 eV, scanning range 40–550 amu, rate = 1 scan ms^{-1}. To assess carryover and retention time shifts, we injected a sample of solvent containing two internal standards (piperine and limonine) between every Protieae sample. These 'blank' samples were analysed under identical chromatographic conditions to those described above.

Data preprocessing. Gas chromatography mass spectrometry chromatograms were processed using the approach in refs [37,38]. Briefly, we assessed chemical similarity between all sampled species by building a mass spectra library containing all chromatographic features for each species (one library per species). The libraries of each species were then cross-referenced across all species using the automated mass spectral deconvolution and identification system to identify common and unique features based on mass spectra, molecular weight and expected retention time[41–44]. This methodology yielded a species-pair matrix of chemical similarity between all sampled Protieae species.

Using the chemical similarity data from the above methodology, we performed a hierarchical clustering analysis (Ward's algorithm, R package pvclust[45,46]) to construct a dendrogram based on species chemical similarity. Subsequently, we extracted a species-pair matrix of chemical distances from the dendrogram.

Finally, the mass spectra of the different compounds in the samples were compared with National Institute of Standards and Technology/Environmental Protection Agency/National Institutes of Health and MassBank databases[43], as well as with the primary literature. Metabolites that did not have a match from the available mass spectra databases or in the available literature were classified as unknown.

High-molecular-weight metabolite chemistry. Sample extraction**.** Leaf samples from marked trees of each of our 31 study species were further dried under moderate vacuum (100 mTorr) at ambient temperatures for 36 h. Each sample was pulverized using a Mini-BeadBeater (BioSpec Products). For each study species, 100 mg samples of three expansion-phase and three fully expanded leaves were independently extracted with 1.5 ml of 3× hexane, 4× 4:1 (v/v) aqueous ethanol:0.5% acetic acid and 3× 7:3 (v/v) aqueous acetone:0.5% acetic acid. For each sample, aqueous ethanol and acetone extracts were combined in tared vials and dried first under a stream of nitrogen at 37 °C and then under moderate vacuum for 36 h. The mass of each extract was recorded before analysis by high-performance liquid chromatography (HPLC).

Chromatography**.** HPLC analyses were carried out using two systems. The first was a Hitachi LaChrom Elite (Hitachi High-Technologies America) equipped with a photodiode array (PDA) detector (L-2455) configured in tandem with an evaporative light-scattering (ELS) detector (SEDEX 75; SEDERE). The second was an Agilent 1100 system configured for electro-spray ionization mass spectrometry (ESIMS) using an ion trap mass detector (LCQ Fleet; Thermo Fisher Scientific). For each analyte, the PDA provided absorption data, the ELS provided relative abundance data and ESIMS provided molecular weight data. In all cases, separation was performed using an Atlantis T3 2 × 150 mm 3 µ ODS HPLC column (Waters) that was maintained at 40 °C.

HPLC samples were prepared by dissolving the extracts in a standard solution at the rate of 100 µg µl^{-1}. The standard solution was a 1:1 mixture of methanol and methyl sulfoxide that was acidified with 0.1% (v/v) formic acid and contained the following retention time standards (in order of elution; all 200 mM concentration): gallic acid, 4-hydroxy benzoic acid, sinapinic acid, flavone, chrysin and α-tocopherol. Injection volumes were 2 µl (PDA and ELS analyses) or 3 µl (ESIMS analyses).

All HPLC analyses were carried out with a solvent system consisting of a 1:1 mixture of methanol and acetonitrile (A) and water (B), both acidified with 0.1% (v/v) formic acid. A conservative elution programme was employed. It had the following linear gradient steps. At time (t, min) = 0, 5% A in B; at t = 10, 15% A in B; at t = 60, 30% A in B; at t = 75, 70% A in B; at t = 100, 95% A in B; at t = 135, 100% A. This was followed by an isocratic step of 100% A to t = 150. With this gradient elution programme, all phenolic compounds eluted by 80 min and saponins eluted by 110 min. PDA data were collected between 250 and 700 nm. ESIMS data were collected in the negative ion mode from t = 0–80, in the positive ion mode from t = 80–114 and in the negative ion mode from t = 114–150.

Data preprocessing**.** HPLC raw data were processed using MZmine (see complete analytical workflow in Supplementary Materials; ref. [47]). The final peak list was used to construct a plant-by-metabolite matrix. Due to time and financial constraints, we were limited to analysing only 6–9 samples per species, which represents 30% of all 860 plants that were surveyed for herbivores. Consequently, to reduce the potential bias caused by intraspecific variation in subsequent analyses, we transformed the plant metabolite abundance data into four abundance categories to assess the effect that chemical abundance had on herbivore interactions (1 = below the twenty-fifth percentile; 2 = between the twenty-fifth and fiftieth percentile; 3 = between the fiftieth and seventy-fifth percentile; 4 = above the seventy-fifth percentile).

Data analysis. Chemical diversity and Protieae phylogeny**.** To determine whether Protieae secondary compounds show non-random phylogenetic patterns, we assessed the phylogenetic signal of chemical composition for both individual metabolites and overall species. For individual chemical compounds, we assessed Blombergs' K index using its implementation in the Picante R package[48,49]. For overall chemical composition, we used the R package Vegan[50] to calculate the Bray–Curtis chemical similarity distance between each Protiaeae species (vegdist). Next, we used a Mantel test (Ade4 package, 10,000 iterations; ref. [51]) to correlate species' overall chemical similarity and phylogenetic distance.

Herbivore diversity and Protieae chemistry**.** To investigate the relationship between plant chemistry and plant–herbivore interactions for our system, we used a comparative species-level approach by contrasting the secondary chemical composition and herbivore fauna across 31 Protieae species. However, because herbivores can be affected by the host plant's particular chemical makeup as well as the concentration of plant secondary metabolites, we tested the relationship between plant chemistry and herbivore fauna in both quantitative and qualitative ways.

Qualitatively, we assessed the relationship between chemical richness (the total number of chemical compounds detected across all our chemical analysis) and herbivore species richness and abundance via linear regression.

Similarly, for our quantitative analysis we assessed the relationship between species' total chemical investment (the average leaf dry-mass percentage of all secondary metabolites) and herbivore species richness and abundance via linear regression.

To assess whether the polyphagous Protieae herbivores showed strong preferences for specific host species (Fig. 2b), we calculated the expected encounter rate for every plant–herbivore species pair under a Poisson distribution model taking into account plant and herbivore abundances. We then contrasted the expected encounter rate with the observed encounter rate from our field data. If the observed encounter rate of a particular plant–herbivore species pair was different from their expected encounter probability by at least 3 s.d., the plant–herbivore interaction was classified as a strong non-random preference (see Fig. 2).

To determine whether closely related herbivores had preferences for closely related Protieae hosts, we calculated the 'phylogenetic diet similarity' between all herbivore species pairs. By treating each herbivore species diet as an independent community (including abundance data), we could calculate the phylogenetic β diversity between the diets of each herbivore species-pair (Picante package, comdist). We then used a Mantel test (10,000 iterations) to correlate this measure of phylogenetic diet similarity with the herbivore phylogenetic similarity.

To assess whether the chemical composition of the herbivores' diets was correlated with herbivore phylogenetic relationships, we first calculated diet chemical similarity between all herbivore species pairs. Using a chemical distance matrix instead of the Protieae phylogenetic distance matrix as input for the comdist function, we could assess the 'chemical β diversity' between herbivores diets. We then used a Mantel test (10,000 iterations) to compare herbivore chemical diet similarity and herbivore phylogenetic distance.

Additionally, to assess whether chemically or phylogenetically similar Protieae species harboured a similar community of herbivores, we correlated both chemical and phylogenetic Protieae similarity with both simple herbivore community similarity (vegdist, Bray–Curtis algorithm) and herbivore community phylogenetic similarity (comdist).

Effective chemical defence diversity. LASSO regression. To determine how specific secondary metabolites affected the attack rate from specific herbivores on each Protieae species (number of feeding events per herbivore per plant host species over 64 weeks), we used LASSO regression—a statistical learning regularization technique that can handle high-dimensional data via penalized shrinkage-based variable selection[23]. Regularization approaches like LASSO use a penalization term (we used the common L1 penalization norm) that shrinks unassociated predictor variables to zero and therefore performs a simultaneous variable selection. The amount of penalization (λ) is estimated for each dataset via K-fold cross-validation. When the number of explanatory variables (that is, chemicals) is much larger than the response sample size (that is, host-plant species), the shrinkage can be quite strong, resulting in the selection of only those variables with clear predictive roles. For the present study, we used the LASSO implementation from the Glmnet package[52]. Briefly, for each of the top 50 herbivores, we used a LASSO Poisson model and estimated the LASSO λ parameter via K-fold cross-validation. This resulted in a model for each herbivore species that predicts the expected number of visits to each host-plant species as a function of host-plant chemical composition. Given that sample sizes for each plant host were slightly different, we used sample size (log) as a regression offset in the cross-validation and final LASSO models.

Metabolite 'net value'**.** The effectiveness of a particular secondary metabolite as a plant defence is likely to vary from one herbivore to another. Some compounds are likely to serve a defensive function against herbivores; nevertheless, if detoxification evolves in a particular herbivore species, some plant compounds can become cues in the search and recognition of compatible plant hosts[15,53,54]. Furthermore, other compounds can even confer protection to herbivores against third-trophic-level attack; for example, sequestration[55–57]. Finally, given that biological activity can be a very rare natural occurrence[58], most compounds are likely to have no significant effect on any given herbivore species. Consequently, it is difficult to assess the true ecological and evolutionary value of a particular secondary metabolite based on the effect it has on a single herbivore species. Thus, to determine whether the phylogenetic patterns of Protieae secondary metabolites are associated with their effect on herbivores, we calculated the average effect (henceforth: net value) that each secondary metabolite had across all of the top 50 herbivores (which accounted for 87% of all herbivore feeding records). Our LASSO approach yielded a herbivore-species-by-secondary-metabolites matrix

populated with regression coefficients. These coefficients represent the relative strength and sign of the association between the presence of these compounds on a plant host and changes in herbivore attack rates. Thus, to estimate the net value of a particular metabolite in terms of defence against herbivores, we averaged the effects (that is, coefficients) of each metabolite across all of the top 50 herbivores. Finally, to facilitate the interpretation and visualization of the results, we multiplied all net values by −1 so that positive values indicated repellent effects and negative values indicated attractive effects. As a result, in the final analysis, compounds that have a strong positive effect for a plant's fitness (a significant reduction in overall herbivore pressure) are represented by positive net effect values, and vice versa. Compounds with a non-zero net value (HAMs) were used for subsequent downstream analysis.

Phylogenetic patterns of Protieae active secondary metabolites. To determine whether there was a relationship between the defence effectiveness of secondary metabolites and phylogenetic patterns consistent with positive or negative selection, we used four complementary approaches.

(1) We performed a linear regression between metabolite net value and the phylogenetic frequency of each metabolite across all Protieae species. Phylogenetic frequency is defined here as the number of taxa that have a specific trait (in our case, the number of Protieae species where a particular metabolite was detected). This is the most robust approach of the four listed here. Although this approach does not yield any information on specific evolutionary processes, and does not discriminate between single versus multiple independent evolutions of any particular metabolite across Protieae, the strong positive relationship found between the phylogenetic frequency of a metabolite and its net value represents compelling evidence that herbivores play a key role on the evolution of these specific metabolites (Fig. 4a; main text).

(2) To determine whether the net value of the LASSO-selected HAMs was associated with different degrees of positive or negative selection, we tested the fit of the phylogenetic patterns of trait variation for each HAM to two macro-evolutionary continuous-time Markov models of trait evolution. The first model tested was a symmetrical evolutionary model where the probability or rate of evolving from state 0 (HAM not expressed) to state 1 (HAM expressed) is equal to the probability of evolving in the opposite direction (we call this the equal rates model). The second model fitted to each HAM was an asymmetrical model where the two rates (losing or gaining a HAM) are allowed to differ (that is, the all rates differ (ARD) model). If HAMs are evolving in a directional fashion (that is, they are either selected in favour or against), we expect the ARD model to better fit their phylogenetic distribution. To use a conservative approach and remove the potential bias that intraspecific chemical variation could have on the analysis, we performed the analysis using discrete characters (presence or absence of the metabolite). To test the fit of the models to the data, we used the function *fitDiscrete* from the R package Geiger (a maximum-likelihood approach). A total of 57 HAMs fit the ARD model based on the Akaike information criterion values of the models and thus showed some level of directional selection. For all HAMs that fitted the ARD model, we used the same R function to estimate the rates of character gain (q10) and loss (q01) for each metabolite given the data. As a proxy of selection strength and direction, we used the following index: log(q10/q01). In this index, negative values correspond to metabolites with a higher rate of character loss than gain (q10 < q01) meaning that the phylogenetic distribution of said metabolite is consistent with negative selection across the Protieae phylogeny. In contrast, positive values correspond to metabolites with a higher rate of character gain compared with character loss (q10 > q01), indicating some degree of positive selection. Finally, we regressed this index against the HAM's net value to determine the relationship between each HAM's effect on herbivore attack and their estimated strength and direction of selection across the Protieae phylogeny (Supplementary Fig. 7). We found a very strong positive association between the selection strength index (log(q10/q01)) and the metabolite net value, suggesting that HAMs with a strong association with reduced herbivore attacks also showed phylogenetic distribution patterns consistent with strong positive selection across the phylogeny.

(3) To determine whether closely related Protieae species were more similar in HAM chemical composition than distantly related species, we used a Mantel test between species HAM chemical similarity and phylogenetic distance (10,000 iterations).

(4) See Supplementary Materials for an additional maximum parsimony character reconstruction approach that we conducted.

Relationship between HAMs and chemical investment. To assess the relationship between HAM diversity and overall plant chemical investment, we used a linear regression between host-species HAM richness (the total number of HAMs detected in a particular Protieae species) and host-plant species total chemical investment (measured as the average leaf dry-mass percentage of secondary compounds).

In addition, to determine the relationship between metabolite net value and metabolite concentration, we employed a linear regression between each metabolite's net value and its average mass percentile (equivalent to a mass rank of all metabolites). Mass percentiles were used instead of chromatogram signals or intensities to incorporate all compounds into a single analysis. Given that the different chromatographic approaches used here to characterize Protieae species chemistry are intrinsically different (for example, the relationship between signal intensity and metabolite abundance), the use of ranked metabolite concentrations represents a conservative yet robust approach to assessing relative chemical investment across these techniques.

Assessment of the relative importance of secondary metabolite traits to herbivore diversity. To disentangle the relative importance of total chemical diversity, total chemical investment and HAM chemical diversity on the herbivore diversity of Protieae species, we performed a simple general linear model predicting herbivore species richness with total chemical richness, total chemical investment, HAM richness and the interaction between HAM richness and total chemical richness (Poisson distribution, log link functions). All traits showed a significant negative effect on host-species herbivore diversity. The strongest predictor of herbivore diversity was total chemical investment, followed by HAM richness, total chemical richness and finally the interaction between total chemical richness and HAM richness (see Supplementary Table 2).

Research permits. Research permits for this study were provided to P.V.A.F. by the Ministerio del Ambiente and Servicio Nacional de Área Naturales Protegidas por el Estado, Peru.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. All R code used for the final analysis and figures is available on request.

Data availability. The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request. Copies of the results from the COI gene sequencing for the Protieae herbivores are deposited in GenBank (see Supplementary Table 1 for accession numbers of herbivore samples and GenBank record numbers). Datasets generated for this study include a herbivore–plant interaction dataset and a full set of chromatograms for all *Protium* species analysed using the following chromatographic techniques: gas chromatography mass spectrometry, liquid chromatography mass spectrometry, liquid chromatography with a diode array detector and liquid chromatography with an ELS detector. Also available is a collection of photographs of almost all morphospecies of herbivores observed (although please note that the chromatographic data and the photographs of herbivores are large (~12 GB total) and may require special arrangements to be transferred). All 2,300 herbivore voucher collections are deposited in the Essig Museum of Entomology on the campus of the University of California, Berkeley (http://essig.berkeley.edu).

Received: 1 September 2017; Accepted: 3 April 2018;
Published online: 14 May 2018

References

1. Lokvam, J. & Kursar, T. A. Divergence in structure and activity of phenolic defenses in young leaves of two co-occurring *Inga* species. *J. Chem. Ecol.* **31**, 2563–2580 (2005).
2. Ehrlich, P. & Raven, P. H. Butterflies and plants: a study in coevolution. *Evolution* **18**, 586–608 (1964).
3. Speed, M. P., Fenton, A., Jones, M. G., Ruxton, G. D. & Brockhurst, M. A. Coevolution can explain defensive secondary metabolite diversity in plants. *New Phytol.* **208**, 1251–1263 (2015).
4. Berenbaum, M. R. Chemical mediation of coevolution: phylogenetic evidence for Apiaceae and associates. *Ann. Mo. Bot. Gard.* **88**, 45–59 (2001).
5. Edger, P. P. et al. The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl Acad. Sci. USA* **112**, 8362–8366 (2015).
6. Alejo-Armijo, A. et al. Antimicrobial and antibiofilm activities of procyanidins extracted from laurel wood against a selection of foodborne microorganisms. *Int. J. Food Sci. Technol.* **52**, 679–686 (2017).
7. Wise, M. J. & Rausher, M. D. Evolution of resistance to a multiple-herbivore community: genetic correlations, diffuse coevolution, and constraints on the plant's response to selection. *Evolution* **67**, 1767–1779 (2013).
8. Volf, M. et al. Community structure of insect herbivores is driven by conservatism, escalation and divergence of defensive traits in *Ficus*. *Ecol. Lett.* **21**, 83–92 (2018).
9. Lankau, R. A. Specialist and generalist herbivores exert opposing selection on a chemical defense. *New Phytol.* **175**, 176–184 (2007).
10. Volf, M., Hrcek, J., Julkunen-Tiitto, R. & Novotny, V. To each its own: differential response of specialist and generalist herbivores to plant defence in willows. *J. Anim. Ecol.* **84**, 1123–1132 (2015).
11. Janzen, D. H. When is it coevolution? *Evolution* **34**, 611–612 (1980).
12. Iwao, K. & Rausher, M. D. Evolution of plant resistance to multiple herbivores: quantifying diffuse coevolution. *Am. Nat.* **149**, 316–335 (1997).
13. Futuyma, D. J. & Agrawal, A. A. Macroevolution and the biological diversity of plants and herbivores. *Proc. Natl Acad. Sci. USA* **106**, 18054–18061 (2009).

14. Opitz, S. E. W. & Müller, C. Plant chemistry and insect sequestration. *Chemoecology* **19**, 117–154 (2009).
15. Reinecke, A. & Hilker, M. in *Annual Plant Reviews: Insect–Plant Interactions* Vol. 47 (eds Voelckel, C. & Jander, G.) 115–153 (John Wiley & Sons, Chichester, 2014).
16. Johnson, M. T. J., Ives, A. R., Ahern, J. & Salminen, J. P. Macroevolution of plant defenses against herbivores in the evening primroses. *New Phytol.* **203**, 267–279 (2014).
17. Kursar, T. A. et al. The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proc. Natl Acad. Sci. USA* **106**, 18073–18078 (2009).
18. Richards, L. A. et al. Phytochemical diversity drives plant–insect community diversity. *Proc. Natl Acad. Sci. USA* **112**, 10973–10978 (2015).
19. Basset, Y. Diversity and abundance of insect herbivores foraging on seedlings in a rainforest in Guyana. *Ecol. Entomol.* **24**, 245–259 (1999).
20. Dyer, L. A. et al. Host specificity of Lepidoptera in tropical and temperate forests. *Nature* **448**, 696–699 (2007).
21. Novotny, V. et al. Low beta diversity of herbivorous insects in tropical forests. *Nature* **448**, 692–695 (2007).
22. Jones, C. G. & Lawton, J. H. Plant chemistry and insect species richness of British umbellifers. *J. Anim. Ecol.* **60**, 767–777 (1991).
23. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Stat. Methodol.* **58**, 267–288 (1996).
24. Rasmann, S., Bennett, A., Biere, A., Karley, A. & Guerrieri, E. Root symbionts: powerful drivers of plant above- and belowground indirect defenses. *Insect Sci.* **24**, 947–960 (2017).
25. Langenheim, J. H. *Plant Resins: Chemistry, Evolution, Ecology, and Ethnobotany* (Timber Press, Portland, 2003).
26. Andrew, R. L., Peakall, R., Wallis, I. R. & Foley, W. J. Spatial distribution of defense chemicals and markers and the maintenance of chemical variation. *Ecology* **88**, 716–728 (2007).
27. Firn, R. D. & Jones, C. G. in *Phytochemical Diversity and Redundancy in Ecological Interactions* (eds Romeo, J. T., Saunders, J. A. & Barbosa, P.) 295–312 (Springer, New York, 1996).
28. Coley, P. D., Endara, M.-J. & Kursar, T. A. Consequences of interspecific variation in defenses and herbivore host choice for the ecology and evolution of *Inga*, a speciose rainforest tree. *Oecologia* https://doi.org/10.1007/s00442-018-4080-z (2018).
29. Fine, P. V. A., Daly, D. C., Villa, F. G., Mesones, I. & Cameron, K. M. The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae trees in the western Amazon. *Evolution* **59**, 1464–1478 (2005).
30. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit 1 from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
31. Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl Acad. Sci. USA* **101**, 14812–14817 (2004).
32. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
33. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
34. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
35. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)* 1–8 (IEEE, 2010).
36. Miller, M. A., Pfeiffer, W. & Schwartz, T. Extreme digital discovery. In *Proc. 2011 TeraGrid Conference* 1–8 (ACM, 2011).
37. Salazar, D., Jaramillo, A. & Marquis, R. J. The impact of plant chemical diversity on plant–herbivore interactions at the community level. *Oecologia* **181**, 1199–1208 (2016).
38. Salazar, D., Jaramillo, A. & Marquis, R. J. Chemical similarity and local community assembly in the species rich tropical genus *Piper*. *Ecology* **97**, 3176–3183 (2016).
39. Lokvam, J. & Fine, P. V. A. An oxidized squalene derivative from *Protium subserratum* Engl. (Engl.) growing in Peru. *Molecules* **17**, 7451–7457 (2012).
40. Lokvam, J., Metz, M. R., Takeoka, G. R., Nguyen, L. & Fine, P. V. A. Habitat-specific divergence of procyanidins in *Protium subserratum* (Burseraceae). *Chemoecology* **25**, 293–302 (2015).
41. Davies, T. The new automated mass spectrometry deconvolution and identification system (AMDIS). *Spectrosc. Eur.* **10**, 24–27 (1998).
42. *The NIST Mass Spectral Search Program for the NIST/EPA/NIH Mass Spectral Library* (National Institute of Standards and Technology, Gaithersburg, 2017).
43. Horai, H. et al. MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **45**, 703–714 (2010).
44. Stein, S. E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.* **10**, 770–781 (1999).
45. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
46. R Development Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2015).
47. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
48. Kembel, S. W. et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
49. Blomberg, S. P., Garland, T. & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).
50. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
51. Dray, S. & Dufour, A.-B. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).
52. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
53. Meiners, T. Chemical ecology and evolution of plant–insect interactions: a multitrophic perspective. *Curr. Opin. Insect Sci.* **8,** 22–28 2015).
54. Wäschke, N., Meiners, T. & Rostás, M. in *Chemical Ecology of Insect Parasitoids* (eds Wajnberg, E. & Colazza, S.) 37–63 (John Wiley & Sons, Chichester, 2013).
55. Erb, M. & Robert, C. A. M. Sequestration of plant secondary metabolites by insect herbivores: molecular mechanisms and ecological consequences. *Curr. Opin. Insect Sci.* **14**, 8–11 (2016).
56. Heckel, D. G. in *Annual Plant Reviews: Insect–Plant Interactions* Vol. 47 (eds Voelckel, C. & Jander, G.) 77–114 (John Wiley & Sons, Chichester, 2014).
57. Pentzold, S., Zagrobelny, M., Roelsgaard, P. S., Møller, B. L. & Bak, S. The multiple strategies of an insect herbivore to overcome plant cyanogenic glucoside defence. *PLoS ONE* **9**, e91337 (2014).
58. Jones, C. G., Firn, R. D. & Malcolm, S. B. On the evolution of plant secondary chemical diversity [and discussion]. *Phil. Trans. R. Soc. Lond. B* **333**, 273–280 (1991).

Acknowledgements

We thank R. Marquis, M. Metz, N. Whiteman and C. Marshall for comments on the manuscript; G. Takeoka (USDA, Albany), P. Oboyski (Essig Museum), L. Smith (Evolutionary Genetics Laboratory at the University of California, Berkeley) and N. Tsutsui for advice, laboratory space and sample archiving; E. Hendrickson, B. Ho, C. Chong, S. Visvanathan, E. Suh and S. Sharma for help with the laboratory work; and D. Vásquez and C. Villacorta for field assistance. We also thank C. Rivera for help with research permits at the Allpahuayo-Mishana National Reserve. Funding for this project was provided by the NSF DEB (award number 1254214).

Author contributions

D.S., J.L., P.dV., P.V.A.F. and I.M. contributed to the design, analysis and preparation of the manuscript. I.M., M.V.P. and J.M.A.Z. coordinated the fieldwork and data collection. D.S. and P.V.A.F. wrote the article with contributions from all co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0552-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Diego Salazar Amoretti

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Sample size was limited by the nature and scale of the work. We aimed to assess the herbivore species diversity of 31 species of amazonian tropical trees. To assess the later, we needed to visit each species enough across the year to capture not only the diversity of herbivores at each time of year, but also how herbivores changed along the seasonal changes along the year. Thus, each species was sampled once a week for more than a year (64 consecutive weeks). Each week we selected ~30 individuals of each species to be visited and surveyed for herbivores. With 31 species, we samples more than 900 individuals per week. For 64 weeks that is more than 57000 plant surveyed. This yielded more than 4500 insect records.

2. Data exclusions

   Describe any data exclusions.

   Two kinds of data where excluded from the final analysis. First, chemical compounds found only on one plant species. The reason behind this was that these compounds will not be informative to the LASSO regression in order to assess which compounds are candidates to have an effect of herbivore attack rates.
   The second kind of data excluded was herbivores for which we had 19 records or less. We wanted to be sure we had a good idea of the diet breath of the herbivores and how much preferences they had for each plant species.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   Data in this project is correlative.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   Data in this project is correlative, nonetheless, samples in the field and specific transect locations where chosen at random.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Data in this project is correlative.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Data was analyzed via R from publicly available packages. References for these packages are included in the Materials and Methods and Supplement.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	None of our analytical approach used restricted, proprietary or hard-to-find materials.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	n/a

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	n/a
b. Describe the method of cell line authentication used.	n/a
c. Report whether the cell lines were tested for mycoplasma contamination.	n/a
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	n/a

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	n/a

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	n/a