

OPEN

## Phytochemistry reflects different evolutionary history in traditional classes versus specialized structural motifs

Kathryn A. Uckele<sup>1,2,3,8</sup>, Joshua P. Jahner<sup>1,2,8</sup>✉, Eric J. Tepe<sup>4</sup>, Lora A. Richards<sup>1,2,3</sup>, Lee A. Dyer<sup>1,2,3,5</sup>, Kaitlin M. Ochsenrider<sup>6</sup>, Casey S. Philbin<sup>3</sup>, Massuo J. Kato<sup>7</sup>, Lydia F. Yamaguchi<sup>7</sup>, Matthew L. Forister<sup>1,2,3</sup>, Angela M. Smilanich<sup>1,2</sup>, Craig D. Dodson<sup>6</sup>, Christopher S. Jeffrey<sup>1,3,6</sup> & Thomas L. Parchman<sup>1,2</sup>

Foundational hypotheses addressing plant–insect codiversification and plant defense theory typically assume a macroevolutionary pattern whereby closely related plants have similar chemical profiles. However, numerous studies have documented variation in the degree of phytochemical trait lability, raising the possibility that phytochemical evolution is more nuanced than initially assumed. We utilize proton nuclear magnetic resonance (<sup>1</sup>H NMR) data, chemical classification, and double digest restriction-site associated DNA sequencing (ddRADseq) to resolve evolutionary relationships and characterize the evolution of secondary chemistry in the Neotropical plant clade *Radula* (*Piper*; Piperaceae). Sequencing data substantially improved phylogenetic resolution relative to past studies, and spectroscopic characterization revealed the presence of 35 metabolite classes. Metabolite classes displayed phylogenetic signal, whereas the crude <sup>1</sup>H NMR spectra featured little evidence of phylogenetic signal in multivariate tests of chemical resonances. Evolutionary correlations were detected in two pairs of compound classes (flavonoids with chalcones; *p*-alkenyl phenols with kavalactones), where the gain or loss of a class was dependent on the other's state. Overall, the evolution of secondary chemistry in *Radula* is characterized by strong phylogenetic signal of traditional compound classes and weak phylogenetic signal of specialized chemical motifs, consistent with both classic evolutionary hypotheses and recent examinations of phytochemical evolution in young lineages.

Plant secondary chemistry affects plant–herbivore interactions at various stages throughout an insect's lifespan: mixtures of compounds can shape adult oviposition preferences<sup>1</sup>, specific chemical compounds can stimulate larval feeding<sup>2</sup>, specific chemotypes can deter insect herbivores via toxicity or physiological disruptions<sup>3</sup>, and sequestered metabolites can alter immune function against natural enemies<sup>4</sup>. Plants capable of developing novel chemical defenses are hypothesized to accrue higher fitness due to enemy release<sup>5</sup>, potentially resulting in the diversification of plant lineages with conserved chemical phenotypes (the escape and radiate hypothesis<sup>6</sup>). Coevolutionary hypotheses and plant defense theory have yielded clear predictions that herbivory, additional trophic interactions, and resource availability shape the evolution of plant defenses, including secondary metabolites<sup>7,8</sup>. However, an evolutionary response to these biotic and abiotic pressures could be complex and highly context-dependent.

Due in part to the enzymatic complexity of metabolic biosynthesis, phylogenetic conservatism is the null hypothesis for the evolution of plant secondary chemistry<sup>9,10</sup>. Indeed, expectations of phylogenetic conservatism appear to hold at deep evolutionary scales; for example, the family Solanaceae is characterized by the presence

<sup>1</sup>Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, NV 89557, USA. <sup>2</sup>Department of Biology, University of Nevada, Reno, NV 89557, USA. <sup>3</sup>Hitchcock Center for Chemical Ecology, University of Nevada, Reno, NV 89557, USA. <sup>4</sup>Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221, USA. <sup>5</sup>Sección Invertebrados, Museo Ecuatoriano de Ciencias Naturales, Quito, Ecuador. <sup>6</sup>Department of Chemistry, University of Nevada, Reno, NV 89557, USA. <sup>7</sup>Department of Fundamental Chemistry, Institute of Chemistry, University of São Paulo, São Paulo, Brazil. <sup>8</sup>These authors contributed equally: Kathryn A. Uckele and Joshua P. Jahner. ✉email: [jjjahner@gmail.com](mailto:jjjahner@gmail.com)

of tropane alkaloids<sup>11</sup>, though they are consistently present in only 3 of 19 tribes (Datoreae, Hyoscyameae, Mandragoreae) and sporadically found elsewhere<sup>12</sup>. Further, recent work suggests that classes of secondary metabolites are more likely to be phylogenetically conserved in large seed plant clades (e.g., eudicots and superasterids) than at lower taxonomic scales (e.g., orders and families)<sup>13</sup>. At shallower scales, numerous studies provide evidence for evolutionary lability in chemical traits within genera<sup>7,14–16</sup>, suggesting that surveys of phytochemical variation within young plant lineages might yield variable perspectives on the evolution of secondary chemistry. Adding further complexity, many studies have found evidence for strong evolutionary associations among chemical classes<sup>16,17</sup>. For example, Johnson et al.<sup>18</sup> found a strong positive correlation between flavonoids and phenolic diversity and a strong negative correlation between ellagitannins and flavonoids across a phylogeny of 26 evening primroses (*Oenothera*: Onagraceae). Such associations are relevant because they may reflect evolutionary constraints, and their causes may be varied. For example, positive associations may be associated with chemical defense syndromes<sup>9,19</sup> or synergistic effects of multiple classes on herbivore deterrence<sup>20</sup>. Alternatively, negative associations might be consistent with evolutionary tradeoffs or at least different optima in defense space<sup>18,19</sup>. By leveraging advances in organic chemistry and genomics, we stand to increase metabolomic and phylogenetic resolution to provide novel insight into the evolution of phytochemistry.

Recent advances in chemical ecology have improved perspectives on phytochemical diversity across a broad range of taxonomic groups and metabolite classes<sup>21,22</sup>. High throughput processing of plant tissue, rapid advances in spectroscopy, and improved ordination and network analyses have enabled characterization of metabolomic variation across plant communities<sup>10,15,22–24</sup> and stand to enhance our understanding of phytochemical evolution across taxonomic scales<sup>21</sup>. Additionally, structural spectroscopic approaches like <sup>1</sup>H NMR can provide improved resolution of structural variation across a wide range of metabolite classes. Selection on the plant metabolome is inherently multivariate, arising from diverse herbivore communities and environmental conditions<sup>10,25</sup>, and even relatively small structural changes can impart disproportionate shifts in bioactivity. Thus, approaches that capture a larger proportion of the structural variation underlying phytochemical phenotypes could be well suited to addressing hypotheses concerning evolutionary patterns.

Next-generation sequencing data has reinvigorated phylogenetic analyses of traditionally challenging groups characterized by recent or rapid diversification<sup>26</sup>. Reduced representation DNA sequencing approaches [e.g., ddRADseq; genotyping-by-sequencing (GBS)] have been increasingly utilized in phylogenetic studies due to their ability to effectively sample large numbers of orthologous loci throughout the genomes of non-model organisms without the need for prior genomic resources<sup>27</sup>. Nearly all such studies have reported increased topological accuracy and support compared with past phylogenetic inference based on smaller numbers of Sanger-sequenced loci<sup>28,29</sup>, especially when applied to diverse radiations<sup>30,31</sup>. While reduced representation approaches have clear phylogenetic utility at relatively shallow time scales, they have also performed well for moderately deep divergence<sup>29,32</sup>.

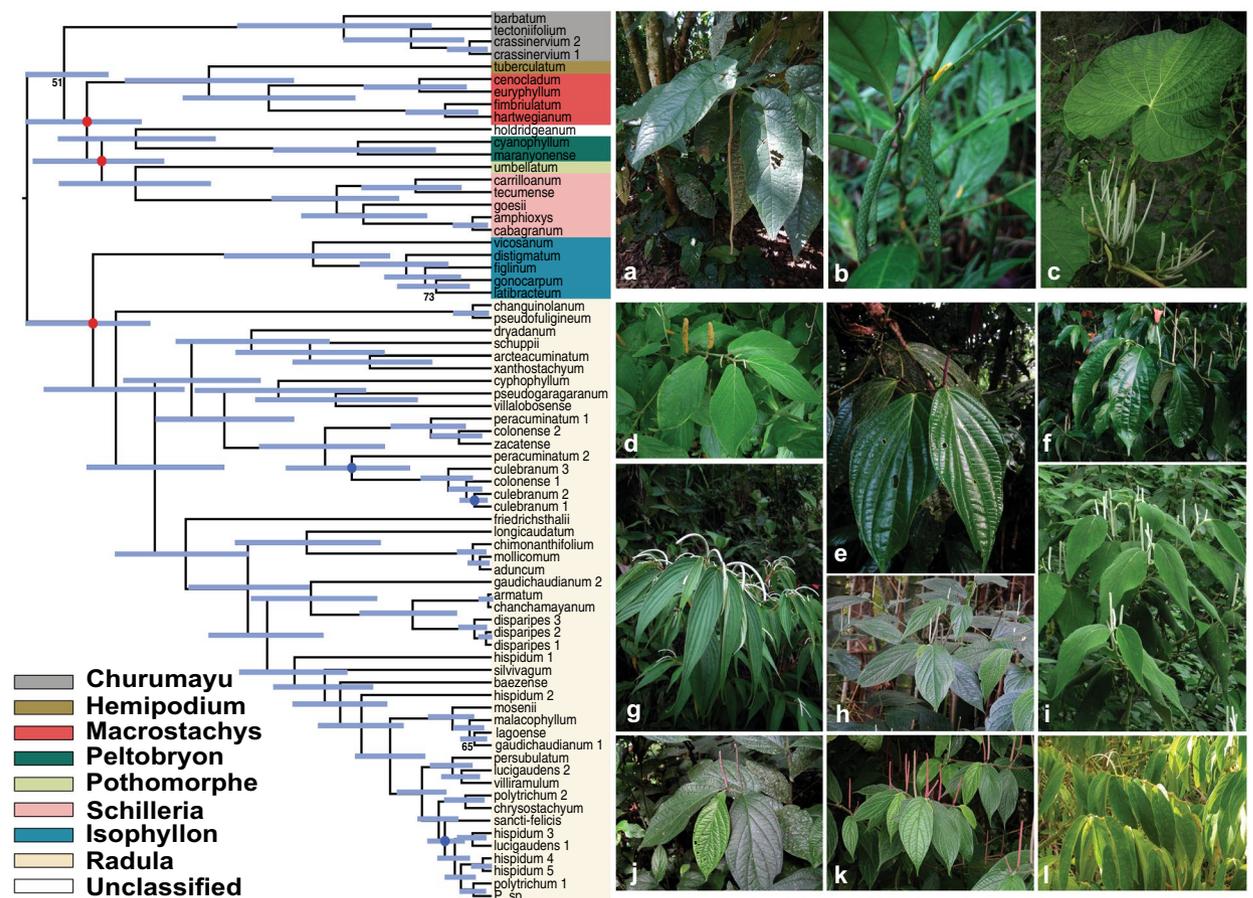
*Piper* (Piperaceae) is a highly diverse, pantropical genus of nearly 2,600 accepted species<sup>33</sup>, with the highest diversity occurring in the Neotropics<sup>34</sup>. Chemically, *Piper* is impressively diverse<sup>35–37</sup>: chemical profiling in a modest number of taxa has yielded 667 different compounds from 11 distinct structural classes thus far<sup>35,36,38,39</sup>. This phytochemical diversity has likely contributed to the diversification of several herbivorous insect lineages that specialize on *Piper*, including the geometrid moth genus *Eois*<sup>40</sup> (Larentiinae). Furthermore, phytochemical diversity in *Piper* communities has been shown to shape tri-trophic interactions and the structure of tropical communities<sup>36,39,41</sup>. As a species-rich genus with abundant and ecologically consequential phytochemical diversity, *Piper* represents a valuable system for understanding how complex diversification histories underlie the evolution of phytochemical diversity.

*Piper* is an old lineage (~72 Ma), yet most of its diversification occurred in the Neotropics during the last 30–40 My following Andean uplift and the emergence of Central America<sup>34,42</sup>. The largest clade of *Piper*, *Radula*, exemplifies this pattern, as much of its extant diversity (~450 species) arose relatively recently during the Miocene<sup>34</sup>. Such bouts of rapid and recent diversification have limited the efficacy of traditional Sanger sequencing methods to resolve the timing and tempo of diversification in *Piper*<sup>42,43</sup>. Past phylogenetic analyses utilizing Sanger-sequenced nuclear and chloroplast regions have consistently inferred eleven major clades within *Piper*; however, phylogenetic resolution within these clades has been elusive<sup>42–45</sup>. Phylogenetic inference based on genome-wide data spanning a range of genealogical histories should facilitate an understanding of evolutionary patterns of phytochemical diversity in *Piper* and their consequences for plant–insect codiversification.

We leveraged complementary phylogenomic, metabolite classification, and <sup>1</sup>H NMR data sets to generate a *Piper* phylogeny and explore the evolution of secondary chemistry within the largest *Piper* clade (*Radula*). We used reduced representation sequencing (ddRADseq) to generate genome-wide data for 71 individuals, spanning eight *Piper* clades but focusing on *Radula*, for phylogenetic analyses. Due to its ability to characterize subtle structural variation across a wide range of compound classes, we used nuclear magnetic resonance (<sup>1</sup>H NMR) spectroscopy to quantify phytochemical diversity in the same individuals. Our goals were to: 1) resolve the evolutionary relationships within the *Radula* clade of *Piper* included in this study; 2) characterize metabolomic variation across the genus and within *Radula* in particular; and 3) quantify the strength of phylogenetic signal and test for evolutionary associations in *Radula* secondary chemistry. Because secondary chemistry is an emergent composite phenotype of many traits that can evolve semi-independently, we expected to detect mixed strengths of phylogenetic signal and strong associations among a subset of traits over evolutionary time.

## Results

**Phylogenetic analyses.** After contaminant filtering and demultiplexing, we retained ~313 million Illumina reads for phylogenetic analyses. Initial clustering, variant calling, and filtering assembled reads into 362,169 ddRADseq loci. There was a high proportion of missing data, presumably due to allelic dropout increasing with

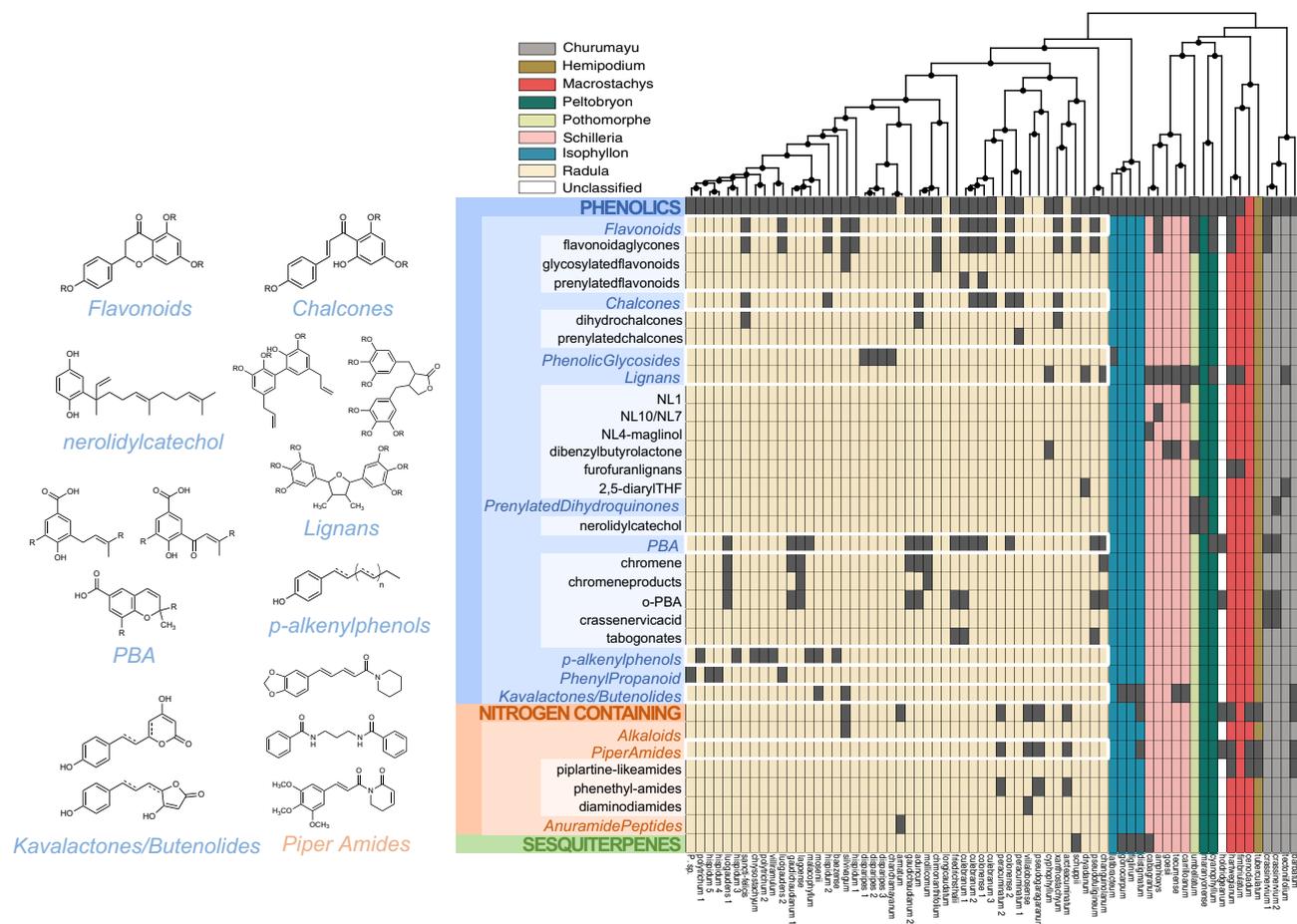


**Figure 1.** Maximum clade credibility tree of 48 samples from the Radula clade of *Piper* and 23 outgroup species inferred with a Bayesian analysis of 641 concatenated ddRADseq loci (55,298 base pairs) comprising 9113 genetic variants (of which 4674 are parsimony informative). The outgroup taxa were sampled across multiple *Piper* clades: Churumayu, Isophyllon, Hemipodium, Macrostachys, Peltobryon, Pothomorphe, and Schilleria. All nodes are supported by at least 95% posterior support except where noted with circles or labels. Blue circles indicate support values between 85 and 95%. Red circles indicate support values between 75 and 85%. Three nodes with less than 75% posterior support were given numerical support values. Blue bars at each node denote the 95% highest posterior density interval on relative node ages. The photos to the right of the tree showcase a sample of *Piper* diversity, including a few of the species which were included in this study: (a) *Piper hillianum* (Macrostachys), (b) *P. acutifolium* (Peltobryon), (c) *P. umbellatum* (Pothomorphe), (d) *P. pseudofulgineum* (Radula), (e) *P. concepcionis* (Radula), (f) *P. disparipes* (Radula), (g) *P. friedrichsthalii* (Radula), (h) *P. dilatatum* (Radula), (i) *P. bredemeyeri* (Radula), (j) *P. immutatum* (Radula), (k) *P. erubescentspicum* (Radula), and (l) the widespread and often weedy *P. aduncum* (Radula). (Photo credits: E. J. Tepe).

high levels of divergence among *Piper* clades. For Bayesian phylogenetic inference, we mitigated the influence of missing data by removing loci absent in > 30% of samples. The final dataset for phylogenetic analysis consisted of 641 ddRADseq loci (~ 86 bp in length each) that housed 9113 genetic variants (51% parsimony informative). Aligned loci were concatenated into a nexus alignment with missing data at 18.9% of sites.

Bayesian phylogenetic analysis of ddRADseq data resolved eight major Neotropical *Piper* clades with high posterior support (Fig. 1). While past phylogenetic studies supported the monophyly of seven of these eight clades (Macrostachys, Radula, Peltobryon, Pothomorphe, Hemipodium, Isophyllon, and Schilleria)<sup>34,43</sup>, our analysis resolved an additional clade, Churumayu. Notably, Isophyllon and Churumayu were highly supported, monophyletic clades and not nested within Radula, as was inferred in previous analyses<sup>43</sup>. Contrary to previous phylogenetic hypotheses of *Piper*<sup>34,43</sup>, our analysis might suggest Churumayu is the most basal clade, but we caution that this node had very low posterior support (51%). Intrageneric relationships below the clade level were highly resolved, with nearly all nodes exhibiting greater than 95% posterior support, including within the diverse Radula clade (Fig. 1). Our phylogenetic hypothesis for Radula indicates three species (*P. hispidum*, *P. colonense*, *P. lucigaudens*) may be paraphyletic.

**Phytochemical diversity in *Piper*.** All but four individuals included in the inferred *Piper* tree were successfully chemically extracted and profiled. Nearly all common compound classes that have been previously reported in *Piper*<sup>46</sup> were observed from our compound characterization analysis (see Table S2). This analysis



**Figure 2.** Patterns of chemical variation are displayed for individuals in this study. Taxa comprise the columns of the matrix and are ordered according to their inferred phylogenetic relationships. Groups of columns are colored according to their designated *Piper* clade. Black circles within the phylogenetic tree designate nodes with posterior support values greater than 85%. Each row of the matrix represents a metabolite class that was detected from  $^1\text{H}$  NMR and MS-based methods, and dark grey cells indicate the presence of that class in that taxa. Classes are hierarchically nested; capitalized font signifies the three classes at the highest level (and coarsest resolution), italicized font signifies the intermediate level, and black font signifies the lowest level (and highest resolution). Rows outlined in white indicate traits that were analyzed for phylogenetic signal in *Radula*. To the left of the matrix are representative compounds for a subset of metabolite classes that were detected in our samples.

revealed the presence of broad metabolite classes that are ubiquitous across plant families (e.g., lignans, flavonoids/chalcones, etc.) as well as classes that are specifically common in *Piper* (e.g., amides) (Fig. 2, Table S2). Specific compound characterization revealed genus specific compounds and compound classes (piplartine, cenocladamide, crassinervic acid, kava lactones), as well as metabolites that are more rarely reported in plants (putrescine diamides, nerolidyl catechol, alkenyl phenols, anuramide peptides) (Fig. 2, Table S2). Alternative methods, such as sampling across a species' ontogeny, sampling reproductive parts or roots, and storing freshly collected tissue in methanol rather than air drying would add to a more comprehensive picture of variation in phytochemical diversity across and within species, but our sampling was standardized to allow for initial comparisons across species, some of which were collected in remote regions.

**Metabolite phylogenetic signal and evolutionary associations.** We recovered 35 metabolite classes, of which only eight were sufficiently present across our taxa to afford tests of phylogenetic signal and correlated evolution. For all eight metabolite classes, estimates of  $D$  did not deviate from a null distribution generated under a scenario of Brownian motion (Table 1), consistent with phylogenetic signal. Two of the eight traits, phenolic glycosides and lignans, exhibited strong phylogenetic signal ( $D < 0$ ), while the remaining six traits exhibited weak phylogenetic signal ( $0 < D < 1$ ). Further, all but one of the metabolite classes had observed values of  $D$  that differed from a null distribution generated under a phylogenetic randomness scenario (Table 1). The mean of the observed  $D$  estimates for the metabolite classes was 0.06, with the largest  $D$  statistic observed for the chalcone class ( $D = 0.62$ ) and the smallest observed for the phenolic glycosides ( $D = -1.18$ ) (Table 1).

Of the 28 pairwise tests of correlated evolution, only two were significant based on a significance level of 0.05. Evidence for correlated evolution was detected in two pairs of metabolite classes: (1) flavonoids and chalcones;

Metabolite class	Observed D	$\Sigma d_{obs}$	Randomness ( $H_0$ : D = 1)		Brownian ( $H_0$ : D = 0)	
			Mean( $\Sigma d_r$ )	P	Mean( $\Sigma d_b$ )	P
Flavonoids	0.49	14.18	17.56	0.012	11.01	0.093
Chalcones	0.62	11.59	13.33	0.095	8.79	0.088
Phenolic glycosides	-1.18	3.11	7.01	0.000	5.19	0.950
Lignans	-0.02	4.16	5.47	0.036	4.19	0.564
PBA	0.22	12.40	17.51	0.001	10.96	0.293
<i>p</i> -alkenyl phenols	0.33	9.47	12.30	0.010	8.19	0.265
Kavalactones/butenolides	0.02	5.17	6.99	0.027	5.18	0.504
Piper amides	0.1	5.37	7.00	0.033	5.18	0.482

**Table 1.** Estimates of phylogenetic signal (D)<sup>57</sup> for a subset of metabolite classes (see “Methods” for explanation of subset). To ask whether traits evolved under scenarios of Brownian motion (D = 0) or phylogenetic randomness (D = 1), observed values of D were compared to null distributions of D modeled under each scenario.

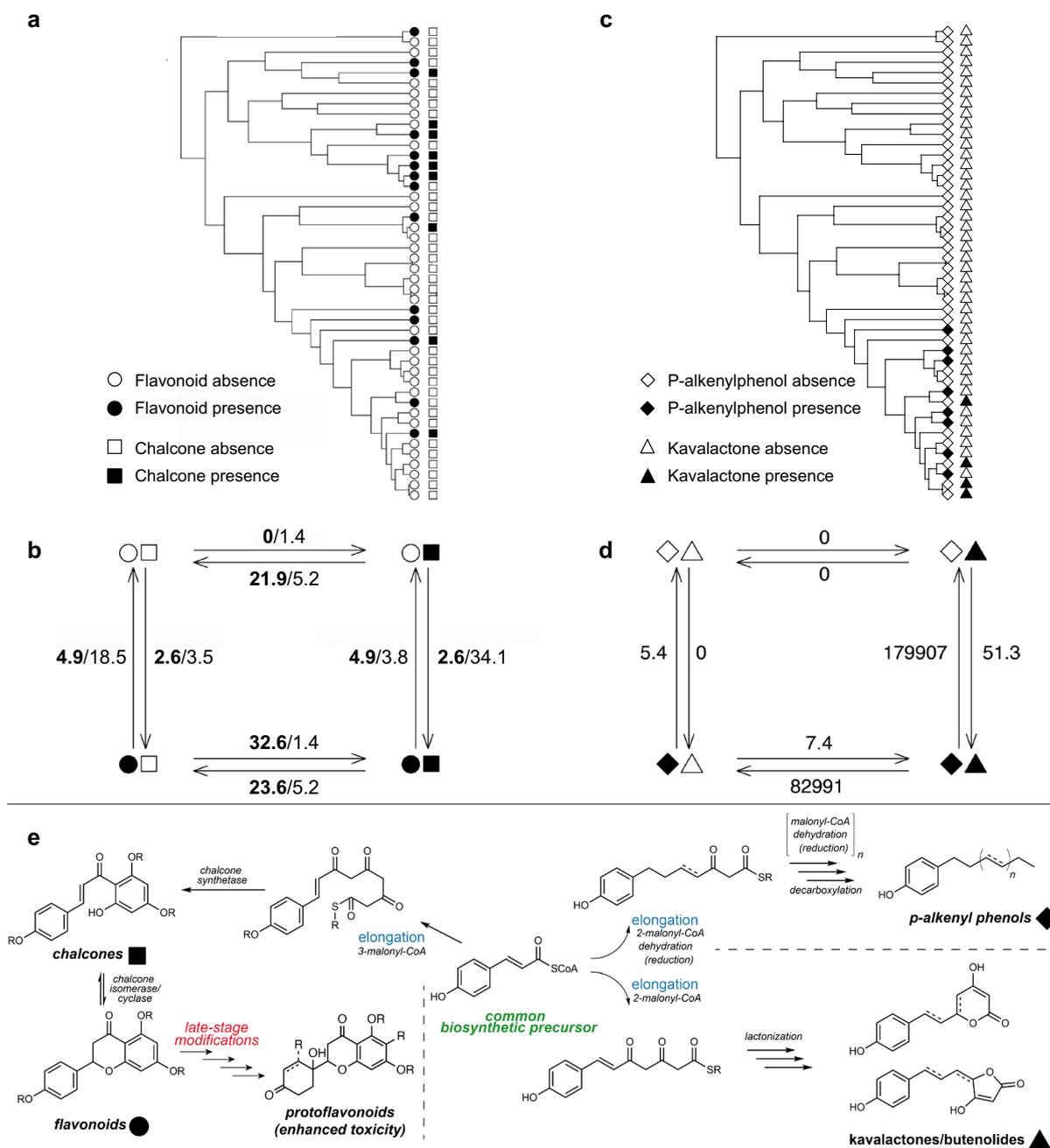
Comparison	Model	AIC	$\Delta$ AIC	AIC weight
Chalcones, flavonoids	Chalcones contingent on flavonoids	95.58	0	0.51
	Flavonoids contingent on chalcones	96.02	0.44	0.41
	Interdependent evolution	99.84	4.26	0.06
	Independent evolution	102.46	6.88	0.02
Kavalactones/butenolides, <i>p</i> -alkenyl phenols	Interdependent evolution	62.35	0	0.95
	Kavalactones/butenolides contingent on <i>p</i> -alkenyl phenols	69.65	7.29	0.03
	<i>p</i> -alkenyl phenols contingent on kavalactones/butenolides	70.61	8.26	0.02
	Independent evolution	71.57	9.22	0.01

**Table 2.** Correlated evolution was detected in two pairs of metabolite classes with Pagel’s method<sup>76</sup>: (1) chalcones and flavonoids; and (2) kavalactones/butenolides and *p*-alkenyl phenols. A model comparison framework was employed to evaluate four potential models of trait evolution using AIC: interdependent evolution (transition rate in one trait depends on state at another, and vice versa); contingent change (transition rate in one trait depends on state at another, but not the converse); and independent evolution.

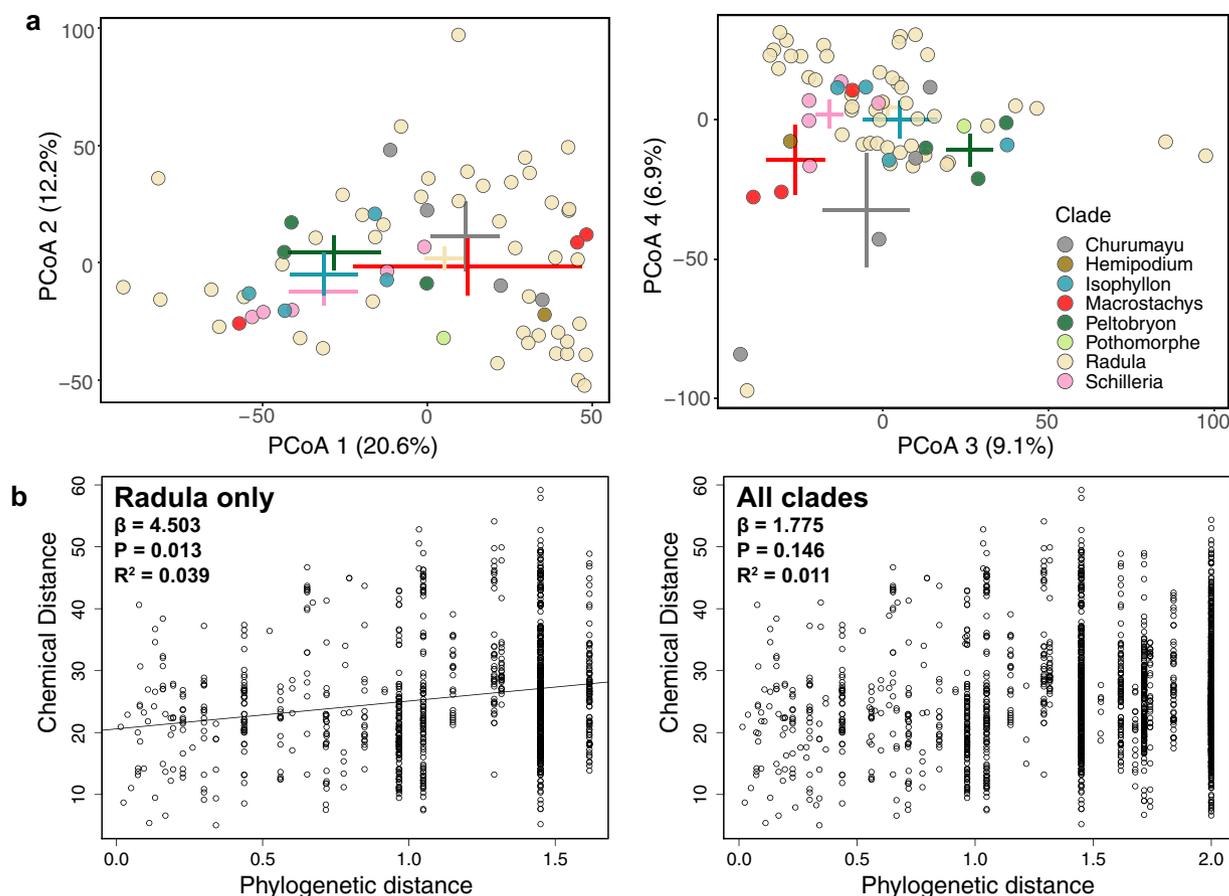
and (2) *p*-alkenyl phenols and kavalactones/butenolides. For the first pair of traits, a model of contingency in which changes in chalcones depend on the state of flavonoids provided the best fit to the data (Table 2). In this model, when flavonoids are present, chalcone gains are 1.4 times more probable than chalcone losses; however, when flavonoids are absent, chalcone losses are much more probable than chalcone gains (Fig. 3). The alternative contingency model for this pair of traits (i.e., changes in flavonoids depend on the state of chalcones) was also a good fit to the data (Table 2). According to this model, when chalcones are present, flavonoid gains are approximately nine times more probable than flavonoid losses. Alternatively, when chalcones are absent, flavonoid losses are approximately five times more probable than flavonoid gains (Fig. 3). For the second pair of traits, *p*-alkenyl phenols and kavalactones/butenolides, the best fit model was one of interdependent evolution in which changes in *p*-alkenyl phenol depend on the state of kavalactones/butenolides, and vice versa (Table 2). When kavalactones/butenolides are present, *p*-alkenyl phenol transitions are more probable than when they are absent, with the loss of *p*-alkenyl phenols being much more probable than the gain of *p*-alkenyl phenols under both scenarios. Alternatively, when *p*-alkenyl phenols are present, the loss of kavalactones/butenolides is extremely probable relative to the gain of kavalactones/butenolides, which is rarely observed. When *p*-alkenyl phenols are absent, kavalactones/butenolides are rarely gained or lost (Fig. 3).

**Phylogenetic signal in high-dimensional metabolomic data.** While the eight metabolite classes uniformly exhibited at least moderate levels of phylogenetic signal, evidence for phylogenetic signal in multivariate analyses of the crude <sup>1</sup>H NMR data was largely absent. PCo axes 1 & 2 and 3 & 4 explained 32.8% and 16.0% of variance in the <sup>1</sup>H NMR data, respectively, but showed little clustering by clade (Fig. 4a). Permutational multivariate analyses of variance were not significant for combinations of either PCo 1 & 2 ( $P=0.407$ ) nor 3 & 4 ( $P=0.142$ ), suggesting that different clades do not form distinct clusters in chemospace based on their <sup>1</sup>H NMR spectra.

According to the MRM models, phylogenetic distance significantly predicts phytochemical distance within *Radula* ( $\beta=4.503$ ,  $P=0.013$ ) but not across all clades ( $\beta=1.775$ ,  $P=0.146$ ) (Fig. 4b). It is important to note that the



**Figure 3.** Evolutionary associations were detected in two pairs of traits according to Pagel's test<sup>76</sup> of correlated evolution: (1) flavonoids and chalcones and (2) *p*-alkenyl phenols and kavalactones/butenolides. Filled shapes indicate presences and unfilled shapes indicate absences of flavonoids (circles), chalcones (squares), *p*-alkenyl phenols (diamonds), and kavalactones/butenolides (triangles), respectively. The shapes used in the phylogenetic plots (a,c) are repeated below (b,d) to depict four states comprising all combinations of presences and absences in the pair of traits. Arrows represent transition rates between states. (b) As both models of contingent change provided good fits to the flavonoid and chalcone data, both sets of transition rates are displayed, with the first set of values (bolded) corresponding to the best supported model (chalcone evolution contingent on flavonoid state) and the second set of values corresponding to the alternative contingency model (flavonoid evolution contingent on chalcone state). (d) The best fit model to the *p*-alkenyl phenol and kavalactone/butenolide data was one of interdependent evolution, where *p*-alkenyl phenol evolution is dependent on the state at the kavalactone/butenolide trait, and vice versa. Panel (e) illustrates the enzymatic processes and branch points along biosynthetic pathways that give rise to the four classes of metabolites. Chalcones are immediate biosynthetic precursors of flavonoids, where the inherent reactivity of the chalcone moiety permits cyclization to the flavonoid scaffold. Subtle structural changes to the flavonoid scaffold caused by late-stage oxidation can produce protoflavonoids, a rare class of metabolite with potent cytotoxic activity. In contrast, the pathways of *p*-alkenyl phenols and kavalactones diverge much earlier and embark on distinct chain elongation pathways that lead to long-chain lipophilic substituent characteristic of the *p*-alkenyl phenols in one case, and lactones (kavalactones and butenolides) in the other case.



**Figure 4.** (a) Chemospace of all 67 *Piper* samples constructed with the crude  $^1\text{H}$  NMR data across 263 peaks. Point colors were chosen according to clade designation as portrayed in the phylogenetic tree in Fig. 1. A legend is provided in the rightmost PCoA plot. Error bars depict estimates of ordination mean  $\pm$  3 standard errors for each clade. (b) MRM analyses recovered a significant positive relationship between phylogenetic and chemical distances calculated among samples from the *Radula* clade (left), but did not recover a significant relationship when calculated among all samples from all clades (right).

proportion of variance explained by the significant MRM model is low ( $R^2 = 0.039$ ), suggesting that the majority of variation in NMR data cannot be explained by phylogenetic distance.

Analyses with the generalized  $K$  statistic ( $K_{\text{mult}}$ ) indicated lower levels of phylogenetic signal in the metabolomic data than expected under a Brownian motion model of evolution for *Piper* generally ( $K_{\text{mult}} = 0.1606$ ,  $P = 0.001$ ) and for *Radula* specifically ( $K_{\text{mult}} = 0.1803$ ,  $P = 0.001$ ). Still, the observed  $K_{\text{mult}}$  was higher than all  $K_{\text{mult}}$  values obtained with permutations of the  $^1\text{H}$  NMR dataset (Fig. S1). Additionally, few  $K_{\text{mult}}$  tests of the permuted data yielded significant  $P$ -values (4.4% of permutations), indicating that the estimate we observed, though subtle and lower than Brownian motion expectations, was real and not a statistical artifact of zero-inflation in the data.

## Discussion

*Piper* is a hyper-diverse lineage in which phytochemical diversity has influenced evolutionary and ecological processes and shaped complex tropical communities<sup>15,39</sup>. However, limitations in both the degree of phylogenetic resolution and the understanding of phytochemical diversity in this group have precluded analyses of phylogenetic signal and correlated evolution of phytochemistry. Phylogenies inferred here with ddRADseq data substantially improved resolution and support compared to past studies of *Piper*, which were limited by interspecific variation in small numbers of Sanger-sequenced loci<sup>34,42,43</sup>. Although the data set did not include members from all previously recognized groups, analyses resolved eight monophyletic Neotropical *Piper* clades, seven of which have been inferred in previous analyses of the genus based on chloroplast *psbJ-petA* and *nrITS*<sup>34,43</sup>. Two of the eight clades, *Churumayu* and *Isophyllon*, had been previously nested within *Radula*<sup>43</sup>, however, our results suggest that they are independent monophyletic lineages (Fig. 1). Despite low support for several deep divergences, the phylogeny inferred here had strong resolution and support for recent relationships, including within *Radula* (Fig. 1), consistent with other recent reduced representation sequencing studies that have generated high quality phylogenies at shallow time scales<sup>28,31,32</sup>. However, a potential limitation of such sequencing designs may include the recovery of fewer loci shared by more distantly related samples due to allelic dropout<sup>47</sup>. It is possible that allelic dropout, potentially exacerbated by strict filtering based on missing data, led to weak support values for deep splits in the phylogeny, many of which occurred early in the history of the Neotropical

*Piper* lineage<sup>34</sup>. Nonetheless, the resulting subset of data (641 loci; 9113 SNPs) was sufficient for inferring a largely resolved phylogeny, highlighting the potential promise of reduced representation sequencing for resolving evolutionary histories even in groups spanning moderately deep divergence. Although our sampling was limited to 44 of 450 estimated species within *Radula*, the extent of sampling is a substantial improvement over past phylogenetic analyses for the group<sup>42,43</sup>.

Comparative studies have taken diverse approaches to analyzing metabolomic data, each providing a unique perspective on the evolution of specialized metabolites<sup>10,24</sup>. Here, we first characterized the presence/absence of 35 metabolite classes commonly used to categorize plant secondary compounds that are hierarchically nested into three levels of structural resolution. Specific categories at the lowest level of the hierarchy, representing specialized structural motifs or specific molecules, were rare across species and precluded tests of phylogenetic signal and correlated evolution at our level of taxonomic sampling (Fig. 2). Despite not being able to test for phylogenetic signal, clustering is evident for more specific categories, such as crassinervic acid and prenylated flavonoids, which are only present in small subclades but include particularly effective defenses<sup>36,46</sup>. Alternatively, broader metabolite classes at intermediate and high positions in the hierarchy that are directly tied to fundamental secondary metabolite biosynthetic pathways were more abundant across species and exhibited moderately high levels of phylogenetic signal across *Radula* (Table 1, Fig. 2). This pattern may be expected if initial biosynthetic steps are conserved over longer evolutionary scales, permitting the abundance of broad chemical classes, yet later stage modifications of these core structures are more evolutionarily labile, causing structural similarity to be low even among related species. Flavonoids are a good example of this pattern, with pathways that form the flavonoid scaffold being very conserved, as they are catalyzed by modified enzymes from ubiquitous metabolic pathways, but then subsequent biosynthetic steps (e.g., those catalyzed by p450 enzymes) modify these scaffolds<sup>48</sup>, yielding unique molecules towards the tips of evolutionary trees (Fig. 3E). For example, late-stage modification of common flavonoid scaffolds can result in the production of non-aromatic protoflavonoids. These compounds rarely occur across the plant kingdom and have only recently been found in one species of *Piper*<sup>49</sup>, but this type of subtle structural modification that leaves most of the flavonoid scaffold intact dramatically enhances the cytotoxic properties compared to that of the parent flavonoid<sup>50,51</sup>.

One key prediction from the escape and radiate hypothesis is that adaptive defensive traits should be phylogenetically conserved within the lineage they evolved, but this prediction has mostly been evaluated with broad classes of secondary metabolites at high taxonomic scales<sup>6,13,48</sup> rather than specific compounds in recent diversifications<sup>7,10,16</sup>. A growing number of studies conducted at shallow evolutionary scales suggests low phylogenetic signal in many chemical traits<sup>14,15,18</sup>. While evidence for low phylogenetic signal is often attributed to high evolutionary rates (i.e., evolutionary lability), simulations under various evolutionary processes and conditions indicate that the relationship between phylogenetic signal and rate of trait evolution is not necessarily straightforward, and evidence for low phylogenetic signal is not an indication of any single evolutionary process<sup>52</sup>. Nonetheless, understanding how phylogenetic signal responds to variation in phylogenetic scale is informative in a comparative sense, especially among different traits or classes of traits generated with different levels of analytical resolution. Phylogenetic signal is also a useful starting point for developing insights into the drivers of herbivorous insect radiations, as codiversification in many of these lineages is structured in part by chemical defense and biotic interactions<sup>40,53</sup>. Our results are generally consistent with the predictions of moderately strong signal for broad classes of compounds, as well as the lack of signal for specific structures captured by <sup>1</sup>H NMR data.

The <sup>1</sup>H NMR data address a different set of hypotheses than data from categorization of individual molecules—peaks represent resonances associated with particular molecular structures rather than individual compounds, and the chemical shift (frequency), shape, and abundance of these resonances are extremely sensitive to subtle structural changes. <sup>1</sup>H NMR spectroscopy easily detects a great range and subtle differences in compositional and structural complexity, including increasing size, asymmetry and oxidation states, that might be predicted to evolve in response to divergent selection across plant populations responding to different suites of enemies<sup>22</sup>. Low levels of phylogenetic signal in the <sup>1</sup>H NMR data is also likely due to the fact that many molecular features of small defensive molecules have potentially evolved in a convergent manner across *Piper*, such as the kavalactones, *p*-alkenyl phenols, pipartine, oxidized prenylated benzoic acids, chromenes, anuramide peptides, and phenethyl amides.

There are numerous limitations that could affect estimates of phylogenetic signal in comparative studies<sup>54</sup> that are relevant to the analyses presented here. First, incomplete taxon sampling likely influenced our results to some degree, but sampling was conducted randomly, and the probability that a particular species was sampled was unlikely related to any aspect of its chemical phenotype<sup>55</sup>. Low sampling proportion in clades other than *Radula* may have reduced our power to detect phylogenetic signal across all our sampled clades<sup>55</sup> (Fig. 4b). However, despite only sampling approximately 10% of the *Radula* clade of *Piper*, our sample size should provide sufficient power to infer phylogenetic signal in this clade if present<sup>56,57</sup> (Fig. 4b). Second, while topological errors and small sample size may have reduced our power to detect phylogenetic signal at deeper time scales<sup>58</sup>, more comprehensive genomic sampling produced enhanced phylogenetic resolution of the *Radula* clade, where we focused the majority of phylogenetic comparative methods. In addition, we were unable to quantify the measurement error associated with the chemical traits within species, which can decrease the statistical power for detecting phylogenetic signal<sup>56,59,60</sup>. It is also possible that environmental effects on our chemical traits could bias estimates of phylogenetic signal and correlations<sup>59</sup>.

The causes of correlated evolution, including linkage, epistasis, and selection, are difficult to detect without careful approaches in quantitative genetics and population genomics. Nevertheless, one advantage of examining the presence/absence of multiple classes of defensive compounds in a phylogenetic context is that it is possible to test for expected patterns of correlated evolution due to shared metabolic pathways (e.g., flavonoids and cardenolides<sup>7</sup>) or due to adaptive advantages of specific mixtures. Recent studies detecting evolutionary

associations among chemical traits<sup>17,18</sup> have posited that the branching structure of metabolic pathways could potentially drive this pattern. If metabolite classes share a common precursor, one might expect evolutionary tradeoffs and negative covariation. Alternatively, if metabolite classes lie along the same metabolic pathway, an increase in one class may be concomitant with increases in another (or vice versa), causing positive covariation among the classes. There are also numerous empirical examples supporting the hypotheses that correlations may be driven by functional redundancy<sup>61</sup> or selection for synergistic effects on herbivores<sup>20</sup> rather than the structural constraints of metabolism. Suites of covarying defensive traits, or defense syndromes, have been detected in several plant genera<sup>9,53</sup> and plant communities<sup>62</sup>, and have been predominantly used to describe covariation among mechanical and chemical defenses. It is interesting to note the correlated evolution of the flavones/chalcones and the *p*-alkenyl phenols/kavalactones could be due to metabolic constraints, as well as possible adaptations via synergistic (e.g., kavalactones in *P. methysticum*) or other mixture-associated defensive attributes<sup>22</sup>. Flavonoids and chalcones are directly linked biosynthetically, such that the inherent reactivity of the chalcone moiety permits the enzymatic processes that result in cyclization to the flavonoid scaffold (Fig. 3e). This strong biosynthetic tie yields a clear prediction that the presence of one would depend on the other, and indeed our structural analysis found many cases where both metabolite classes co-occurred in the same sample. Revealing the relationship between the kavalactones and *p*-alkenyl phenols is more tenuous because both classes are less prevalent across our samples. Kavalactones and *p*-alkenyl phenols are dramatically different compounds that diverge at a much earlier branch point from a common cinnamic/coumaric acid precursor. Whereas one polyacetate chain extension pathway leads to the long-chain lipophilic substituent, characteristic of the *p*-alkenyl phenols, the other chain extension pathway conserves oxidation states through the chain extension process to produce the lactones (kavalactones or butenolides) through cyclization reactions (Fig. 3e). The overall outcome is different than the chalcone-flavonoid relationship; in this case, two dramatically different compounds are produced by divergence from a common early-stage biosynthetic precursor in contrast to the immediate biosynthetic precursor relationship between chalcones and flavonoids. Broader sampling across *Piper* and *Radula* will be necessary to confirm this unexpected relationship between kavalactones and *p*-alkenyl phenols.

## Conclusion

Here we sought to advance understanding of phylogenetic relationships within *Piper* while simultaneously investigating the mode and manner of phytochemical evolution in this group. In addition to generating a well-resolved phylogeny, our results support theoretical expectations that broad classes of compounds display higher degrees of phylogenetic signal than molecular features revealed by <sup>1</sup>H NMR data. In addition, trait associations observed in *Radula* can be used to pose functional hypotheses about genetic constraints or biases on phytochemical evolution and how these factors structure plant-animal interactions. Such investigations are one of the emerging frontiers in terrestrial ecology, and we hope that our study provides one example of how collaborative and multi-disciplinary research can progress in this area.

## Methods

**Study system and sample collection.** For phylogenetic and chemical analyses, we collected leaf material from 71 individuals representing 65 Neotropical *Piper* species from the following clades: Churumayu ( $N=3$ ), Hemipodium ( $N=1$ ), Isophyllon ( $N=5$ ), Macrostachys ( $N=4$ ), Peltobryon ( $N=2$ ), Pothomorphe ( $N=1$ ), *Radula* ( $N=44$ ), and *Schilleria* ( $N=5$ ). This study complied with all local and national regulations/guidelines, and vouchers for all collections were deposited in herbaria in the country of origin as stipulated in the permit documents (Table S1). Brazilian collections were made under permit No. 15780-6 from the Sistema de Autorização e Informação em Biodiversidade (SISBIO). Costa Rican collections were made under the permits R-054-2018-OT-CONAGEBIO and R-055-2018-OT-CONAGEBIO from the Ministerio del Ambiente y Energía (MINAE). Collections from Ecuador were conducted under the permit 03-IC-FAU/FLO-DNP/MA granted by the Ministerio del Ambiente. Collections from Panamá were covered by the permit SE/AP-15-13 from the Autoridad Nacional del Ambiente (ANAM). Finally, Peruvian collections were covered by the permit 288-2015-SERFOR-DGGSPFFS granted by the Servicio Nacional Forestal de Fauna Silvestre (SERFOR). All collections were identified by E.J.T. in the field, and confirmed with vouchers in the herbarium using regional keys, where available, comparison with type specimens, and experience with the genus. For chemical profiling and DNA sequencing, we collected the youngest, fully expanded leaves and dried them immediately with silica gel. While drying on silica gel may not inhibit enzymatic activity and could limit our analyses to relatively stable molecules, this is not an issue for the phylogenetic analyses described below. Collections were only made from mature individuals in the field. Vouchers were pressed, dried, and deposited in one or more herbaria for future reference and species verification (Table S1). To investigate the evolution of phytochemistry at a relatively shallow evolutionary scale, we conducted the majority of our sampling within *Radula*<sup>34</sup>.

**Phylogenetic analyses.** Genome-wide polymorphism data was generated for 71 individuals for phylogenetic analyses. Either the same accession sampled for chemical analysis, or an individual from the same population as the one sampled, were sequenced with a genotyping-by-sequencing approach<sup>63</sup> that is analogous to ddRADseq<sup>64</sup>. Briefly, genomic DNA was digested with two restriction enzymes, *EcoRI* and *MseI*. Sample-specific barcoded oligos containing Illumina adaptors were annealed to the *EcoRI* cut sites, and oligos containing the alternative Illumina adaptor were annealed to the *MseI* cut sites. Fragments were PCR amplified and pooled for sequencing. The library was size-selected for fragments between 350 and 450 base pairs (bp) with the Pip-pin Prep System (Sage Sciences, Beverly, MA), and sequenced on two lanes of an Illumina HiSeq 2500 at the University of Texas Genome Sequencing and Analysis Facility (Austin, TX). Single-end, 100 bp, raw sequence data were filtered for contaminants (*E. coli*, *PhiX*, Illumina adaptors or primers) and low quality reads using

bowtie2\_db<sup>65</sup> and a pipeline of bash and perl scripts (<https://github.com/ncgr/tapioca>). We used custom perl scripts to demultiplex our reads by individual and trim barcodes and restriction site-associated bases.

Assembly and initial filtering was conducted with ipyRAD v.0.7.30<sup>66</sup>. ipyRAD was specifically designed to assemble ddRADseq data for phylogenetic applications, permits customization of clustering and filtering, and allows for indel variation among samples<sup>66</sup>. Because a suitable *Piper* genome was not available at the time of analysis, we generated a de novo consensus reference of sampled genomic regions with ipyRAD. Briefly, nucleotide sites with phred quality scores lower than 33 were treated as missing data. Sequences were clustered within individuals according to an 85% similarity threshold with vsearch<sup>67</sup> and aligned with muscle<sup>68</sup> to produce stacks of highly similar ddRADseq reads (hereafter, ddRADseq loci). The sequencing error rate and heterozygosity were jointly estimated for all ddRADseq loci with a depth > 6, and these parameters informed statistical base calls according to a binomial model. Consensus sequences for each individual in the assembly were clustered once more, this time across individuals, and discarded if possessing > 8 indels (max\_Indels\_locus), > 50% heterozygous sites (max\_shared\_Hs\_locus), or > 20% variable sites (max\_SNPs\_locus). To reduce the amount of missing data in our alignment matrix, ddRADseq loci were retained if they were present in at least 50 of 71 samples. The nexus file of concatenated consensus sequences for each individual, including invariant sites, was used as input for the Bayesian phylogenetic methods described below. Individual FASTQ files, nexus alignment, and complete information on additional parameter settings for this analysis are archived at Dryad (<https://doi.org/10.5061/dryad.j6q573nc7>).

To resolve patterns of diversification and to provide a foundation for investigating variation in patterns of phytochemical evolution, we estimated a rooted, calibrated tree according to a relaxed clock model in RevBayes v.1.0.12<sup>69</sup>, which provides the ability to specify custom phylogenetic models for improved flexibility compared with other Bayesian approaches. The prior distribution on node ages was defined by a birth–death process in which the hyper priors on speciation and extinction rates were exponentially distributed with  $\lambda = 10$ . We relaxed the assumption of a global molecular clock by allowing each branch-rate variable to be drawn from a lognormal distribution. After comparing the relative fits of JC, HKY, GTR, and GTR + Gamma nucleotide substitution models with Bayes factors, we modeled DNA sequence evolution according to the best-fit HKY model. Eight independent MCMC chains were run for 100,000 generations with a burn-in of 1,000 generations and sampled every 10 generations. Chains were visually assessed for convergence with Tracer v.1.7.1<sup>70</sup> and numerically assessed with effective sample sizes (ESS), the Gelman – Rubin convergence diagnostic<sup>71</sup>, and by comparing the posterior probabilities of clades sampled between MCMC chains. The maximum clade credibility (MCC) tree provided the ultrametric fixed tree topology and relative node ages for phylogenetic comparative methods described below.

**Chemical profiling.** Crude proton nuclear magnetic resonance (<sup>1</sup>H NMR) spectroscopy was chosen for chemotype mapping due to its ability to characterize subtle structural variation across a wide range of compound classes in a single, reproducible, non-destructive analysis<sup>39</sup>. Briefly, after leaf samples were ground to fine powder, approximately 100.0–2000.0 mg of leaf material were ground and transferred to a glass screw cap test tube with 10 ml of methanol, sonicated for 10 min, and filtered. This step was repeated and both filtrates were combined in a pre-weighed 20 ml scintillation vial. The solvent was removed *in vacuo* and dissolved in 0.6 ml methanol-*d*<sub>4</sub> for <sup>1</sup>H NMR analysis. Crude <sup>1</sup>H NMR solutions were standardized to 13.1 ± 3.8 mg/mL when possible and analyzed on a Varian 400 MHz solution state NMR spectrometer with autosampler. Data were processed using MestReNova software (Mestrelab Research, Santiago de Compostela, Spain). Spectra from the crude extracts were aligned with the solvent peak (CD<sub>3</sub>,  $\delta = 3.31$  ppm), baseline corrected, and phase corrected. Solvent and water peaks were removed and the binned spectra were normalized to a total area of 100. This data set is referred to as “crude <sup>1</sup>H NMR”.

In addition to crude <sup>1</sup>H NMR spectral chemotyping, we further annotated samples based upon the presence or absence of compound classes. To further gain structural resolution across the crude extracts that were sampled, aliquots of the <sup>1</sup>H NMR extracts were diluted and subjected to GC–MS and LC–MS analysis (see Supplementary Information for additional details). Crude extracts were classified using chemotaxonomic classifications outlined in Parmar’s comprehensive review of *Piper* phytochemistry<sup>35</sup>, and our rationale for assigning chemical classes is outlined for each species in Table S2. Briefly, phenolic compounds were identified from high-resolution matches to the METLIN mass spectrometry database<sup>72</sup>. Database hits were then confirmed by agreement of crude <sup>1</sup>H NMR chemical shifts with literature values for phenolics known to be found in *Piper*, but not always *Radula* species. Many compounds identified by LC–MS as flavonoids and chalcones had multiple possible METLIN matches, which confounded NMR confirmation. In these cases, we were still able to differentiate flavonoids from chalcones by characteristic UV spectra ( $I_{\max} \sim 350$  nm). Phenylpropanoids and *p*-alkenyl phenols were identified based on characteristic GC–MS fragmentation for these compound classes known to be found in *Piper*. *Piper* amides were characterized in a similar fashion, starting from high-resolution mass spectrometric matches and confirming with known <sup>1</sup>H NMR data from the literature. In some cases, crude 2D-NMR analysis (COSY, HSQC) was used to confirm structural classifications. CORrelated Spectroscopy (COSY) was used to identify <sup>1</sup>H NMR that were contained within the same molecule, while Heteronuclear Single Quantum Coherence (HSQC) spectroscopy was used to identify the carbon (<sup>13</sup>C) resonances associated with certain proton (<sup>1</sup>H) signals to verify the presence of specific functional groups<sup>73</sup>. Only the most abundant and spectroscopically apparent compounds were classified due to the low sensitivity of NMR. 35 total classes were identified at three levels of structural resolution. At the coarsest level of resolution, we identified compounds as phenolics, nitrogen-containing, or sesquiterpenes. Within the phenolics, we identified nine intermediate and 17 high-resolution subclasses. Within the nitrogen-containing compounds, we identified three intermediate and three high-resolution subclasses. Finer resolution

was not characterized for the sesquiterpene class. This hierarchical set of 35 traits is referred to as “metabolite classes” (Fig. 2). Additional details on chemical profiling can be found in the Supplementary Information.

**Phylogenetic signal and evolution of metabolite classes.** To assess whether metabolite classes were phylogenetically conserved across *Radula*, we quantified phylogenetic signal in these binary traits using the D statistic<sup>57</sup>. The D statistic calculates the sum of sister-clade differences,  $\Sigma d_{obs}$  for an observed tree and binary trait, and scales this value with the distributions of sums expected under two disparate evolutionary models, random and Brownian motion ( $\Sigma d_r$  and  $\Sigma d_b$ , respectively), using the following equation:

$$D = \frac{[\Sigma d_{obs} - \text{mean}(\Sigma d_b)]}{[\text{mean}(\Sigma d_r) - \text{mean}(\Sigma d_b)]}$$

Thus, D is expected to equal 1 when the observed binary trait is distributed randomly, lacking phylogenetic signal, and is expected to equal 0 when it exhibits phylogenetic signal as expected under Brownian motion. As tests of phylogenetic signal with the D statistic are most accurate when the ratio of presences and absences is closer to 1:1<sup>57</sup>, we tested for phylogenetic signal in eight of the 35 metabolite classes (outlined in white in Fig. 2) which were present in a sufficient proportion of taxa. We used the *phylo.d* function in the *caper* package<sup>74</sup> in R v.4.0.0<sup>75</sup> to calculate the observed D for a subset of binary traits that were sufficiently present across the phylogeny. This value was compared to a distribution of D values simulated under models of phylogenetic randomness ( $D = 1$ ) and pure Brownian motion ( $D = 0$ ) to determine whether the observed D differed from either zero or one.

To detect evolutionary associations among pairs of metabolite classes within *Radula*, we used Pagel’s method<sup>76</sup> that models evolutionary changes in two binary traits, X and Y, as continuous-time Markov processes in which the probabilities of state transition at one trait may depend on the state at the other trait. We tested all pairwise associations among the eight metabolite classes that were represented by a sufficient number of *Radula* taxa to provide accurate tests of evolutionary associations ( $N = 28$ ). Significant tests of correlated evolution were followed by tests of contingency, in which changes at X depend on the state of Y, or vice versa. Model fits, comparisons, and plots were performed with the *fitPagel* function in the *phytools* package<sup>77</sup> in R.

**Multivariate analyses of phylogenetic signal with crude <sup>1</sup>H NMR spectra.** While the analyses above based on broad classifications of structurally determined metabolites provide a coarse view of phytochemical evolution, these classifications are anchored to the foundations of plant secondary metabolite biosynthesis. Using <sup>1</sup>H NMR spectra as a raw chemotype should allow a more detailed multivariate perspective on phytochemical diversity. Studies on other plant taxa have typically detected some signal and evolutionary correlations for broad classes of compounds but not necessarily for specific compounds or biologically active moieties, both of which can be inferred from <sup>1</sup>H NMR data. Multivariate approaches to phylogenetic comparative methods have provided insight into covarying suites of related traits, while simultaneously increasing the statistical power to detect phylogenetic signal<sup>78</sup> and differences in trait means among taxa<sup>79</sup>. Indeed, these multivariate approaches might be particularly useful when exploring the evolution of complex phenotypes, like the plant metabolome, which exhibit trait covariances due to metabolomic or functional associations<sup>20</sup>. Here we utilize three multivariate methods to detect patterns of phylogenetic signal for 263 resonances found in the crude <sup>1</sup>H NMR data representing all 35 metabolite classes: (1) principal coordinate analyses (PCoA); (2) multiple regression on distance matrices (MRM); and (3) multivariate estimation of phylogenetic signal.

To visualize patterns of chemotypic variation across all sampled species from all clades, we first analyzed the <sup>1</sup>H NMR data with PCoA. First we calculated the Manhattan distances between all pairwise species with the *dist* function in R, and then conducted PCoA on the distance matrix using the *pcoa* function in R. If the major axes of metabolomic variation are phylogenetically conserved, the plotted species scores should be clustered by clade in a rotated principal coordinate (PCo) space. Alternatively, if metabolomic variation is randomly distributed across the phylogeny, there should be little to no clustering by clade<sup>80</sup>. The degree to which plant clade predicted chemical similarity was assessed using permutational multivariate analysis of variance (permanova)<sup>81</sup> in the *vegan* package<sup>82</sup> in R based on Euclidean distances of the first four PCo axes.

Mantel tests have been frequently used to assess the degree of phylogenetic signal in multivariate data<sup>10,83,84</sup> by estimating the relationship between phylogenetic and phenotypic distances. Simulations under scenarios of measurement error have found instances where Mantel tests outperform traditional univariate methods in detecting phylogenetic signal, especially as the number of traits increases<sup>60</sup>. Because we were unable to account for measurement error in our study, we utilized Multiple Regression on distance Matrices (MRM)<sup>85</sup> to examine the relationship between metabolomic and phylogenetic distance at two evolutionary scales (within *Radula* and across all clades). Euclidean distances were calculated from the crude <sup>1</sup>H NMR spectra using the *dist* function in R, and phylogenetic distances for *Radula* only and all clades were calculated using the *cophenetic.phylo* function in the *ape* package<sup>86</sup> in R. MRM analyses were implemented using the *MRM* function with 1000 permutations in the *ecodist* package<sup>87</sup> in R.

Since Blomberg’s  $K^{26}$  statistic exhibits higher statistical power to detect phylogenetic signal relative to Mantel tests<sup>88</sup>, we quantified phylogenetic signal of the crude <sup>1</sup>H NMR at both evolutionary scales using a multivariate generalization of the K statistic ( $K_{mult}$ )<sup>89</sup> with the *physignal* function in the *geomorph* package<sup>90</sup> in R. Similar to the aforementioned D statistic, the K statistic compares the observed variation to that expected under Brownian motion, but the K statistic does not scale this comparison by the variation exhibited under a completely random evolutionary model<sup>56,89</sup>. Values of K greater than 1 indicate phylogenetic signal greater than expected under Brownian motion, whereas values between 0 and 1 indicate less signal than expected under Brownian motion. Significance for the generalized K statistic was assessed by permuting the <sup>1</sup>H NMR peak data among the tips of the phylogeny for 999 iterations. To determine whether the zero-inflated nature of the <sup>1</sup>H NMR data influenced the

detection of phylogenetic signal, we permuted our  $^1\text{H}$  NMR data set over 1000 iterations by randomly indexing our original  $^1\text{H}$  NMR data matrix. This permutation method preserves the original proportion of zeros in the matrix while obfuscating any observed phylogenetic signal. The generalized  $K$  statistic test was calculated for each permutation, and our observed generalized  $K$  statistic was compared to the null distribution of permuted values.

Received: 19 February 2021; Accepted: 15 July 2021

Published online: 26 August 2021

## References

1. Thompson, J. N. & Pellmyr, O. Evolution of oviposition behavior and host preference in Lepidoptera. *Annu. Rev. Entomol.* **36**, 65–89 (1991).
2. Bowers, M. D. Iridoid glycosides and host-plant specificity in larvae of the buckeye butterfly, *Junonia coenia* (Nymphalidae). *J. Chem. Ecol.* **10**, 1567–1577 (1984).
3. Zagrobelny, M. *et al.* Cyanogenic glucosides and plant–insect interactions. *Phytochemistry* **65**, 293–306 (2004).
4. Richards, L. A. *et al.* Synergistic effects of iridoid glycosides on the survival, development and immune response of a specialist caterpillar, *Junonia coenia* (Nymphalidae). *J. Chem. Ecol.* **38**, 1276–1284 (2012).
5. Berenbaum, M. Toxicity of a furanocoumarin to armyworms: A case of biosynthetic escape from insect herbivores. *Science* **201**, 532–534 (1978).
6. Ehrlich, P. R. & Raven, P. H. Butterflies and plants: A study in coevolution. *Evolution* **18**, 586–608 (1964).
7. Agrawal, A. A., Salminen, J. P. & Fishbein, M. Phylogenetic trends in phenolic metabolism of milkweeds (*Asclepias*): Evidence for escalation. *Evolution* **63**, 663–673 (2009).
8. Maron, J. L., Agrawal, A. A. & Schemske, D. W. Plant-herbivore coevolution and plant speciation. *Ecology* **100**, e02704 (2019).
9. Agrawal, A. A. & Fishbein, M. Plant defense syndromes. *Ecology* **87**, S132–S149 (2006).
10. Salazar, D. *et al.* Origin and maintenance of chemical diversity in a species-rich tropical tree lineage. *Nat. Ecol. Evol.* **2**, 983 (2018).
11. Griffin, W. J. & Lin, G. D. Chemotaxonomy and geographical distribution of tropane alkaloids. *Phytochemistry* **53**, 623–637 (2000).
12. Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3–19 (2003).
13. Zhang, Y. *et al.* Phylogenetic patterns suggest frequent multiple origins of secondary metabolites across the seed plant “tree of life”. *Natl. Sci. Rev.* **7**, 964–977 (2020).
14. 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).
15. Salazar, D., Jaramillo, M. A. & Marquis, R. J. Chemical similarity and local community assembly in the species rich tropical genus *Piper*. *Ecology* **97**, 3176–3183 (2016).
16. Allevato, D. M., Groppo, M., Kiyota, E., Mazzafera, P. & Nixon, K. C. Evolution of phytochemical diversity in *Pilocarpus* (Rutaceae). *Phytochemistry* **163**, 132–146 (2019).
17. Boachon, B. *et al.* Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Mol. Plant* **1**, 1084–1096 (2018).
18. Johnson, M. T., Ives, A. R., Ahern, J. & Salminen, J. P. Macroevolution of plant defenses against herbivores in the evening primroses. *New Phytol.* **203**, 267–279 (2014).
19. Agrawal, A. A. Macroevolution of plant defense strategies. *Trends Ecol. Evol.* **22**, 103–109 (2007).
20. Richards, L. A., Dyer, L. A., Smilanich, A. M. & Dodson, C. D. Synergistic effects of amides from two *Piper* species on generalist and specialist herbivores. *J. Chem. Ecol.* **36**, 1105–1113 (2010).
21. Sedio, B. E. Recent breakthroughs in metabolomics promise to reveal the cryptic chemical traits that mediate plant community composition, character evolution and lineage diversification. *New Phytol.* **214**, 952–958 (2017).
22. Dyer, L. A. *et al.* Modern approaches to study plant–insect interactions in chemical ecology. *Nat. Rev. Chem.* **2**, 50–64 (2018).
23. Richards, L. A. *et al.* Phytochemical diversity and synergistic effects on herbivores. *Phytochem. Rev.* **15**, 1153–1166 (2016).
24. Sedio, B. E., Parker, J. D., McMahon, S. M. & Wright, S. J. Comparative foliar metabolomics of a tropical and a temperate forest community. *Ecology* **99**, 2647–2653 (2018).
25. Fine, P. V. A. *et al.* The growth–defense trade-off and habitat specialization by plants in Amazonian forests. *Ecology* **87**, S150–S162 (2006).
26. Léveillé-Bourret, É., Chen, B. H., Garon-Labrecque, M. É., Ford, B. A. & Starr, J. R. RAD sequencing resolves the phylogeny, taxonomy and biogeography of Trichophoreae despite a recent rapid radiation (Cyperaceae). *Mol. Phylogenet. Evol.* **145**, 106727 (2020).
27. Parchman, T. L., Jahner, J. P., Uckele, K. A., Galland, L. M. & Eckert, A. J. RADseq approaches and applications for forest tree genetics. *Tree Genet. Genomes* **14**, 39 (2018).
28. Massatti, R., Reznicek, A. A. & Knowles, L. L. Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in *Carex* sect. *Racemosae*. *Am. J. Bot.* **103**, 337–347 (2016).
29. Du, Z. Y., Harris, A. J. & Xiang, Q. Y. J. Phylogenomics, co-evolution of ecological niche and morphology, and historical biogeography of buckeyes, horsechestnuts, and their relatives (Hippocastaneae, Sapindaceae) and the value of RAD-seq for deep evolutionary inferences back to the Late Cretaceous. *Mol. Phylogenet. Evol.* **145**, 106726 (2020).
30. Fernández-Mazuecos, M. *et al.* Resolving recent plant radiations: power and robustness of genotyping-by-sequencing. *Syst. Biol.* **67**, 250–268 (2017).
31. Paetzold, C., Wood, K. R., Eaton, D., Wagner, W. L. & Appelhans, M. S. Phylogeny of Hawaiian *Melicope* (Rutaceae): RAD-Seq resolves species relationships and reveals ancient introgression. *Front. Plant Sci.* **10**, 1074 (2019).
32. Eaton, D. A., Spriggs, E. L., Park, B. & Donoghue, M. J. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Syst. Biol.* **66**, 399–412 (2017).
33. Callejas-Posada, R. Piperaceae. in *Flora Mesoamericana* Vol. 2, pt. 2 (eds. Davidse, G., Ulloa Ulloa, C., Hernández, H. M. & Knapp, S.) 1–618 (Missouri Botanical Garden Press, 2020).
34. Martínez, C., Carvalho, M. R., Madriñán, S. & Jaramillo, C. A. A late Cretaceous *Piper* (Piperaceae) from Colombia and diversification patterns for the genus. *Am. J. Bot.* **102**, 273–289 (2015).
35. Parmar, V. S. *et al.* Phytochemistry of the genus *Piper*. *Phytochemistry* **46**, 597–673 (1997).
36. Dyer, L. A. & Palmer, A. D. N. *Piper: A Model Genus for Studies of Phytochemistry, Ecology, and Evolution*. (Kluwer Academic/Plenum Publishers, 2004).
37. Richards, L. A. *et al.* Phytochemical diversity drives plant–insect community diversity. *Proc. Natl. Acad. Sci. USA* **112**, 10973–10978 (2015).
38. Kato, M. J. & Furlan, M. Chemistry and evolution of the Piperaceae. *Pure Appl. Chem.* **79**, 529–538 (2007).
39. Richards, L. A., Oliveira, C. & Dyer, L. A. Shedding light on chemically mediated tri-trophic interactions: A  $^1\text{H}$ -NMR network approach to identify compound structural features and associated biological activity. *Front. Plant Sci.* **9**, 1155 (2018).

40. Jahner, J. P. *et al.* Host conservatism, geography, and elevation in the evolution of a Neotropical moth radiation. *Evolution* **71**, 2885–2900 (2017).
41. Glassmire, A. E. *et al.* Intraspecific phytochemical variation shapes community and population structure for specialist caterpillars. *New Phytol.* **212**, 208–219 (2016).
42. Smith, J. F., Stevens, A. C., Tepe, E. J. & Davidson, C. Placing the origin of two species-rich genera in the late cretaceous with later species divergence in the tertiary: a phylogenetic, biogeographic and molecular dating analysis of *Piper* and *Peperomia* (Piperaceae). *Plant Syst. Evol.* **275**, 9 (2008).
43. Jaramillo, M. A. *et al.* A phylogeny of the tropical genus *Piper* using ITS and the chloroplast intron psbJ–petA. *Syst. Bot.* **33**, 647–660 (2008).
44. Molina-Henao, Y. F., Guerrero-Chacón, A. L. & Jaramillo, M. A. Ecological and geographic dimensions of diversification in *Piper* subgenus *Ottonia*: A lineage of Neotropical rainforest shrubs. *Syst. Bot.* **41**, 253–262 (2016).
45. Asmarayani, R. Phylogenetic relationships in Malesian-Pacific *Piper* (Piperaceae) and their implications for systematics. *Taxon* **67**, 693–724 (2018).
46. Salehi, B. *et al.* *Piper* species: A comprehensive review on their phytochemistry, biological activities and applications. *Molecules* **24**, 1364 (2019).
47. Cariou, M., Duret, L. & Charlat, S. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecol. Evol.* **3**, 846–852 (2013).
48. Yonekura-Sakakibara, K., Higashi, Y. & Nakabayashi, R. The origin and evolution of plant flavonoid metabolism. *Front. Plant Sci.* **10**, 943 (2019).
49. Freitas, G. C. *et al.* Cytotoxic non-aromatic B-ring flavanones from *Piper carnioconnectivum* C. DC. *Phytochemistry* **97**, 81–87 (2014).
50. Hunyadi, A., Martins, A., Danko, B., Chang, F. R. & Wu, Y. C. Protoflavones: A class of unusual flavonoids as promising novel anticancer agents. *Phytochem. Rev.* **13**, 69–77 (2014).
51. Latif, A. D. *et al.* Protoflavone-chalcone hybrids exhibit enhanced antitumor action through modulating redox balance, depolarizing the mitochondrial membrane, and inhibiting ATR-dependent signaling. *Antioxidants* **9**, 1–18 (2020).
52. Revell, L. J., Harmon, L. J. & Collar, D. C. Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* **57**, 591–601 (2008).
53. Endara, M. J. *et al.* Coevolutionary arms race versus host defense chase in a tropical herbivore-plant system. *Proc. Natl. Acad. Sci. USA* **114**, E7499–E7505 (2017).
54. Kamilar, J. M. & Cooper, N. Phylogenetic signal in primate behaviour, ecology and life history. *Philos. Trans. R. Soc. B* **368**, 20120341 (2013).
55. Garamszegi, L. Z. & Møller, A. P. Nonrandom variation in within-species sample size and missing data in phylogenetic comparative studies. *Syst. Biol.* **60**, 876–880 (2011).
56. 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).
57. Fritz, S. A. & Purvis, A. Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* **24**, 1042–1051 (2010).
58. Sakamoto, M. & Venditti, C. Phylogenetic non-independence in rates of trait evolution. *Biol. Lett.* **14**, 20180502 (2018).
59. Ives, A. R., Midford, P. E. & Garland, T. Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.* **56**, 252–270 (2007).
60. Hardy, O. J. & Pavoine, S. Assessing phylogenetic signal with measurement error: A comparison of Mantel tests, Blomberg *et al.*'s K, and phylogenetic histograms. *Evolution* **66**, 2614–2621 (2012).
61. Romeo, J. T., Saunders, J. A. & Barbosa, P. *Phytochemical Diversity and Redundancy in Ecological Interactions*, Vol. 30. (Springer, 2013).
62. Kursar, T. A. & Coley, P. D. Convergence in defense syndromes of young leaves in tropical rainforests. *Biochem. Syst. Ecol.* **31**, 929–949 (2003).
63. Parchman, T. L. *et al.* Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* **21**, 2991–3005 (2012).
64. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135 (2012).
65. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
66. Eaton, D. A. PyRAD: Assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844–1849 (2014).
67. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
68. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
69. Höhna, S. *et al.* RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).
70. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
71. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7**, 457–472 (1992).
72. Guijas, C. *et al.* METLIN: a technology platform for identifying known and unknowns. *Anal. Chem.* **90**, 3156–3164 (2018).
73. Crews, P., Rodríguez, J. & Jaspars, M. *Organic Structure Analysis* (Oxford University Press, 2010).
74. Orme, D. *et al.* caper: Comparative analyses of phylogenetics and evolution in R. R package version 1.0.1. <https://CRAN.R-project.org/package=caper> (2018)
75. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, <https://www.R-project.org/>, 2020).
76. Pagel, M. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proc. R. Soc. B* **255**, 37–45 (1994).
77. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
78. Zheng, L. *et al.* New multivariate tests for phylogenetic signal and trait correlations applied to ecophysiological phenotypes of nine *Manglietia* species. *Funct. Ecol.* **23**, 1059–1069 (2009).
79. Clavel, J., Escarguel, G. & Merceron, G. mvmorph: An R package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* **6**, 1311–1319 (2015).
80. Klingenberg, C. P. & Gidaszewski, N. A. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Syst. Biol.* **59**, 245–261 (2010).
81. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
82. Oksanen, J. *et al.* vegan: Community Ecology Package, R package version 2.5-7. <https://CRAN.R-project.org/package=vegan> (2020)
83. Cardini, A. & Elton, S. Does the skull carry a phylogenetic signal? Evolution and modularity in the guenons. *Biol. J. Linn. Soc.* **93**, 813–834 (2008).
84. Easson, C. G. & Thacker, R. W. Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. *Front. Microbiol.* **5**, 532 (2014).
85. Lichstein, J. W. Multiple regression on distance matrices: A multivariate spatial analysis tool. *Plant Ecol.* **188**, 117–131 (2007).

86. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
87. Goslee, S. C. & Urban, D. L. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* **22**, 1–19 (2007).
88. Harmon, L. J. & Glor, R. E. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution* **64**, 2173–2178 (2010).
89. Adams, D. C. A generalized *K* statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Syst. Biol.* **63**, 685–697 (2014).
90. Adams, D. C. & Otárola-Castillo, E. geomorph: An R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* **4**, 393–399 (2013).

## Acknowledgements

This research was funded by the National Science Foundation (DEB-1145609, DEB-1442103, DEB-1442075, and DEB-1146119) to C.S.J., L.A.D., L.A.R., M.L.F., T.L.P., A.M.S., and E.J.T. by the National Science Foundation Graduate Research Award (Award No. 1650114) to K.A.U., and by FAPESP (Award No 2014/50316-7) to M.J.K. Fellowship support for K.A.U., K.M.O., and C.S.P. and funding for chemical instrumentation and analysis was provided by the Hitchcock Center for Chemical Ecology at the University of Nevada, Reno. We thank Jennifer L. McCracken for her assistance with the collection of GC-MS data for the categorical chemical characterization, and we thank Chris Feldman, Beth Leger, and Steve Vander Wall for their guidance during the earliest stages of this project.

## Author contributions

M.L.F., L.A.D., A.M.S., C.S.J., L.A.R., T.L.P., and E.J.T. developed the original idea for the research and secured funding. E.J.T., M.J.K., and L.F.Y. collected specimens. E.J.T. extracted DNA from plant specimens. K.A.U. and T.L.P. generated genotyping-by-sequencing libraries. K.A.U. and J.P.J. analyzed the genetic data. K.M.O. and L.A.R. performed chemical extractions and analyses. C.S.J., C.S.P., and C.D.D. executed chemical annotation and structure determination. K.A.U. and J.P.J. wrote the first draft of the manuscript, and all authors contributed to subsequent revisions.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-96431-3>.

**Correspondence** and requests for materials should be addressed to J.P.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021