nature ecology & evolution

Article

https://doi.org/10.1038/s41559-024-02444-2

# Tree phytochemical diversity and herbivory are higher in the tropics

Received: 16 May 2023

Accepted: 20 May 2024

Published online: 27 June 2024

Check for updates

Lu Sun [1], Yunyun He [1,2], Min Cao [1], Xuezhao Wang [1,2], Xiang Zhou [3], Jie Yang [1] ✉ & Nathan G. Swenson [4]

A long-standing but poorly tested hypothesis in plant ecology and evolution is that biotic interactions play a more important role in producing and maintaining species diversity in the tropics than in the temperate zone. A core prediction of this hypothesis is that tropical plants deploy a higher diversity of phytochemicals within and across communities because they experience more herbivore pressure than temperate plants. However, simultaneous comparisons of phytochemical diversity and herbivore pressure in plant communities from the tropical to the temperate zone are lacking. Here we provide clear support for this prediction by examining phytochemical diversity and herbivory in 60 tree communities ranging from species-rich tropical rainforests to species-poor subalpine forests. Using a community metabolomics approach, we show that phytochemical diversity is higher within and among tropical tree communities than within and among subtropical and subalpine communities, and that herbivore pressure and specialization are highest in the tropics. Furthermore, we show that the phytochemical similarity of trees has little phylogenetic signal, indicating rapid divergence between closely related species. In sum, we provide several lines of evidence from entire tree communities showing that biotic interactions probably play an increasingly important role in generating and maintaining tree diversity in the lower latitudes.

The dramatic increase in species richness from high to low latitudes is one of the few rules in ecology[1,2]. This pattern is particularly striking in tree communities, where the number of species locally co-occurring in a tropical rainforest can be more than two orders of magnitude larger than the number of co-occurring species in a temperate forest[3,4]. Beginning with some of the earliest work contemplating these patterns, the role of biotic interactions has been emphasized[5,6]. For example, specialized pests and pathogens have been proposed as major regulators of tropical tree populations via conspecific negative density dependence, which should promote species co-existence[7,8]. Other researchers have focused on the belief that tropical climates are, and have been, relatively benign compared with the temperate zone and that biotic interactions probably play a larger role in producing and maintaining tropical tree diversity than abiotic factors[5,6]. Potential mechanisms underlying this framework include escape-and-radiate[9] and red queen[10] dynamics between plants and pests. These influential works (and others[11–13]) have produced the expectation that biotic interactions are more important in diverse tropical tree assemblages than in less diverse temperate assemblages.

Despite the general interest in biotic interactions along tropical-to-temperate tree species richness gradients, several foundational predictions remain poorly tested[14–16]. Among these is the expectation that plant phytochemistry should be more diverse within and between communities in the tropics due to increased,

[1]CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, China. [2]University of Chinese Academy Sciences, Beijing, China. [3]School of Ethnic Medicine, Key Lab of Chemistry in Ethnic Medicinal Resources, State Ethnic Affairs Commission & Ministry of Education of China, Yunnan Minzu University, Kunming, China. [4]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA. ✉e-mail: yangjie@xtbg.org.cn

https://doi.org/10.1038/s41559-024-02444-2

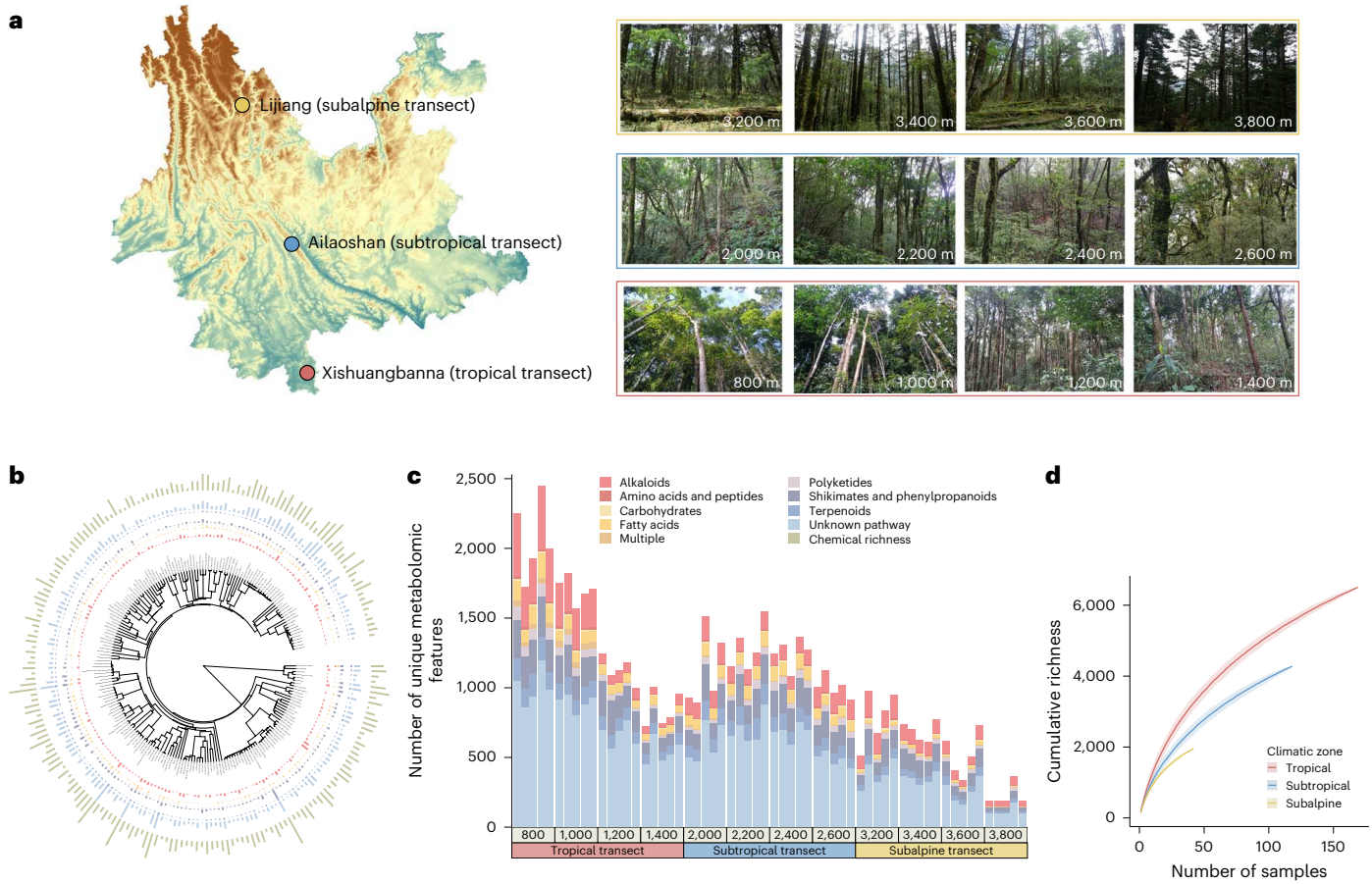



**Fig. 1 | Geographic location, forest landscapes, and phylogenetic and elevational distributions of chemical profiles across the three forest types. a**, The geographical locations and forest landscapes of the study region, established in tropical (Xishuangbanna: 800 m, 1,000 m, 1,200 m and 1,400 m), subtropical (Ailaoshan: 2,000 m, 2,200 m, 2,400 m and 2,600 m) and subalpine (Lijiang: 3,200 m, 3,400 m, 3,600 m and 3,800 m) forests. **b**, The phytochemical richness of 206 seed plant species arranged by their phylogenetic relationships. The size of the multicoloured bars represents the number of unique metabolomic features of different metabolite groups (entire metabolites and seven biosynthetic pathways) detected in each tree species. **c**, The number of unique metabolomic features belonging to each biosynthetic pathway category found in each elevational belt across three climatic zones. **d**, Cumulative metabolomic feature richness as a function of the number of samples examined within three climatic zones. The lines and shaded bands show the average richness and the 10th and 90th percentiles of richness. The accumulation curves did not approach saturation for any climatic zone.

particularly specialized, herbivore pressure. Specifically, plant communities are expected to have higher phytochemical diversity in the tropics due to an increase in specialized herbivores, which may drive species to be more dissimilar in their phytochemistry[17–19]. Furthermore, phytochemistry is expected to be extremely evolutionarily labile at finer taxonomic scales because of the rapid evolutionary changes in the metabolome, and it should therefore display little to no phylogenetic signal[20]. As sessile organisms, plants are renowned for their production of a wide variety of chemicals (that is, specialized metabolites) essential for mediating their interactions with the abiotic and biotic environment[21–23]. Characterization of this phytochemical diversity, while essential for testing predictions relating biotic interactions to changes in species richness from the tropics to the temperate zone, has been a persistent challenge[24,25]. Recent advances in plant metabolomics are quickly removing this barrier, resulting in detailed studies of community phytochemical diversity[26,27] and investigations into how phytochemistry has evolved in diverse lineages[28].

Here we address fundamental predictions regarding how phytochemical diversity and herbivore interactions change from species-rich to species-poor tree assemblages. We do this by leveraging recent advances in community metabolomics[24] used in conjunction with forest inventory plot data in tropical, subtropical and subalpine zones in a single geographic region. We use a multi-scale approach that allows us to test the prediction that phytochemical diversity is higher within and among communities in tropical forests than in less species-rich subtropical and subalpine forests. We marry these data with detailed measures of herbivore damage and specialization to test the prediction that these variables have higher values in the tropics. Finally, we quantify phylogenetic signal in the phytochemical similarity between species to test whether closely related species diverge more than expected in their phytochemicals.

## Results and discussion

We investigated tree communities in Yunnan province in China. Yunnan is trisected by three large rivers and is characterized by steep topographic and climatic gradients (Fig. 1a). Combined, this geological complexity results in Yunnan being one of the floristic hotspots of the world, containing an elevational gradient of tropical rainforest to subtropical forest to subalpine forest over a relatively short distance[29,30]. The region is thus ideal for investigations into how forest composition and dynamics change from the tropical to the subalpine biomes while minimizing major floristic changes as much as possible. During 2011 and 2012, we installed long-term forest inventory plots along this gradient. In the lowland tropical rainforest, we installed five 20 m × 20 m forest plots near 800 m, 1,000 m, 1,200 m and 1,400 m above sea level (a.s.l.). In the subtropical forest, we installed five plots near 2,000 m, 2,200 m, 2,400 m and 2,600 m a.s.l. Finally, in the subalpine zone, we installed

five plots near 3,200 m, 3,400 m, 3,600 m and 3,800 m a.s.l. (Fig. 1a). There are thus a total of 60 forest inventory plots in the study system. In each plot, we have measured the diameter and mapped the position of individual trees and identified them to species. The species diversity in the forest plots decreases steadily from the tropical rainforest to the subalpine forest and is correlated with elevation (Supplementary Fig. 1).

This study used a community metabolomics approach[26] to characterize the specialized metabolites of tree communities in the 60 forest inventory plots. We characterized the specialized metabolites of each species in each elevation belt (for example, plots near 800 m a.s.l.) in which it occurred. This was accomplished by pooling leaf samples taken from individuals in the plots at that elevation belt. If a species had more than three individuals in the plots, then only three individuals were sampled. If the species had fewer than three individuals, then all of the individuals present in the plots at that elevation were sampled. While the pooling of multiple individuals per site limited our ability to quantify local intra-specific variation, the number of samples collected is similar to that in previous studies and permits the characterization of and insights into specialized metabolites at the species level[26,27]. Collected leaves were nearly fully expanded, non-lignified, healthy and intact. The leaves were flash-frozen in the field, and the extraction of metabolites followed the protocols described in Sedio et al.[31]. The extractions were analysed using ultra-high-performance liquid chromatography–high-resolution tandem mass spectrometry (UHPLC–MS/MS). The output data were converted to raw spectral data that were deposited to the Global Natural Products Social Molecular Networking (GNPS) platform[32], and a molecular network was generated using classical molecular networking[33] with recommended parameter settings[31,34]. The spectral data were annotated using the publicly available datasets in the GNPS and then were classified into seven biosynthetic pathways using a deep-neural-network-based natural product structural classification tool called NPClassifier[35]. In total, we collected leaves from 979 individuals belonging to 206 species, 122 genera and 56 families representing 329 unique species-by-elevation combinations. Further details regarding this community metabolomic approach can be found in the Methods.

### Overview of the metabolite profiles of tree species

The metabolomic analyses recovered 9,327 metabolomic features, of which 4,070 (43.6%) could be annotated using public GNPS data (Fig. 1c). The most abundant annotated metabolomic biosynthetic pathway was shikimates and phenylpropanoids (32.31%), followed by alkaloids (27.08%), terpenoids (19.07%), fatty acids (8.08%), polyketides (7.71%), multiple (2.51%), carbohydrates (1.77%) and amino acids/peptides (1.47%). Next, using the specialized metabolite composition of the species-by-site combinations, we quantified sampling curves in each of the three forest types. This was done to determine the relative accumulation of phytochemical richness within the three forest types. The results show that tropical forests accumulate phytochemical richness faster on a sample-by-sample basis than subtropical and subalpine forests (Fig. 1d), which was similar to the result from a plot-by-plot basis (Fig. 2b). Furthermore, none of the sampling curves approached an asymptote, indicating the extraordinary phytochemical richness in trees on landscape to regional scales.

### Difference in metabolite composition between climatic zones

Because few metabolomic features are typically shared between species, we next quantified the structural similarity of the metabolomic features across species. This resulted in a chemical structural–compositional similarity (CSCS) matrix, which we converted to a chemical dissimilarity matrix (Methods). To visualize and evaluate the specialized metabolite composition of species-by-site combinations, we used a non-metric multi-dimensional scaling (NMDS) ordination using the chemical dissimilarity matrix (Fig. 2a). The magnitude and significance of the dissimilarity in chemical composition between the species in the tropical, subtropical and subalpine zones were tested using permutational multivariate analysis of variance (PERMANOVA) using 999 permutations of the matrix. The PERMANOVA results show that the chemical composition was different between the three regions ($R^2 = 0.090$, $F = 16.153$, $P = 0.001$). Interestingly, the chemical composition of the tropical species largely overlapped in the NMDS with that of the subtropical species, while also having a clearly distinct region in the NMDS space (Fig. 2a). The subalpine species slightly overlapped with the subtropical and tropical species in the NMDS space (Fig. 2a). Thus, the chemical composition of the tropical and subtropical species was generally divergent from that of the subalpine species. The PERMDISP results show that the group mean dispersion between the three climatic zones was significant (permutation test, $F = 40.025$, $P = 0.001$). We note that groups that are inferred to be different may simply be different due to differences in dispersion, and disentangling these two factors is challenging. Lastly, the average distance of group members to the group centroid was significantly different among the three climatic zones (Supplementary Fig. 4). The much higher average distance to the centroid in the tropical zone indicates that the sample-to-sample variation in structural diversity is higher for tropical communities.

### Difference in phytochemical diversity across climatic zones

We quantified the phytochemical diversity of trees within and between communities (that is, phytochemical alpha and beta diversity). We took this approach because increased biotic pressure should lead to greater phytochemical diversity locally and regionally. Measurements of community diversity can differentially weight the abundance of species, and comparing the results of differential weighting can be informative, particularly for phytochemical diversity[25]. Hill numbers provide an ideal analytical framework for this purpose[25,36]. For each of the 60 plots in the dataset, we calculated the phytochemical diversity of the species in the plot using the species-level phytochemical dissimilarity matrix based on CSCS scores and Hill numbers (${}^{q}D_{\mathrm{alpha}}$) with exponents 0, 1 and 2. When $q = 0$, the metric ignores species abundance; when $q = 1$, the metric weights all species in proportion to their abundance, and ${}^{1}D$ is equal to the exponential of Shannon's diversity; and when $q = 2$, the metric puts more weight on abundant species, and ${}^{2}D$ is equal to the inverse Simpson diversity[36]. One key advantage of Hill numbers is that they can be used to characterize species, trait, phytochemical and/or phylogenetic diversity using a unifying mathematical framework. Comparing diversity values for the same communities while varying the degree to which rare species are weighted (that is, changing $q$) thus provides more refined insights than using a single metric or a series of mathematically incoherent metrics. It is important to note that our approach investigates the overall phytochemical dissimilarity of species and not the dissimilarity of individual compounds, which is how Hill numbers have been previously used in the phytochemical diversity literature[25].

The observed phytochemical alpha diversity was the highest in the tropical plots, followed by the subtropical and then the subalpine plots, using each of the Hill numbers (that is, ${}^{0}D_{\mathrm{alpha}}$, ${}^{1}D_{\mathrm{alpha}}$ and ${}^{2}D_{\mathrm{alpha}}$) (Fig. 2c). We found a similar result when comparing the Hill numbers to the tree species richness in each forest plot (Supplementary Fig. 2). That is, as the tree species richness increased, the phytochemical diversity increased using each of the Hill numbers, which results in coincidental gradients in species richness and phytochemical diversity from the tropics to the subalpine. We next considered the phytochemical alpha diversity for individual chemical biosynthetic pathways. Tropical forests were the most diverse with respect to alkaloids, terpenoids, shikimates and phenylpropanoids, and amino acids/peptides, and this was consistent across most Hill number exponents (Extended Data Fig. 1 and Supplementary Table 3). Subtropical forests were slightly more diverse in polyketides only when using ${}^{1}D_{\mathrm{alpha}}$, and they were

https://doi.org/10.1038/s41559-024-02444-2

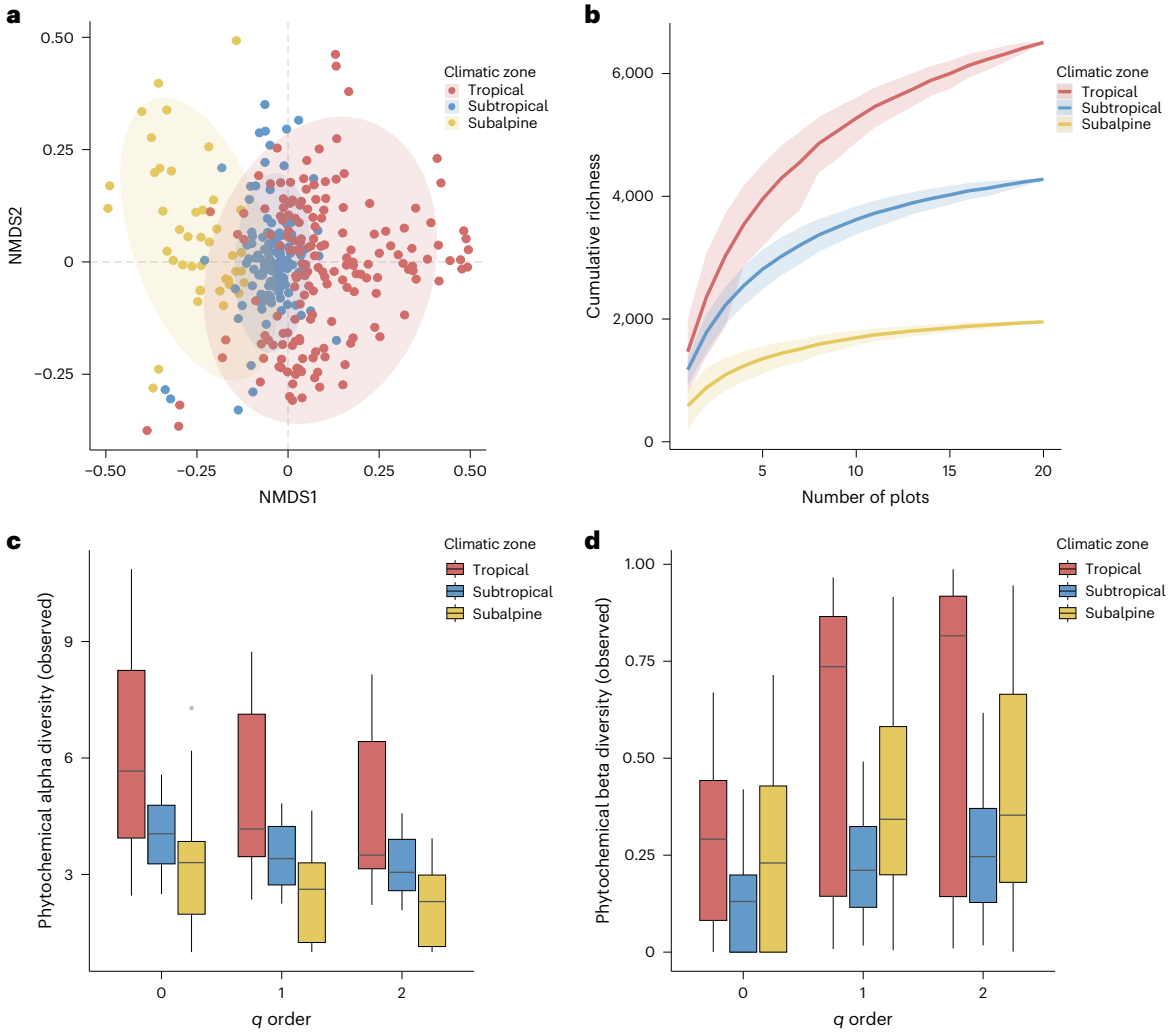



**Fig. 2 | Phytochemical variation and diversity (alpha and beta) among three forest types for whole metabolite profiles. a**, NMDS visualization of the metabolomic features across three climatic zones. The ellipsoids for each climatic zone represent the 95% confidence interval around their mean position in chemical space. The stress value is 0.260. **b**, Cumulative richness of metabolomic features as a function of the number of plots examined across three climatic zones. The lines and shaded bands show the average richness and the 10th and 90th percentiles of richness. **c**, Observed phytochemical alpha diversity within each plot ($n = 20$ in each forest type for one of $q$ exponents) in each climatic zone (tropical, subtropical and subalpine) using three $q$ exponents (0, 1 and 2), which quantify variability in the chemical trait space across trees in a plot on the basis of pairwise chemical distance (from CSCS scores) between tree samples and their weights (species abundance) in forest communities. **d**, Observed phytochemical beta diversity between plots ($n = 190$ in each forest type for one of $q$ exponents) in each climatic zone (tropical, subtropical and subalpine) using three $q$ exponents (0, 1 and 2). The phytochemical beta diversity measures the observed chemical dissimilarity between plots and reflects variability in the chemical trait space from plot to plot, on the basis of pairwise chemical distance (from CSCS scores) between tree samples and their weights (species abundance) in forest communities. In **c**,**d**, the significance of the differences in phytochemical alpha and beta diversity across forest type pairs was tested using an ANOVA with a post-hoc Tukey test. In each box plot, the centre line represents the median, the lower and upper box edges correspond to the 25th and 75th percentiles, and the lower and upper whiskers extend to the lowest and highest points to a limit of 1.5 × the interquartile range from the closest edge.

indistinguishable from tropical forests with respect to fatty acids and carbohydrates.

We also used a Hill number framework to calculate phytochemical beta diversity (${}^{q}D_{\mathrm{beta}}$), using $q = 0$, 1 and 2. The results show that phytochemical beta diversity is the highest among tropical tree plots when weighting by abundance (that is, ${}^{1}D_{\mathrm{beta}}$ and ${}^{2}D_{\mathrm{beta}}$) and indistinguishable from subalpine plots when weighting by presence/absence (that is, ${}^{0}D_{\mathrm{beta}}$) (Fig. 2d). When considering individual biosynthetic pathways, subalpine forests had higher beta diversity in alkaloids, fatty acids and carbohydrates. Tropical forests, in contrast, had higher beta diversity in terpenoids when weighting by abundance, as well as in shikimates and phenylpropanoids and in polyketides (Extended Data Fig. 2 and Supplementary Table 4). Thus, as with the phytochemical alpha diversity results, the phytochemical beta diversity results provide general support for an increased role of biotic interactions in lower latitudes.

Next, we considered species phytochemical beta diversity in a spatial context by quantifying distance decay relationships across the whole system and within each forest type by regressing community phytochemical beta diversity on the elevation distance between plots (Extended Data Fig. 3). Significant distance decay relationships were found across the entire study system and within the forest types. The distance decay slope was generally steeper in tropical plots than in subtropical and subalpine plots, indicating that the phytochemical composition in tropical plots changes much more quickly through space than it does in the other forest types. The steeper distance decay in the tropics was primarily due to rapid changes in some of the biosynthetic pathways that had the most beta diversity in the non-spatial

analyses (that is, terpenoids, shikimates and phenylpropanoids, and polyketides) (Extended Data Figs. 4–10). In sum, these results provide robust support for one of the key expectations of the biotic interactions hypothesis: that phytochemical alpha and beta diversity should be the highest in the tropics and greater than that expected given the underlying species diversity gradient.

While the observed phytochemical alpha and beta diversity results provide support for the hypothesized increase towards tropical rainforests, we were interested in quantifying whether the observed values were higher or lower than that expected given the regional (that is, gamma) phytochemical diversity for a given forest type[37]. That is, we sought to quantify whether the elevated phytochemical diversity in the tropics was primarily due to extraordinary local (that is, alpha) phytochemical diversity or the spatial turnover of (that is, beta) phytochemical diversity. Previous work has demonstrated that disentangling these factors requires a null model the considers gamma diversity in each region[38]. We therefore took a null modelling approach where we constructed null communities by randomizing the names of species-by-site combinations within each region in the phytochemical dissimilarity matrix based on CSCS scores to produce 999 random assemblages from each plot; compared the observed ${}^0D_{\mathrm{alpha}}$, ${}^1D_{\mathrm{alpha}}$ and ${}^2D_{\mathrm{alpha}}$ values to the null distribution; and calculated a standardized effect size (SES). We then performed the same procedure for ${}^0D_{\mathrm{beta}}$, ${}^1D_{\mathrm{beta}}$ and ${}^2D_{\mathrm{beta}}$ (Methods). SES values that significantly deviate positively from zero indicate observed diversity levels that are greater than expected, whereas significantly negative deviations of SES values from zero indicate less observed diversity than expected.

The phytochemical alpha diversity results show that the median SES value in forest types was either indistinguishable from zero or significantly less than zero when considering all phytochemicals and individual biosynthetic pathways (Supplementary Table 5). This indicates that the observed phytochemical alpha diversity in the plots was either no different from that expected given the phytochemical gamma diversity or less than that expected. There were slight differences between the forest types. Specifically, subalpine plots were typically no different from random except when using ${}^2D_{\mathrm{alpha}}$, whereas subtropical and tropical plots were almost always less diverse than expected given the observed gamma diversity.

The phytochemical beta diversity results generally differed from the alpha diversity results. Specifically, when considering all phytochemicals or individual biosynthetic pathways, the turnover in phytochemicals from plot to plot in the tropics was higher than expected, except for amino acids/peptides and, in some instances, abundance-weighted fatty acids, carbohydrates and alkaloids (Supplementary Table 6). The extreme phytochemical gamma diversity observed in the tropics can thus be primarily attributed to an unusually high degree of spatial turnover in phytochemistry between plots.

The subtropical forest type had a mixture of less and more than expected turnover in phytochemistry between plots, whereas the subalpine forest had plots that were generally more similar than expected in their phytochemistry. The main exception were the alkaloids, which always differed more than expected between subalpine plots. From these null modelling results, it is apparent that the increase in phytochemical diversity towards the tropics found in this study is primarily attributable to exceptional spatial turnover in phytochemistry between plots and not local (that is, within a 20 m × 20 m plot) phytochemical diversity.

## Higher herbivore selective pressure in the tropical zone

The leaf phytochemical richness and diversity results implicate the increased importance of biotic interactions in tropical forests relative to subtropical and subalpine forests. Specifically, the high phytochemical diversity within and between tree communities is expected if there is an increase in herbivory and, in particular, herbivory via specialized pests in tropical forests. We tested this expectation by quantifying herbivory rates and herbivore specialization in each species in each of the forest plots. Specifically, we quantified leaf herbivory on ten leaves from three branches taken from three adult individuals of a species in each forest plot ($n = 90$). If the species had fewer than three adult individuals, we sampled ten leaves on each of the three branches selected for all present individuals. The collected leaves were pressed, dried and scanned using a flatbed scanner. We then estimated the amount of herbivore damage using leaf-area-lost measures at the species level. We then calculated a mean per cent leaf area lost in a plot weighting each species by the plot abundance. We used standardized approaches for estimating herbivore specialization on the basis of hole- and marginal-type insect feeding damage[39,40]. From this, we obtained the frequency and proportion of damage caused by generalist, intermediate or specialized herbivores for each species, and we calculated a mean percentage for each of these at the plot level weighting each species by abundance. The results show a clear decline in herbivore-caused leaf damage from tropical to subtropical to subalpine forests (Fig. 3a). Furthermore, the amount of damage attributed to intermediate and specialized hole- and margin-type feeding insects decreased from tropical to subalpine forests (Fig. 3b,c), whereas the damage due to generalist herbivores was lowest in tropical forests and increased towards subalpine forests (Fig. 3d). Thus, along with increased alpha and beta phytochemical diversity in leaves in tropical tree communities, we found an increase in leaf herbivory and the degree of specialized herbivory in the tropics.

## Phylogenetic divergence of phytochemistry between trees

The coinciding increase in herbivory, particularly by specialists, and higher-than-expected phytochemical beta diversity in the tropical forest plots indicates that plant–pest interactions may be critical generators and maintainers of tropical tree diversity. Two macroevolutionary possibilities could explain this outcome. The first is that the evolution of phytochemical diversity is relatively constrained such that the increase in phytochemical diversity observed in a sample (for example, a forest type) is simply due to an increase in the phylogenetic diversity. Under this scenario, inter-specific phytochemical similarity would exhibit phylogenetic signal. The second is that the evolution of phytochemical diversity is highly dynamic within lineages such that closely related species exhibit high degrees of phytochemical divergence. Under this scenario, inter-specific phytochemical similarity would lack phylogenetic signal. We constructed a phylogeny containing all species in our system (Fig. 1c) and quantified the phylogenetic signal in phytochemical similarity (Methods). We found that inter-specific phytochemical similarity lacked phylogenetic signal and this was also generally true across individual chemical biosynthetic pathways (Supplementary Table 7). These results support the hypothesis that phytochemical evolution is highly dynamic, probably in response to biotic interactions, and that the elevated phytochemical diversity observed in tropical forests is not simply the result of these forests containing more phylogenetic diversity or phylogenetic clustering of lineages in individual forest types in this system.

That biotic interactions are stronger or more important in the tropics is one of the most widely believed but poorly tested hypotheses in ecology and evolutionary biology. Here we have provided several lines of evidence that provide broad support for core predictions made from the biotic interactions hypothesis: phytochemical alpha and beta diversity are highest in tropical tree communities, the phytochemical dissimilarity between local communities in the tropics is exceptionally high and is a key generator of regional phytochemical diversity, overall and specialized herbivore damage are higher in the tropics, and closely related species are more divergent in their phytochemistry than expected. The results also underscore the multiple dimensions of tropical biodiversity, which are frequently not quantified, are valuable for human societies and are under threat as global change continues[41,42]. Not only do tropical forests contain more species than their temperate counterparts, but they house spectacular phytochemical

https://doi.org/10.1038/s41559-024-02444-2

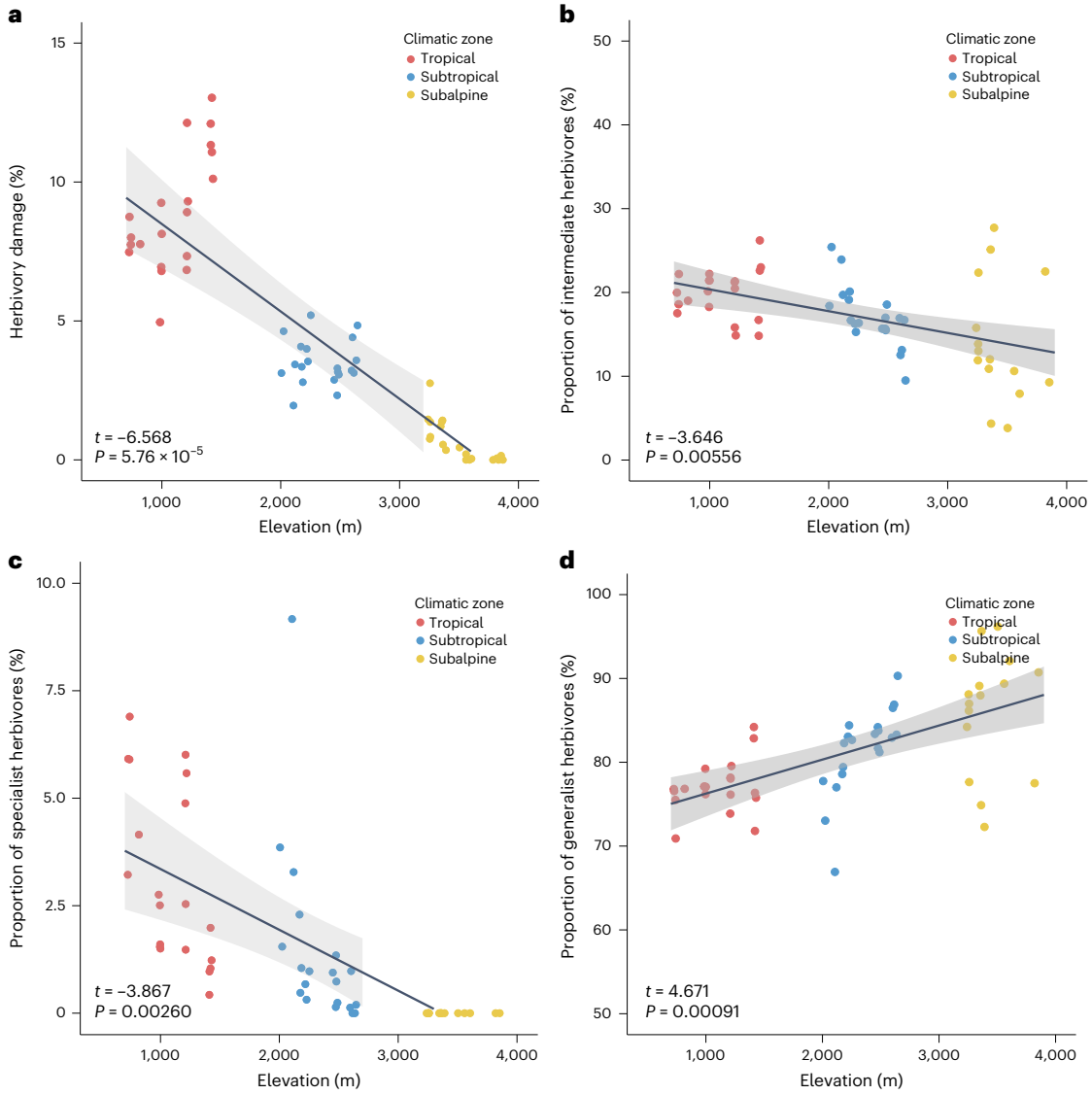



**Fig. 3 | Herbivory damage and specialization across the elevational gradient.** **a**, The plot-level herbivore damage. **b**, The plot-level proportion of intermediate herbivores. **c**, The plot-level proportion of specialist herbivores. **d**, The plot-level proportion of generalist herbivores. All data points are shown as circles and represent individual plots ($n = 20$ plots in each climatic zone). The black lines represent linear fits from regressions, and the grey shaded areas indicate 95% confidence intervals.

diversity, such that a loss of tropical tree species is outpaced by the resulting loss of phytochemical diversity that is foundational for promoting and maintaining biodiversity. We stress that this is necessarily an observational study of hundreds of species from many clades along a substantial climatic and spatial gradient. There are thus many abiotic (for example, temperature and precipitation) and biotic (for example, herbivores, pathogens, neighbourhood composition and diversity) covariates that are also likely to be related to phytochemical diversity and that cannot be cleanly disentangled to elucidate a specific mechanism. That said, here we have provided extensive evidence of gradients in phytochemical diversity and herbivore damage from tropical to subalpine forests that have been predicted or assumed in many core hypotheses in plant ecology and evolution, but have been poorly documented and increasingly questioned.

## Methods

### Study area

No specific permits were required for the described field studies, as no endangered or protected species was involved, and the localities involved are not protected in any way. This study was conducted along an elevational gradient that encompasses three climatic regions—tropical, subtropical and subalpine—in Yunnan Province, a global biodiversity hotspot in southwestern China (Fig. 1a). During 2011 and 2012, four elevational belts were established in each climatic region[43,44]. Specifically, in the tropical region, four elevational belts (800 m, 1,000 m, 1,200 m and 1,400 m) were established in Xishuangbanna National Nature Reserve and are characterized by vegetation that includes tropical seasonal rainforest, tropical montane rainforest and tropical montane evergreen broad-leaved forest. In the subtropical region, four elevational belts (2,000 m, 2,200 m, 2,400 m and 2,600 m) were established in Ailaoshan National Nature Reserve and are characterized by subtropical evergreen broad-leaved forest. In the subalpine region, four elevational belts (3,200 m, 3,400 m, 3,600 m and 3,800 m) were established in Yulong Snow Mountain Nature Reserve and are characterized by subalpine coniferous forest.

In each elevational belt, five 20 m × 20 m replicate plots were established approximately 200 m apart from each other. Locations impacted by anthropogenic and natural disturbances, such as large

canopy gaps, were avoided. Within each plot, all trees with diameter at breast height >5 cm were measured, tagged and identified in previous studies[43,44]. Thus, in total, 60 forest plots from 12 elevational belts across three climatic regions were included in the analyses (Supplementary Table 1). Supplementary Fig. 1 shows the comparison of plot-level species diversity using Hill numbers with diverse $q$ orders ($q$ = 0, 1 and 2) between three climatic zones, and the elevational pattern of plot-level species diversity using Hill numbers with diverse $q$ orders ($q$ = 0, 1 and 2) across three climatic zones. When $q$ = 0, the metric is species richness; when $q$ = 1, the metric is an exponential of Shannon entropy; and when $q$ = 2, the metric is 1 / (1 − Simpson diversity).

### Leaf sampling and metabolite extraction

We collected nearly fully emerged, unlignified and undamaged leaves of trees in each forest plot during the growing season in 2021. In each forest plot, we randomly sampled three individuals of each common species and collected all individuals of species with less than three individuals in the plot. For each individual, two to ten of the expanded leaves were sampled, labelled, wrapped in foil and immediately immersed in liquid nitrogen in the field, and then stored at −80 °C in a laboratory freezer. In total, we collected leaf samples from 979 individuals of 206 species, belonging to 122 genera and 56 families.

Leaf material from each species in a single elevation belt was pooled for downstream metabolomic analyses. Although the pooling of leaf material ignores individual-level variation, our primary goal was to characterize species-level phytochemistry, and recent studies have indicated that intra-specific variation in a local-scale sample is far less than inter-specific variation[45–47]. After pooling samples, across the environmental gradient, 329 unique species–elevation combinations (hereafter, samples) were included in the analysis.

The extraction of leaf metabolites followed the protocol described in Sedio et al.[31], modified slightly to be compatible with our laboratory equipment. Briefly, we weighed ~300 mg of frozen pooled samples, ground them into a fine powder while using liquid nitrogen and extracted the metabolites with 1.8 ml of extraction solution (ethanol:water, 90:10 at pH 4.5). The extracted samples were stored in a refrigerator at 4 °C overnight in a microcentrifuge tube. This solvent extracted small molecules of varying polarity, and the mild acidity facilitated the extraction of alkaloids. The extracted samples were then shaken at 0.14985 $g$ at room temperature for three hours, followed by centrifugation at 23,931.8 $g$ for ten minutes. The supernatants were removed and transferred to new microcentrifuge tubes. After the extraction process, the extraction solution was evaporated using a vacuum concentrator to increase the concentration of the extracted molecules. Subsequently, 500 µl of extraction solution was added to dissolve the samples, which were then filtered with a hand-actuated Millipore filter ($\Phi$ = 0.22 µm) and transferred to 2 ml brown chromatographic vials. Before performing metabolite analysis, we added 800 µl of extra extraction solution to dilute the solution of the samples.

### Metabolomics data analysis

The metabolomic data from the extracted samples were analysed using UHPLC–MS/MS. Chromatographic separation was performed using an Agilent 1290 Infinity Series UHPLC instrument (Agilent Technologies), which was equipped with a quaternary pump (G4204A), a de-gasser, a diode-array detector (G4212B), an autosampler (G4226A) and a column compartment (G1316C). Chromatographic separation was achieved on a Hypersil GOLD aQ column (5.0 µm; Thermo Fisher Scientific) with 250 mm length and 4.6 mm internal diameter. The mobile phase consisted of a mixture of pure water with 0.1% formic acid (A) and acetonitrile (B). The chromatographic elution method was set as follows: 0.00–20.00 min, 100% A; 20.00–50.00 min, a mixture of 80% A and 20% B; 50.00–55.00 min, 100% B; equilibration 55.00–60.00 min, gradient from 100% to 0% B. The flow rate was set at 1.0 ml $min^{-1}$.

Mass spectrometric analysis was carried out on a quadrupole time-of-flight high-resolution mass spectrometer (Q-TOF LC/MS 6540 series, Agilent Technologies) coupled with electrospray ionization. The data acquisition was performed using Mass Hunter Workstation software in positive electrospray ionization mode. The fragmentor voltage was set at 135 V, the capillary was set at 3,500 V, the skimmer was set at 65 V and nitrogen was used as the drying (350 °C, 8 l $min^{-1}$) and nebulizing (30 psi) gas. Higher-energy collision-induced dissociation was performed with a normalized collision energy of 10, 20, 30, 40 and 50 eV in stepped mode. The MS detection was performed in positive electrospray ionization mode over a mass range of 100–1,700 $m/z$ for MS1 and 20–1,700 $m/z$ for MS2. MS/MS spectra were acquired for each sample by running data-dependent acquisition mode, in which all precursor ions from the full mass range were fragmented to yield MS/MS spectra.

The raw MS/MS data files were centroided and converted from the proprietary format (.d) to the $m/z$ extensible markup language format (.mzML) using ProteoWizard (v.3.0.22112, MSConvert tool)[48], and the converted raw spectral data were uploaded to the GNPS platform. A molecular network was generated using classical molecular networking[33] with parameter settings as recommended in Sedio et al.[31] and Aron et al.[34], including a precursor ion mass tolerance of 2.0 Da, a fragment ion mass tolerance of 0.5 Da, a minimum pairs cosine of 0.6, a minimum matched fragment ions of 3, a minimum cluster size of 2, a network topK of 10, a run MSCluster of yes, a maximum connected component size of 0, a search analogues of Do Search and a maximum analogue search mass difference of 100. While processing the molecular network, MS/MS spectra of fragmented molecules were clustered into consensus spectra that represent a single unique molecular structure (hereafter, metabolomic feature).

The MS/MS spectral matching was performed using all publicly available datasets in GNPS to annotate metabolomic features. Each annotated metabolomic feature was classified into different biosynthetic pathway categories (including fatty acids, polyketides, shikimates and phenylpropanoids, terpenoids, alkaloids, amino acids/peptides, and carbohydrates), using a deep-neural-network-based natural product structural classification tool called NPClassifier[35]. Metabolomic features that did not match with the available databases were declared as unknown metabolomic features. Additionally, some metabolomic features were assigned to a mixed pathway (for example, 'alkaloids|amino acids/peptides' and 'polyketides|terpenoids'). We recategorized these metabolites using the following approach. When a mixed metabolomic feature contained a first-order classification (alkaloids, polyketides, shikimates and phenylpropanoids, or terpenoids) and a second-order classification (amino acids/peptides, carbohydrates or fatty acids), we recategorized it using the first-order classification. However, when a metabolomic feature was characterized with two first-order or second-order categories, we called it 'multiple' in the subsequent analyses.

We used consensus spectra derived from molecular networking to approximate the functional representation of phytochemical diversity directly from the organized spectral space[27,49]. We employed a compound-based molecular networking approach, where we first grouped related features into compounds and then generated a species-by-compound abundance matrix and a compound-by-compound MS/MS cosine similarity matrix. As described in Sedio et al.[49], we combined these data and quantified the chemical similarity between each sample pair by computing the CSCS index. The CSCS index was calculated as the weighted average structural similarity of each metabolomic feature pair present in either sample, with the weights determined by the ion intensities of the metabolomic features in the compared samples. We calculated the CSCS index for the entire metabolome and for each of the seven biosynthetic pathway categories separately for the subsequent analyses. The dissimilarity of the chemical profiles between each sample pair was calculated as one minus the chemical similarity.

### Plant specialized metabolite composition across elevation

To evaluate the variation in metabolite composition among the three forest types, we used NMDS ordinations for the entire metabolite profile. This approach was applied using the metaMDS function in the vegan package[50], on the basis of the sample-pair matrix of chemical dissimilarity (from the CSCS index described above). The significance and magnitude of the dissimilarity in chemical composition between the three forest types were tested with a PERMANOVA method, using the adonis2 function in the vegan package[50] and 999 permutations. The $R^2$ value derived from the PERMANOVA was used to index the strength of pairwise chemical dissimilarity among the different climatic transects. Furthermore, a multivariate analysis of the homogeneity of group dispersions (PERMDISP) was also performed using the function betadisper in the vegan package[50] to test whether one or more forest types is more variable than the others. The betadisper function calculated the distances from each sample to the group centroid, and statistical support for differences in dispersion across forest types was assessed using a permutation test ($n$ = 999 permutations) followed by a post-hoc Tukey's honestly significant difference test to assess pairwise differences among forest types.

To visualize chemical diversity across samples in the three forest types, we conducted a sampling curve analysis. Specifically, we used data on the presence or absence of metabolomic features to generate sampling curves in each forest type using the rarep function replicated 999 times, as described in Wetzel & Whitehead[25]. We also performed the analysis using a plot-by-plot accumulation curve for metabolomic features. Specifically, on the basis of the presence/absence data, we merged all metabolomic features extracted from the tree samples in each forest plot. We then built a plot–metabolomic features matrix, with 60 rows (forest plots) and 9,327 columns (metabolomic features). Lastly, three plot-level accumulation curves (for tropical forest, subtropical forest and subalpine forest) were calculated.

### Phytochemical alpha diversity

To compare the local diversity of metabolites across the three climatic zones, we calculated functional Hill numbers as a measure of the phytochemical alpha diversity in each plot, for both the entire chemical profile and the seven biosynthetic pathway categories. Using functional Hill numbers to measure phytochemical diversity has several advantages[36]. First, the parameter $q$ controls the sensitivity of the measure to the relative abundances of species. By adjusting $q$, the behaviour of the index can be controlled to enable a more nuanced measure of diversity. For $q$ = 0, species abundance is ignored, and all species have the same weight. For $q$ = 1, all species are weighed in proportion to their abundance, and ${}^1D$ is equal to the exponential of Shannon's diversity. For $q$ = 2, more weight is put on abundant species, and ${}^2D$ is equal to the inverse Simpson diversity. If the most abundant species have different phytochemicals from the scarcest ones, the $q \geq 1$ indices will have different trends than the one with $q$ = 0. A second advantage is that functional Hill numbers are expressed as the effective number of functionally equally distinct species (or 'virtual functional groups') with all pairwise distances for different species pairs. The within-community phytochemical functional diversity in each plot was calculated on the basis of pairwise chemical distance between tree samples (from the CSCS index) and their weights (species abundance) in forest communities, to represent the variability in the chemical trait space across trees in a plot. The metric thus calculates dissimilarities between species in their entire phytochemistry and does not focus on the dissimilarity of individual compounds. Using the framework described in Chao et al.[36], we used the $q$ exponents 0, 1 and 2 implemented using the alpha.fd.hill function in the mFD package[51]. We compared the Hill numbers for a given $q$ exponent across the three forest types using an ANOVA with a post-hoc Tukey test to assess whether forest type pairs were significantly ($\alpha$ = 0.05) different in their phytochemical alpha diversity.

### Phytochemical beta diversity

To quantify the phytochemical beta diversity (that is, phytochemical dissimilarity) between plots, we also used a Hill number framework using $q$ exponent values of 0, 1 and 2. These calculations were performed using the beta.fd.hill function in the mFD package[51], on the basis of the pairwise chemical distance (from the CSCS index) between tree samples and their weights (species abundance) in forest communities, to represent variability in the chemical trait space from plot to plot. The phytochemical beta diversity between plots was calculated across the entire study system. The phytochemical beta diversity between plots within each forest type was compared across three forest types using an ANOVA with a post-hoc Tukey test.

Next, we examined the distance decay in phytochemical similarity between plots by regressing the phytochemical beta diversity of plot pairs (the chemical dissimilarity among communities) onto their differences in elevation scaled to a maximum of 1. This was done across the entire system and within forest types using generalized linear models with a log link and a quasi-binomial family[52,53]. The phytochemical beta diversity between plots was then rescaled to a range between 0 and 1. A value of 0 indicates identical phytochemical composition, and a value of 1 indicates complete phytochemical dissimilarity. The similarity of chemical composition between forest plots was determined as 1 minus the above chemical dissimilarity between forest plots. The rate of decrease in the observed phytochemical beta diversity with elevational distance was compared across forest types using the slopes of the regressions[54]. The 95% confidence interval of the slope was assessed using the confint function, and confidence intervals for the predictions were calculated using the add_ci function in the ciTools package[55]. All beta diversity analyses, like the alpha diversity analyses, were conducted on all phytochemical information combined and by each of the seven biosynthetic pathways.

### Null model analyses of phytochemical alpha and beta diversity

To assess whether the observed phytochemical alpha and beta diversities were higher than, no different from or lower than expected given the observed phytochemical diversity within a forest type (that is, phytochemical gamma diversity), we conducted null model analyses. The null model used randomized sample names in the chemical dissimilarity matrix (from CSCS scores) within a forest type. This was repeated 999 times, and during each iteration, null phytochemical alpha and beta diversity values were calculated for each of the $q$ exponent values. This approach only randomized the chemical dissimilarity of samples to one another, while maintaining the observed community data matrix. Species occupancy frequencies, abundances and spatial distributions (for example, any patterns of intraspecific aggregation or dispersal limitation) were thus fixed in each randomization. Using the null distributions, we calculated the SES values. The SES values were calculated by subtracting the mean value of the null distribution from the observed value and dividing this difference by the standard deviation of the null distribution[37]. Thus, positive values indicate observed values that are higher than the expected values from the null model, and negative values indicate observed values that are lower than the expected values. Wilcoxon signed-rank non-parametric tests were performed to test whether the SES values deviated from zero (that is, did not differ from a random expectation).

### Measurement of herbivory

To estimate herbivory in the study system, we measured the total leaf damage due to herbivores and the diet breadth of the herbivores in each forest plot. To quantify herbivore damage, we randomly sampled three mature individuals in each forest plot if there were three individuals available or fewer if less than three were available. On each individual, three branches were sampled in different random directions. For each branch, ten leaves were picked in order beginning from the end of the branch and placed into envelopes. The leaves were then transported

to the lab and scanned using a flatbed scanner (Epson Company). The eaten and lost areas of the scanned leaves were manually filled in using Adobe Photoshop software 2021 (Adobe Systems) by referencing an intact leaf of a similar size[39,56]. Next, the area of the original leaf and the area of the leaf with the missing regions filled in were measured using the LeafArea package[57]. The herbivory damage in each branch in a sampled individual was calculated as:

$$\text{Herbivory damage} = 1 - \frac{\text{Total area of original images of ten sampled leaves}}{\text{Total area of filled images of ten sampled leaves}}$$

To determine the species-level herbivore damage in each forest plot, the average herbivore damage across individuals was estimated. The plot-level herbivore damage was then calculated using a community-weighted mean using the dbFD function in the R package FD[58].

The damage types observed in each leaf were carefully scrutinized, identified and recorded on the basis of a published rubric for classifying herbivore damage types[40]. We focused on hole feeding and margin feeding herbivore damage types, which are based on a guide of >140 distinctive patterns of damage more likely caused by chewing and mining herbivores[39], excluding fungal and mechanical leaf damage[40]. Although the established leaf damage typology was originally designed for identifying the insect-mediated damage type in the leaf-fossil record, one benefit of using leaves and damage types to document plant–insect herbivore trophic relationships in modern ecosystems is that it offers an alternative method to quantifying insect herbivore diversity and abundance through insect sampling across landscapes, which has been widely used in previous studies[39,59–62]. For each leaf on each branch on each tree individual, we recorded the presence of all unique damage types. Each recorded damage type was assigned to one of three host specificities following the protocols described in Labandeira et al.[40], including generalist herbivore, intermediate herbivore and specialist herbivore (Supplementary Fig. 3). Next, the frequency and proportion of each host specificity in each individual tree was estimated and then averaged at the species level across individuals. A community-weighted mean of the proportion of each host specificity for each plot was then calculated using the dbFD function[58]. The spatial patterns of herbivorous selection pressure (herbivory damage percentage and the proportion of each host specificity) across the three climatic zones were quantified with linear mixed models using the lmer function in the lme4 package[63]. The elevation value in each plot was a fixed-effect term, and the elevational belt was a random-effect term.

### Quantifying phylogenetic signal in phytochemistry

We quantified phylogenetic signal using a species-level phylogeny containing the 206 tree species in this study. The phylogeny was generated using the phylo.maker function in the V.PhyloMaker R package[64] using the largest dated phylogeny for vascular plants (GBOTB.extended.tre). We calculated the phylogenetic signal in the phytochemical dissimilarity matrix using the $K_{mult}$ metric developed by Adams[65]. This was replicated using all phytochemicals as well as for the seven individual biosynthetic pathways. The $K_{mult}$ metric is a generalized version of the $K$ statistic[66]. Briefly, the metric quantifies the degree to which the observed inter-specific variation in metabolites matches the expected variation in a trait evolving under Brownian motion on the observed phylogeny. A $K_{mult}$ value of 1 indicates observed metabolite variation that matches Brownian motion (that is, phylogenetic signal), a value >1 indicates more phylogenetic signal than expected from a Brownian motion model (that is, closely related species have more similar phytochemistry than expected from a random walk on the phylogeny) and a value <1 indicates a lack of phylogenetic signal (that is, closely related species are more dissimilar in their phytochemistry than expected from a random walk). Because some species, such as *Abies forrestii*, *Camellia forrestii*, *Lithocarpus truncates* and *Phoebe minutiflora*, span multiple forest types, we had more than one metabolomic profile for these species. To resolve this issue, we conducted a resampling process by randomly selecting one representative per species in the analyses. This process was repeated 1,000 times, and the $K_{mult}$ value was calculated separately for each iteration. The phylogenetic tree was visualized with the iTOL tool (v.5)[67]. All statistical modelling analyses were performed in R v.4.2.2 (ref. 68).

### Reporting summary

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

## Data availability

The MS data (.mzML) have been deposited in the MassIVE public repository and are available under accession number MSV000092950. The datasets analysed in the current study, including the molecular network, sample–sample chemical structural and compositional similarity, plot-species-abundance community data, phytochemical richness and the phylogenetic tree of 206 tree species, are available via Figshare at https://doi.org/10.6084/m9.figshare.22758269 (ref. 69).

## Code availability

The R code used in the current study is available via Figshare at https://doi.org/10.6084/m9.figshare.22758269 (ref. 69).

## References


1. Hillebrand, H. On the generality of the latitudinal diversity gradient. *Am. Nat.* **163**, 192–211 (2004).
2. Willig, M. R., Kaufman, D. M. & Stevens, R. D. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Evol. Syst.* **34**, 273–309 (2003).
3. Gentry, A. H. Tree species richness of upper Amazonian forests. *Proc. Natl Acad. Sci. USA* **85**, 156–159 (1988).
4. Valencia, R. et al. in *Tropical Forest Diversity and Dynamism: Findings from a Large-Scale Plot Network* (eds Losos, E. C. & Leigh, E. G.) 609–628 (Univ. Chicago Press, 2004).
5. Dobzhansky, T. Evolution in the tropics. *Am. Sci.* **38**, 209–221 (1950).
6. Fischer, A. G. Latitudinal variations in organic diversity. *Evolution* **14**, 64–81 (1960).
7. Janzen, D. H. Herbivores and the number of tree species in tropical forests. *Am. Nat.* **104**, 501–528 (1970).
8. Connell, J. H. in *Dynamics of Numbers in Populations* (eds Den Boer, P. J. & Gradwell, G. R.) 298–312 (PUDOC, 1971).
9. Ehrlich, P. R. & Raven, P. H. Butterflies and plants—a study in coevolution. *Evolution* **18**, 586–608 (1964).
10. Van Valen, L. The red queen. *Am. Nat.* **111**, 809–810 (1977).
11. Coley, P. D. & Kursar, T. A. On tropical forests and their pests. *Science* **343**, 35–36 (2014).
12. Zvereva, E. L. & Kozlov, M. V. Latitudinal gradient in the intensity of biotic interactions in terrestrial ecosystems: sources of variation and differences from the diversity gradient revealed by meta-analysis. *Ecol. Lett.* **24**, 2506–2520 (2021).
13. Roslin, T. et al. Higher predation risk for insect prey at low latitudes and elevations. *Science* **356**, 742–744 (2017).
14. Moles, A. T. & Ollerton, J. Is the notion that species interactions are stronger and more specialized in the tropics a zombie idea? *Biotropica* **48**, 141–145 (2016).
15. Coley, P. D. & Barone, J. A. Herbivory and plant defenses in tropical forests. *Annu. Rev. Ecol. Evol. Syst.* **27**, 305–335 (1996).
16. Schemske, D. W., Mittelbach, G. G., Cornell, H. V., Sobel, J. M. & Roy, K. Is there a latitudinal gradient in the importance of biotic interactions? *Annu. Rev. Ecol. Evol. Syst.* **40**, 245–269 (2009).

17. Ali, J. G. & Agrawal, A. A. Specialist versus generalist insect herbivores and plant defense. *Trends Plant Sci.* **17**, 293–302 (2012).
18. Carmona, D., Lajeunesse, M. J. & Johnson, M. T. J. Plant traits that predict resistance to herbivores. *Funct. Ecol.* **25**, 358–367 (2011).
19. Jones, C. G. & Firn, R. D. On the evolution of plant secondary chemical diversity. *Phil. Trans. R. Soc. Lond. B.* **333**, 273–280 (1991).
20. Endara, M. J., Forrister, D. L. & Coley, P. D. The evolutionary ecology of plant chemical defenses: from molecules to communities. *Annu. Rev. Ecol. Evol. Syst.* **54**, 107–127 (2023).
21. Kessler, A. & Kalske, A. Plant secondary metabolite diversity and species interactions. *Annu. Rev. Ecol. Evol. Syst.* **49**, 115–138 (2018).
22. Wang, S., Alseekh, S., Fernie, A. R. & Luo, J. The structure and function of major plant metabolite modifications. *Mol. Plant* **12**, 899–919 (2019).
23. Iason, G. R., Dicke, M. & Hartley, S. E. *The Ecology of Plant Secondary Metabolites: From Genes to Global Processes* (Cambridge Univ. Press, 2012).
24. Sedio, B. E. Recent breakthroughs in metabolomics promise to reveal the cryptic chemical traits that mediate plant community composition, character evolution and lineage diversification. *N. Phytol.* **214**, 952–958 (2017).
25. Wetzel, W. C. & Whitehead, S. R. The many dimensions of phytochemical diversity: linking theory to practice. *Ecol. Lett.* **23**, 16–32 (2020).
26. 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).
27. Defossez, E. et al. Spatial and evolutionary predictability of phytochemical diversity. *Proc. Natl Acad. Sci. USA* **118**, e2013344118 (2021).
28. Forrister, D. L. et al. Diversity and divergence: evolution of secondary metabolism in the tropical tree genus *Inga*. *N. Phytol.* **237**, 63–642 (2023).
29. Qian, L. S., Chen, J. H., Deng, T. & Sun, H. Plant diversity in Yunnan: current status and future directions. *Plant Divers.* **42**, 281–291 (2020).
30. Zhu, H. & Tan, Y. H. Flora and vegetation of Yunnan, southwestern China: diversity, origin and evolution. *Diversity* **14**, 340 (2022).
31. Sedio, B. E., Boya P, C. A. & Rojas Echeverri, J. C. A protocol for high-throughput, untargeted forest community metabolomics using mass spectrometry molecular networks. *Appl. Plant Sci.* **6**, e1033 (2018).
32. GNPS: Global Natural Products Social Molecular Networking. *UCSD* https://gnps.ucsd.edu (2023).
33. Wang, M. X. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
34. Aron, A. T. et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* **15**, 1954–1991 (2020).
35. Kim, H. W. et al. NPClassifier: a deep neural network-based structural classification tool for natural products. *J. Nat. Prod.* **84**, 2795–2807 (2021).
36. Chao, A. et al. An attribute-diversity approach to functional diversity, functional beta diversity, and related (dis)similarity measures. *Ecol. Monogr.* **89**, e01343 (2019).
37. Swenson, N. G. *Functional and Phylogenetic Ecology in R* (Springer, 2014).
38. Kraft, N. J. B. et al. Disentangling the drivers of β diversity along latitudinal and elevational gradients. *Science* **333**, 1755–1758 (2011).
39. Wang, X. Z. et al. Niche differentiation along multiple functional-trait dimensions contributes to high local diversity of Euphorbiaceae in a tropical tree assemblage. *J. Ecol.* **110**, 2731–2744 (2022).
40. Labandeira, C. C., Wilf, P., Johnson, K. R. & Marsh, F. Guide to insect (and other) damage types on compressed plant fossils v.3.0. *figshare* https://doi.org/10.6084/m9.figshare.16571441.v1 (2007).
41. Kursar, T. A. et al. Linking bioprospecting with sustainable development and conservation: the Panama case. *Biodivers. Conserv.* **16**, 2789–2800 (2007).
42. 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).
43. Song, X. Y. et al. Different environmental factors drive tree species diversity along elevation gradients in three climatic zones in Yunnan, southern China. *Plant Divers.* **43**, 433–443 (2021).
44. Song, X. Y., Nakamura, A., Sun, Z. H., Tang, Y. & Cao, M. Elevational distribution of adult trees and seedlings in a tropical montane transect, Southwest China. *Mt. Res. Dev.* **36**, 342–354 (2016).
45. Endara, M. J. et al. Divergent evolution in antiherbivore defences within species complexes at a single Amazonian site. *J. Ecol.* **103**, 1107–1118 (2015).
46. Richards, L. A. et al. Phytochemical diversity drives plant–insect community diversity. *Proc. Natl Acad. Sci. USA* **112**, 10973–10978 (2015).
47. Salazar, D. et al. Origin and maintenance of chemical diversity in a species-rich tropical tree lineage. *Nat. Ecol. Evol.* **2**, 983–990 (2018).
48. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
49. Sedio, B. E., Rojas Echeverri, J. C., Boya P, C. A. & Wright, S. J. Sources of variation in foliar secondary chemistry in a tropical forest tree community. *Ecology* **98**, 616–623 (2017).
50. Oksanen, J. et al. vegan: community ecology package. R package v.2.6-4 *CRAN* https://CRAN.R-project.org/package=vegan (2022).
51. Magneville, C. et al. mFD: an R package to compute and illustrate the multiple facets of functional diversity. *Ecography* **2022**, e05904 (2022).
52. Millar, R. B., Anderson, M. J. & Tolimieri, N. Much ado about nothings: using zero similarity points in distance–decay curves. *Ecology* **92**, 1717–1722 (2011).
53. Graco-Roza, C. et al. Distance decay 2.0—a global synthesis of taxonomic and functional turnover in ecological communities. *Glob. Ecol. Biogeogr.* **31**, 1399–1421 (2022).
54. Anderson, M. J. et al. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol. Lett.* **14**, 19–28 (2011).
55. Haan, J. & Avery, M. ciTools: confidence or prediction intervals, quantiles, and probabilities for statistical models. R package v.0.6.1 *CRAN* https://CRAN.R-project.org/package=ciTools (2020).
56. Kurokawa, H. et al. Plant characteristics drive ontogenetic changes in herbivory damage in a temperate forest. *J. Ecol.* **110**, 2772–2784 (2022).
57. Katabuchi, M. LeafArea: an R package for rapid digital image analysis of leaf area. *Ecol. Res.* **30**, 1073–1077 (2015).
58. Laliberté, E. & Legendre, P. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **91**, 299–305 (2010).
59. Carvalho, M. R. et al. Insect leaf-chewing damage tracks herbivore richness in modern and ancient forests. *PLoS ONE* **9**, e94950 (2014).

60. Smith, D. M. & Nufio, C. R. Levels of herbivory in two Costa Rican rain forests: implications for studies of fossil herbivory. *Biotropica* **36**, 318–326 (2004).
61. Azevedo-Schmidt, L., Meineke, E. K. & Currano, E. D. Insect herbivory within modern forests is greater than fossil localities. *Proc. Natl Acad. Sci. USA* **119**, e2202852119 (2022).
62. Wang, X. Z. et al. Phytochemical diversity impacts herbivory in a tropical rainforest tree community. *Ecol. Lett.* **26**, 1898–1910 (2023).
63. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
64. Jin, Y. & Qian, H. V. PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography* **42**, 1353–1359 (2019).
65. 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).
66. Simon, P. B., Theodore Garland, J. R. & Anthony, R. I. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).
67. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
68. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022). https://www.R-project.org/
69. Sun, L. et al. Tree phytochemical diversity and herbivory are higher in the tropics. *figshare* https://doi.org/10.6084/m9.figshare.22758269 (2024).

## Acknowledgements

This research was supported by the NSFC China–US Dimensions of Biodiversity Grant (DEB: no. 32061123003 to M.C.); the National Natural Science Foundation of China (grant nos. 32201318 to L.S. and 31870410 to J.Y.); the Chinese Academy of Sciences Youth Innovation Promotion Association (grant no. Y202080 to J.Y.); the Distinguished Youth Scholar of Yunnan (grant no. 202101AV070005 to J.Y.); the Ten Thousand Talent Plans for Young Top-Notch Talents of Yunnan Province (grant no. YNWR-QNBJ-2018-309 to J.Y.); a Postdoctoral Fellowship of Xishuangbanna Tropical Botanical Garden, CAS, to L.S.; the Postdoctoral Science Foundation of Yunnan Province to L.S.; the 14th Five-Year Plan of the Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (grant nos. XTBG-1450101 and E3ZKFF2B01 to J.Y.); and an NSF US–China Dimensions of Biodiversity Grant (DEB: no. 2124466) to N.G.S. We acknowledge support from Xishuangbanna Station for Tropical Rain Forest Ecosystem Studies, Ailaoshan Station for Subtropical Forest Ecosystem Studies and Lijiang Forest Ecosystem Research Station. We thank the Molecular Biology Experiment Center in Germplasm Bank of Wild Species, Chinese Academy of Sciences, for facilitating the extraction of plant metabolites, and the State Key Laboratory of Phytochemistry and Plant Resources in West China, Chinese Academy of Sciences, for performing the UHPLC–MS/MS analysis. We thank J. Wang, C. Xu, P. Song, T. Liang and many local residents for their assistance in collecting leaf samples. We also thank J. Yang, H. Liu and Y. Tan for their kind assistance during extracting plant metabolites and metabolite analysis.

## Author contributions

J.Y., L.S. and N.G.S. designed the study. M.C. set up the forest inventory plots. L.S., X.Z., Y.H. and X.W. collected and processed the metabolomics data. L.S., Y.H. and X.W. collected and processed the leaf samples. L.S. and J.Y. analysed the data with input from all authors. J.Y., N.G.S. and L.S. wrote the paper. All authors provided feedback on the final version of the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

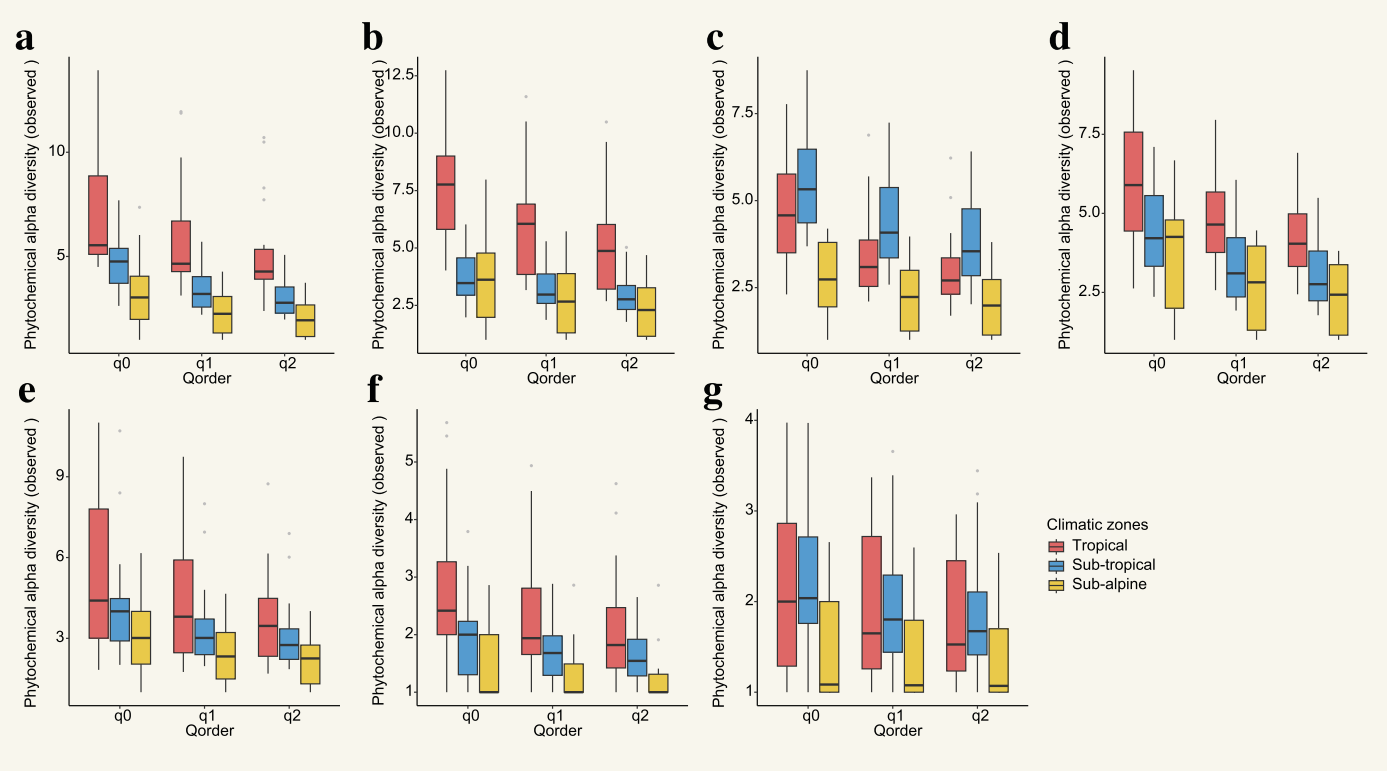
**Extended data** is available for this paper at https://doi.org/10.1038/s41559-024-02444-2.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41559-024-02444-2.

**Correspondence and requests for materials** should be addressed to Jie Yang.

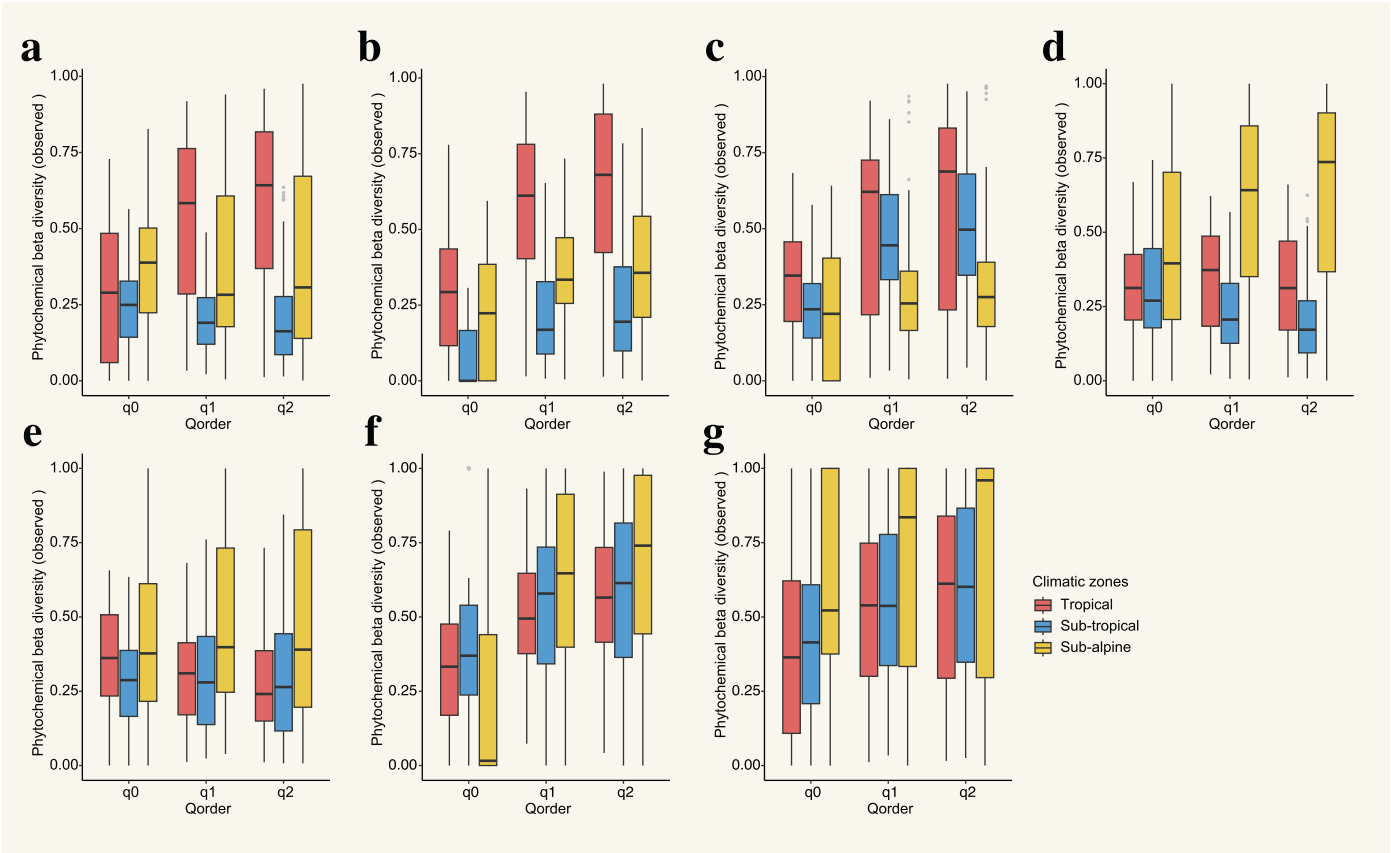
**Peer review information** *Nature Ecology & Evolution* thanks Susan Whitehead and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41559-024-02444-2



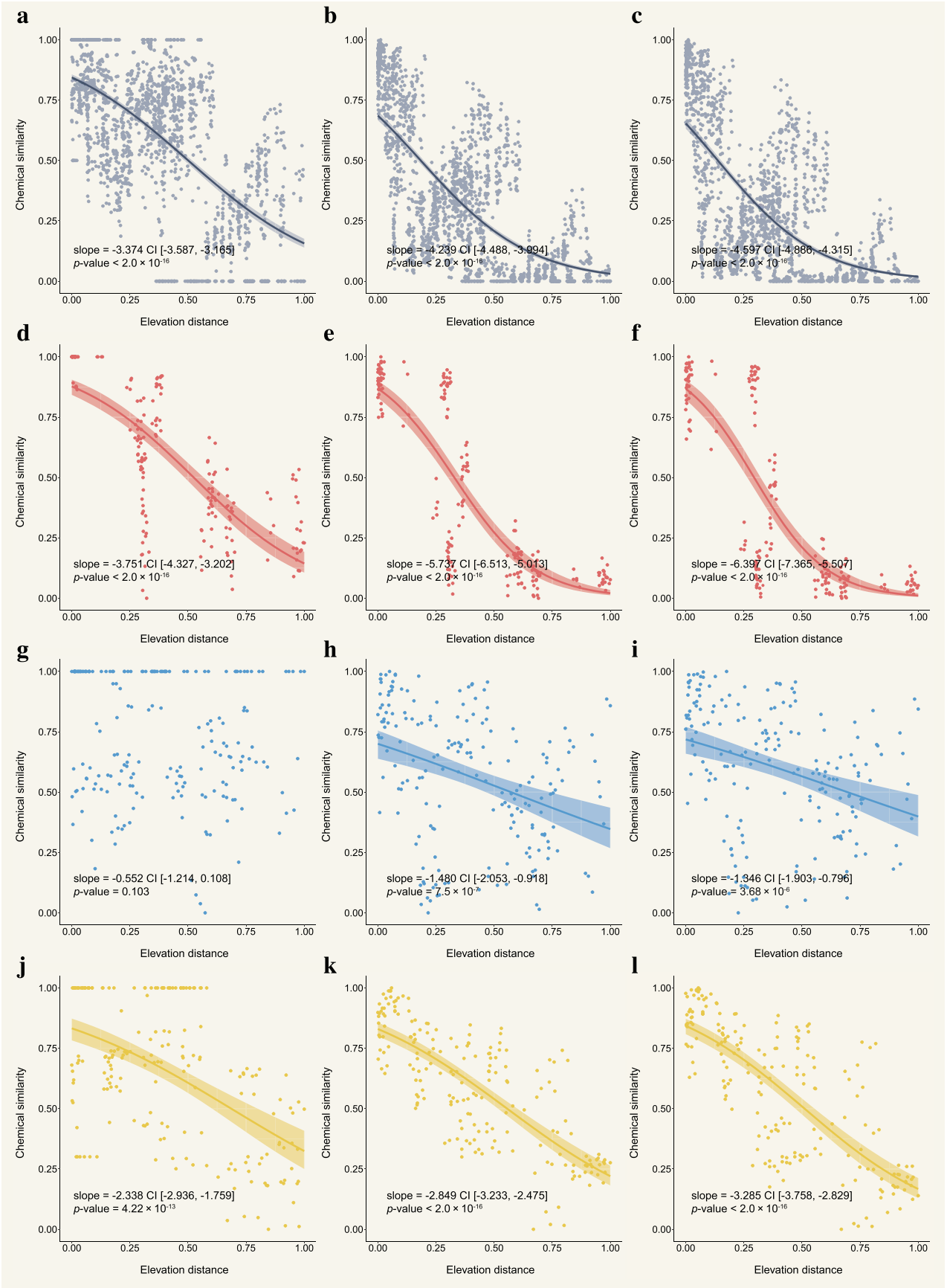

**Extended Data Fig. 1 | Observed phytochemical alpha diversity for seven biosynthetic pathway categories within each climatic zone (tropical, sub-tropical and sub-alpine) with diverse *q* exponents (Qorder = 0, 1, 2).** (**a**) terpenoids (n = 20 in tropical zone, n = 20 in sub-tropical zone, n = 20 in sub-alpine zone, for one of *q* exponents), (**b**) shikimates and phenylpropanoids (n = 20 in tropical zone, n = 20 in sub-tropical zone, n = 20 in sub-alpine zone, for one of *q* exponents), (**c**) polyketides (n = 20 in tropical zone, n = 20 in sub-tropical zone, n = 20 in sub-alpine zone, for one of *q* exponents), (**d**) alkaloids (n = 20 in tropical zone, n = 20 in sub-tropical zone, n = 20 in sub-alpine zone, for one of *q* exponents), (**e**) fatty acids (n = 20 in tropical zone, n = 20 in sub-tropical zone, n = 17 in sub-alpine zone, for one of *q* exponents), (**f**) amino acids/peptides (n = 18 in tropical zone, n = 19 in sub-tropical zone, n = 12 in sub-alpine zone, for one of *q* exponents), (**g**) carbohydrates (n = 17 in tropical zone, n = 20 in sub-tropical zone, n = 17 in sub-alpine zone, for one of *q* exponents). In all panels, the significance of difference of phytochemical alpha diversity across forest type pairs were tested using a one-way ANOVA with a post-hoc Tukey test. In boxplots: the centre line represents the median; the lower and upper hinges correspond to the 25th and 75th percentiles; the lower and upper whiskers extend to the lowest and highest points to a limit of 1.5× the interquartile range from the closest hinge.

https://doi.org/10.1038/s41559-024-02444-2




**Extended Data Fig. 2 | Observed phytochemical beta diversity for seven biosynthetic pathway categories within each climatic zone (tropical, sub-tropical and sub-alpine) with diverse *q* exponents (Qorder = 0, 1, 2).** (**a**) terpenoids (n = 190 in tropical zone, n = 190 in sub-tropical zone, n = 190 in sub-alpine zone, for one of *q* exponents), (**b**) shikimates and phenylpropanoids (n = 190 in tropical zone, n = 190 in sub-tropical zone, n = 190 in sub-alpine zone, for one of *q* exponents), (**c**) polyketides (n = 190 in tropical zone, n = 190 in sub-tropical zone, n = 190 in sub-alpine zone, for one of *q* exponents), (**d**) alkaloids (n = 190 in tropical zone, n = 190 in sub-tropical zone, n = 190 in sub-alpine zone, for one of *q* exponents), (**e**) fatty acids (n = 190 in tropical zone, n = 190 in sub-tropical zone, n = 136 in sub-alpine zone, for one of *q* exponents), (**f**) amino acids/peptides (n = 153 in tropical zone, n = 171 in sub-tropical zone, n = 66 in sub-alpine zone, for one of *q* exponents), (**g**) carbohydrates (n = 136 in tropical zone, n = 190 in sub-tropical zone, n = 136 in sub-alpine zone, for one of *q* exponents). In all panels, the significance of difference of phytochemical beta diversity across forest type pairs were tested using a one-way ANOVA with a post-hoc Tukey test. In boxplots: the centre line represents the median; the lower and upper hinges correspond to the 25th and 75th percentiles; the lower and upper whiskers extend to the lowest and highest points to a limit of 1.5× the interquartile range from the closest hinge.

https://doi.org/10.1038/s41559-024-02444-2

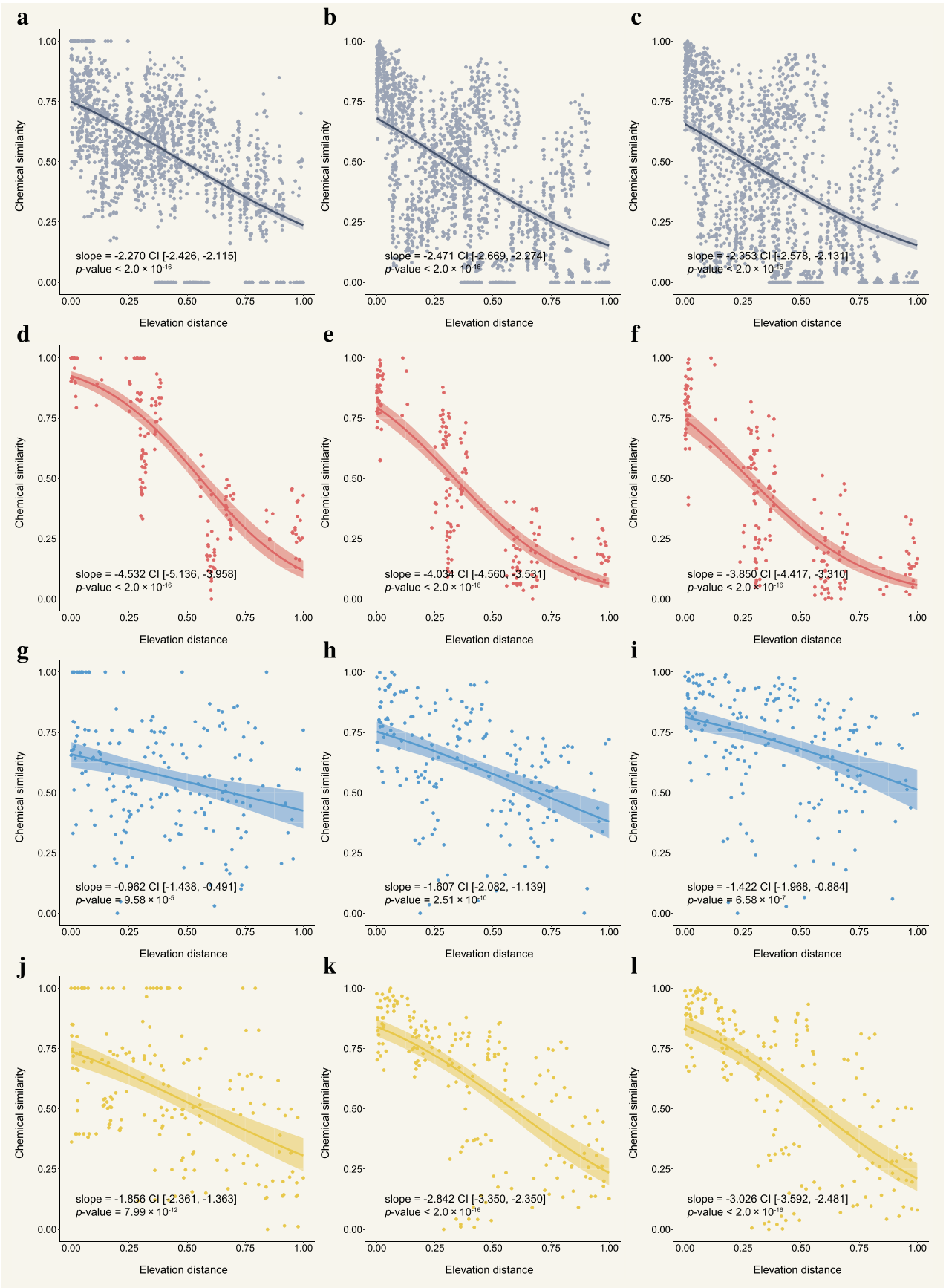



**Extended Data Fig. 3 | See next page for caption.**

https://doi.org/10.1038/s41559-024-02444-2

**Extended Data Fig. 3 | Distance-decay curves for the whole plant specialized metabolites with diverse *q* exponents of 0, 1, 2.** The rate of decay (slope) and corresponding significance level were estimated by regressing the chemical similarity against elevational distance via generalized linear model with link log and a quasi-binomial family. The trend lines represent linear fits from regressions, and coloured shaded areas indicate 95% confidence interval (CI) of the prediction. Colours denote whole study region (grey), tropical zone (red), sub-tropical zone (blue) and sub-alpine zone (yellow). Panels, **a, d, g** and **j** show the slope of the relationship when *q* exponents is 0. Panels, **b, e, h** and **k**. show the slope of the relationship when *q* exponents is 1. Panels, **c, f, i** and **l** show the slope of the relationship when *q* exponents is 2.

https://doi.org/10.1038/s41559-024-02444-2

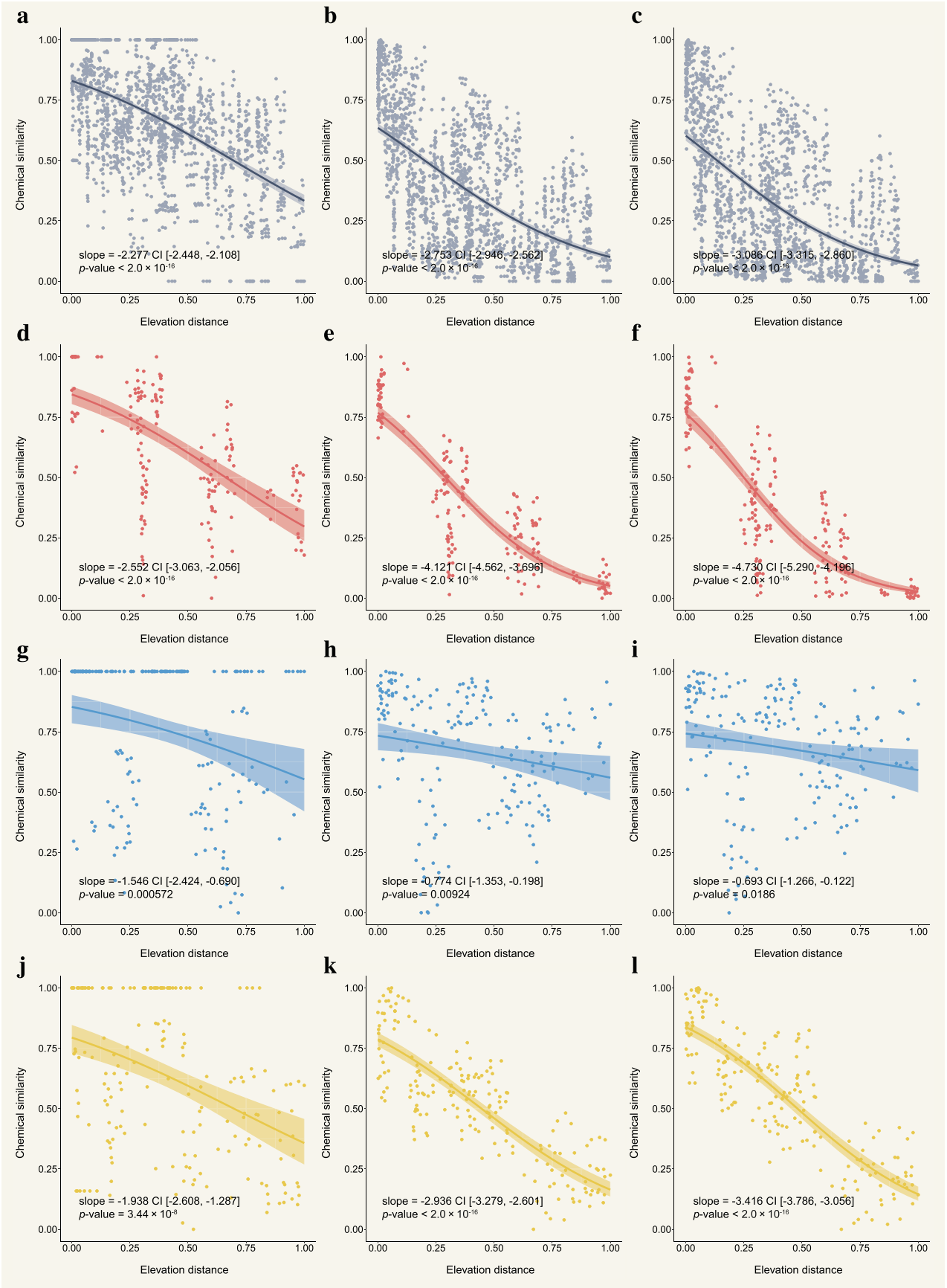



**Extended Data Fig. 4 | See next page for caption.**

https://doi.org/10.1038/s41559-024-02444-2

**Extended Data Fig. 4 | Distance-decay curves for the plant specialized metabolites on terpenoids with diverse *q* exponents of 0, 1, 2.** The rate of decay (slope) and corresponding significance level were estimated by regressing the chemical similarity against elevational distance via generalized linear model with link log and a quasi-binomial family. The trend lines represent linear fits from regressions, and coloured shaded areas indicate 95% confidence interval (CI) of the prediction. Colours denote whole study region (grey), tropical zone (red), sub-tropical zone (blue) and sub-alpine zone (yellow). Panels, **a, d, g** and **j** show the slope of the relationship when *q* exponents is 0. Panels, **b, e, h** and **k**. show the slope of the relationship when *q* exponents is 1. Panels, **c, f, i** and **l** show the slope of the relationship when *q* exponents is 2.

https://doi.org/10.1038/s41559-024-02444-2

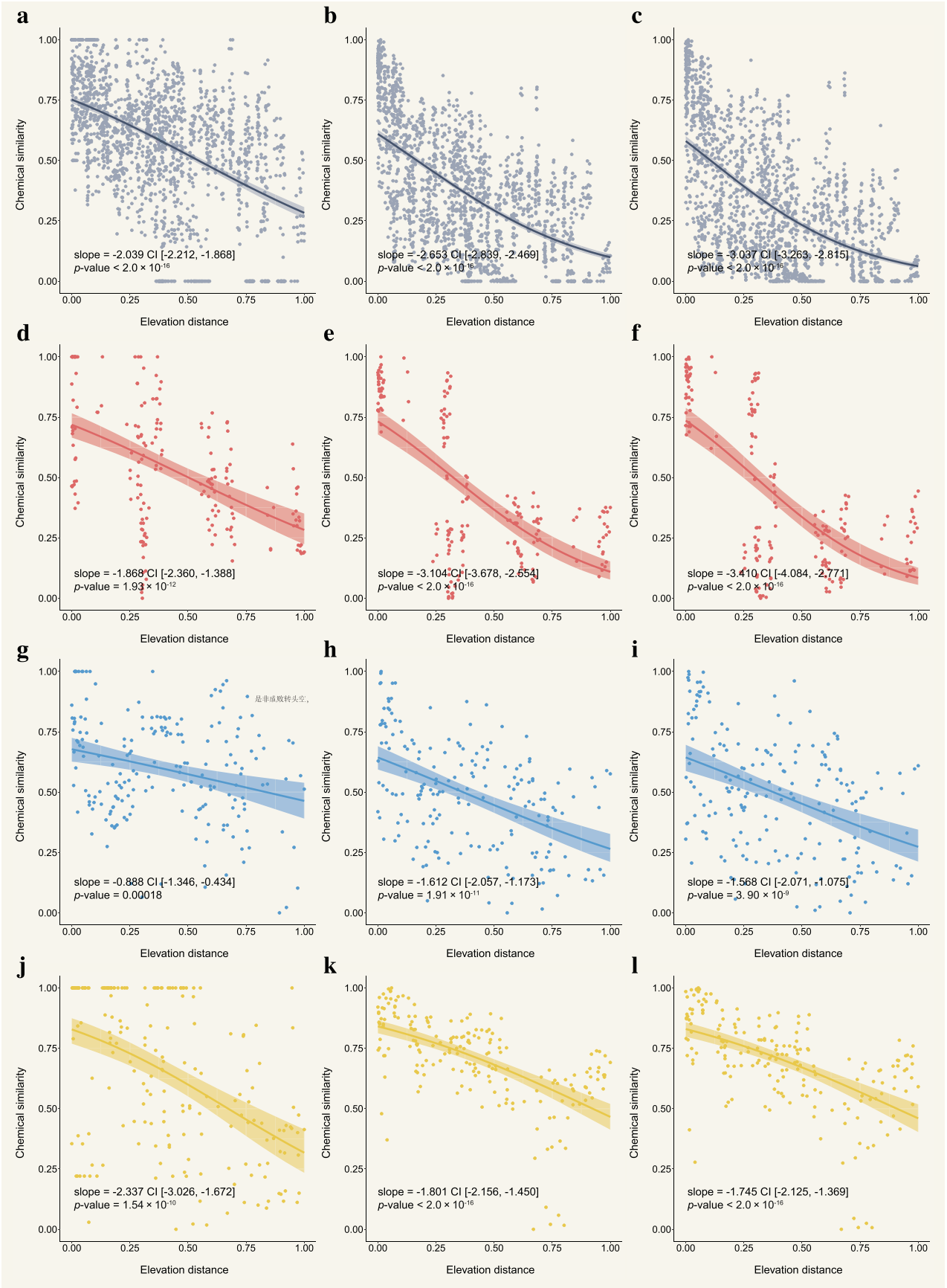



**Extended Data Fig. 5 | See next page for caption.**

https://doi.org/10.1038/s41559-024-02444-2

**Extended Data Fig. 5 | Distance-decay curves for the plant specialized metabolites on shikimates and phenylpropanoids with diverse *q* exponents of 0, 1, 2.** The rate of decay (slope) and corresponding significance level were estimated by regressing the chemical similarity against elevational distance via generalized linear model with link log and a quasi-binomial family. The trend lines represent linear fits from regressions, and coloured shaded areas indicate 95% confidence interval (CI) of the prediction. Colours denote whole study region (grey), tropical zone (red), sub-tropical zone (blue) and sub-alpine zone (yellow). Panels, **a, d, g** and **j** show the slope of the relationship when *q* exponents is 0. Panels, **b, e, h** and **k**. show the slope of the relationship when *q* exponents is 1. Panels, **c, f, i** and **l** show the slope of the relationship when *q* exponents is 2.

https://doi.org/10.1038/s41559-024-02444-2

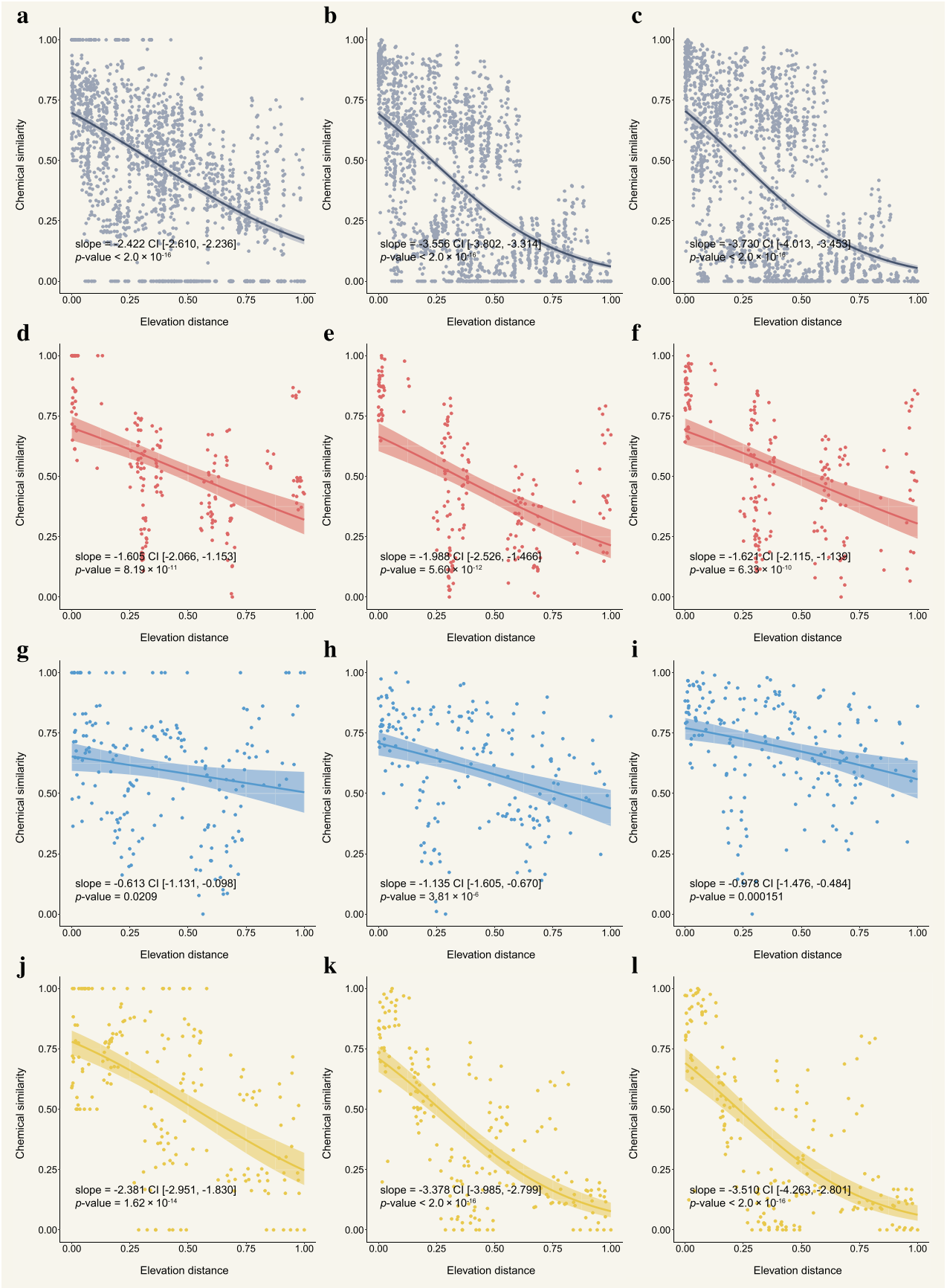



**Extended Data Fig. 6 | See next page for caption.**

https://doi.org/10.1038/s41559-024-02444-2

**Extended Data Fig. 6 | Distance-decay curves for the plant specialized metabolites on polyketides with diverse *q* exponents of 0, 1, 2.** The rate of decay (slope) and corresponding significance level were estimated by regressing the chemical similarity against elevational distance via generalized linear model with link log and a quasi-binomial family. The trend lines represent linear fits from regressions, and coloured shaded areas indicate 95% confidence interval (CI) of the prediction. Colours denote whole study region (grey), tropical zone (red), sub-tropical zone (blue), sub-alpine zone (yellow). Panels, **a, d, g** and **j** show the slope of the relationship when *q* exponents is 0. Panels, **b, e, h** and **k**. show the slope of the relationship when *q* exponents is 1. Panels, **c, f, i** and **l** show the slope of the relationship when *q* exponents is 2.

https://doi.org/10.1038/s41559-024-02444-2

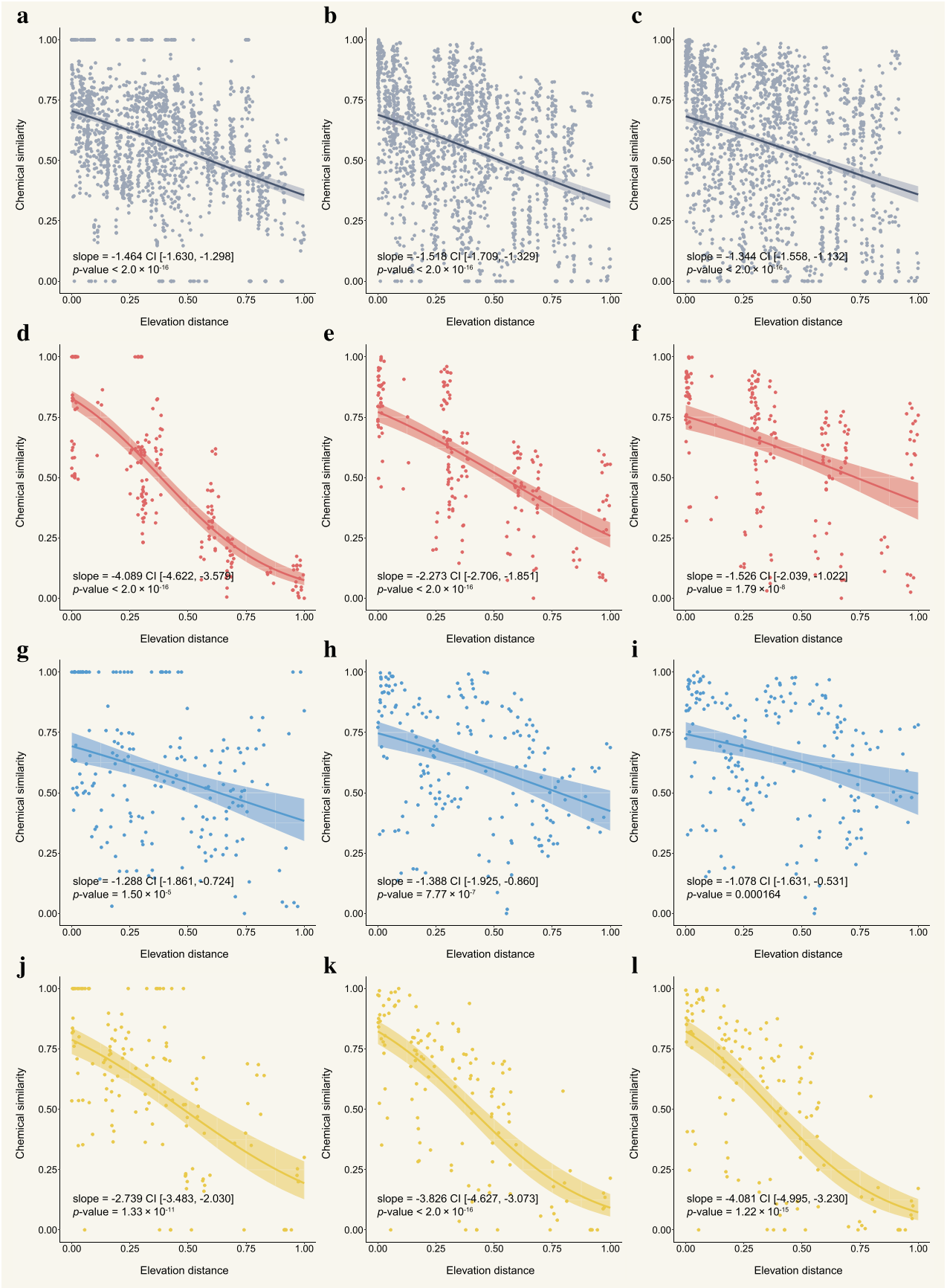



**Extended Data Fig. 7 | See next page for caption.**

https://doi.org/10.1038/s41559-024-02444-2

**Extended Data Fig. 7 | Distance-decay curves for the plant specialized metabolites on alkaloids with diverse *q* exponents of 0, 1, 2.** The rate of decay (slope) and corresponding significance level were estimated by regressing the chemical similarity against elevational distance via generalized linear model with link log and a quasi-binomial family. The trend lines represent linear fits from regressions, and coloured shaded areas indicate 95% confidence interval (CI) of the prediction. Colours denote whole study region (grey), tropical zone (red), sub-tropical zone (blue) and sub-alpine zone (yellow). Panels, **a, d, g** and **j** show the slope of the relationship when *q* exponents is 0. Panels, **b, e, h** and **k**. show the slope of the relationship when *q* exponents is 1. Panels, **c, f, i** and **l** show the slope of the relationship when *q* exponents is 2.

https://doi.org/10.1038/s41559-024-02444-2

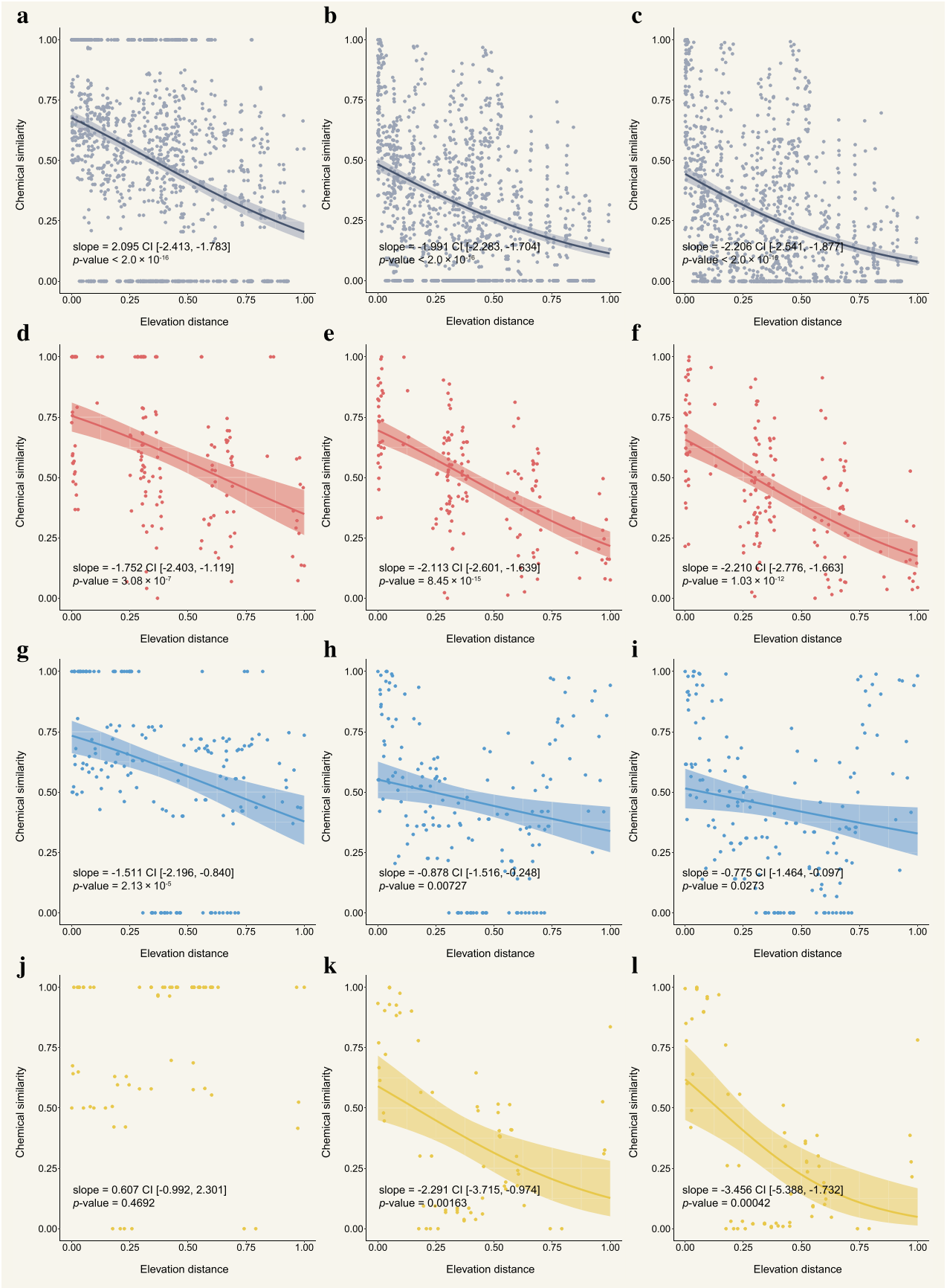



**Extended Data Fig. 8 | See next page for caption.**

https://doi.org/10.1038/s41559-024-02444-2

**Extended Data Fig. 8 | Distance-decay curves for the plant specialized metabolites on fatty acids with diverse *q* exponents of 0, 1, 2.** The rate of decay (slope) and corresponding significance level were estimated by regressing the chemical similarity against elevational distance via generalized linear model with link log and a quasi-binomial family. The trend lines represent linear fits from regressions, and coloured shaded areas indicate 95% confidence interval (CI) of the prediction. Colours denote whole study region (grey), tropical zone (red), sub-tropical zone (blue) and sub-alpine zone (yellow). Panels, **a, d, g** and **j** show the slope of the relationship when *q* exponents is 0. Panels, **b, e, h** and **k**. show the slope of the relationship when *q* exponents is 1. Panels, **c, f, i** and **l** show the slope of the relationship when *q* exponents is 2.

https://doi.org/10.1038/s41559-024-02444-2

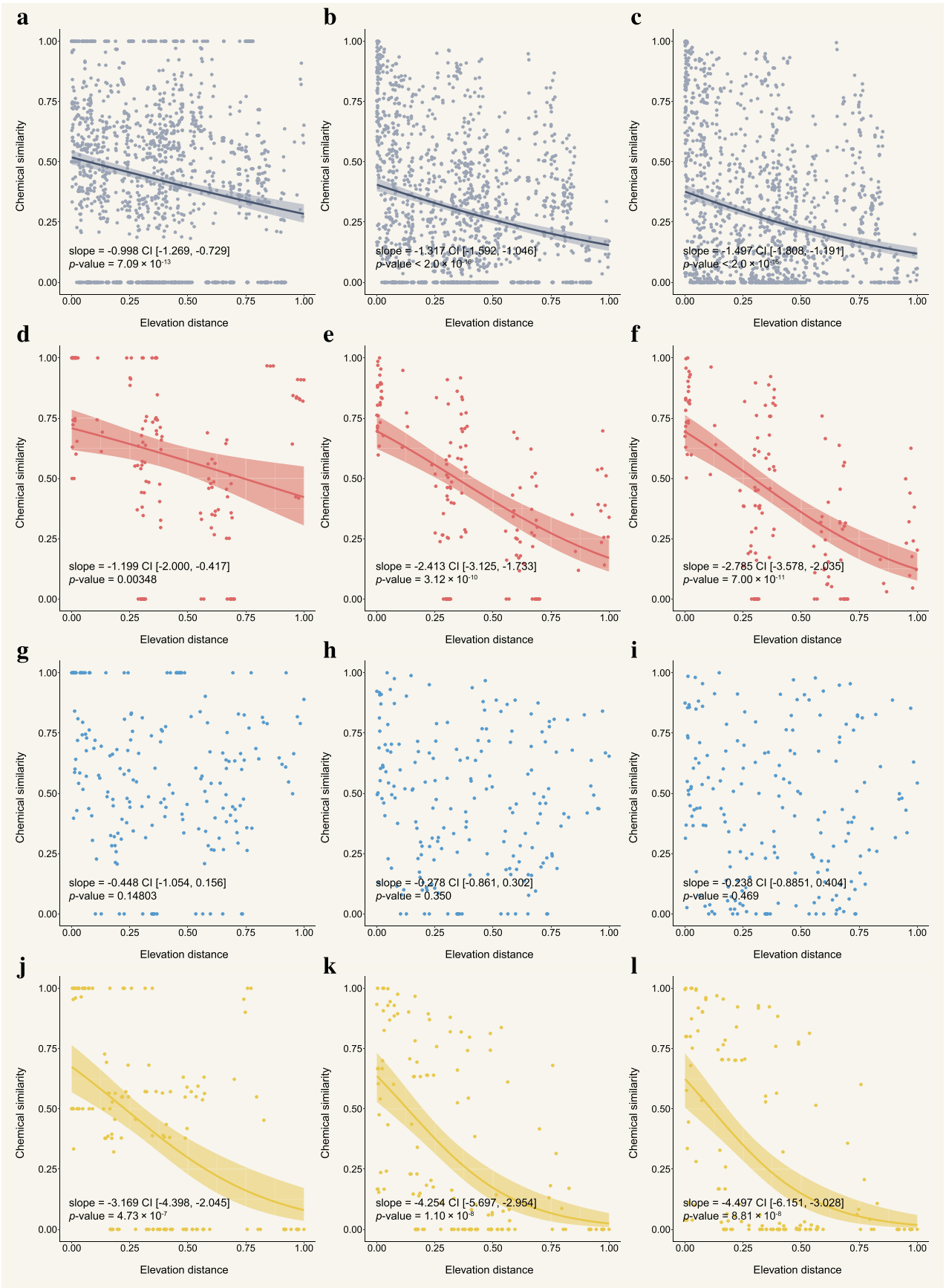



**Extended Data Fig. 9 | See next page for caption.**

https://doi.org/10.1038/s41559-024-02444-2

**Extended Data Fig. 9 | Distance-decay curves for the plant specialized metabolites on amino acids/ peptides with diverse q exponents of 0, 1, 2.** The rate of decay (slope) and corresponding significance level were estimated by regressing the chemical similarity against elevational distance via generalized linear model with link log and a quasi-binomial family. The trend lines represent linear fits from regressions, and coloured shaded areas indicate 95% confidence interval (CI) of the prediction. Colours denote whole study region (grey), tropical zone (red), sub-tropical zone (blue) and sub-alpine zone (yellow). Panels, **a, d, g** and **j** show the slope of the relationship when q exponents is 0. Panels, **b, e, h** and **k**. show the slope of the relationship when q exponents is 1. Panels, **c, f, i** and **l** show the slope of the relationship when q exponents is 2.

https://doi.org/10.1038/s41559-024-02444-2




**Extended Data Fig. 10 | See next page for caption.**

https://doi.org/10.1038/s41559-024-02444-2

**Extended Data Fig. 10 | Distance-decay curves for the plant specialized metabolites on carbohydrates with diverse *q* exponents of 0, 1, 2.** The rate of decay (slope) and corresponding significance level were estimated by regressing the chemical similarity against elevational distance via generalized linear model with link log and a quasi-binomial family. The trend lines represent linear fits from regressions, and coloured shaded areas indicate 95% confidence interval (CI) of the prediction. Colours denote whole study region (grey), tropical zone (red), sub-tropical zone (blue) and sub-alpine zone (yellow). Panels, **a, d, g** and **j** show the slope of the relationship when *q* exponents is 0. Panels, **b, e, h** and **k**. show the slope of the relationship when *q* exponents is 1. Panels, **c, f, i** and **l** show the slope of the relationship when *q* exponents is 2.

nature portfolio

Corresponding author(s): Jie Yang

Last updated by author(s): Apr 24, 2024

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Code to support our analyses are available here: http://dx.doi/org/10.6084/m9.figshare.22758269. |
| Data analysis | All statistical analyses were performed in R version 4.2.2. All packages used are decribed in the methods and supplementary materials. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The mass spectrometry data (.mzML) were deposited on the MassIVE public repository and are available under the dataset accession number MSV000092950. The datasets used in the current study are available from Figshare repository: http://dx.doi/org/10.6084/m9.figshare.22758269.The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files. Should any raw data files be needed in another format they are available from the corresponding author upon reasonable request. Source data are provided with this paper.

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | *Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.*<br>*Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.*<br>*Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.* |
| Reporting on race, ethnicity, or other socially relevant groupings | *Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).*<br>*Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)*<br>*Please provide details about how you controlled for confounding variables in your analyses.* |
| Population characteristics | *Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."* |
| Recruitment | *Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.* |
| Ethics oversight | *Identify the organization(s) that approved the study protocol.* |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | This study used a community metabolomics approach to characterize the specialized metabolites of tree communities in the 60 forest inventory plots across a gradient of tropical rain forest to sub-tropical forest to sub-alpine forest. |
| Research sample | we collected leaves from 979 individuals belonging to 206 species, 122 genera and 56 families representing 329 unique species-by-elevation combinations. |
| Sampling strategy | We collected the nearly fully emerged, unlignified and undamaged leaves of trees in each forest plots during the growing season in 2021. In each forest plot, we randomly sampled three individuals of each common species and collected all individuals of species with less than 3 individuals in the plot. For each individual, 2 to 10 of the expanded leaves were sampled, labelled, wrapped in foil, and immediately immersed in liquid nitrogen in the field, and then stored at -80 degree centigrade in a laboratory freezer. In total, we collected leaf samples from 979 individuals of 206 species, belonging to 122 genera and 56 families. |
| Data collection | For each individual, 2 to 10 of the expanded leaves were sampled, labelled, wrapped in foil, and immediately immersed in liquid nitrogen in the field, and then stored at -80 degree centigrade in a laboratory freezer. Leaf material from each species in a single elevation belt were pooled for downstream metabolomic analyses. Although the pooling of leaf material ignores individual-level variation, our primary goal was to characterize species-level phytochemistry and recent studies have indicated that intra-specific variation in a local scale sample is far less than inter-specific variation. After pooling samples, across the elevational gradient, 329 unique species-elevation combinations were included in the analysis. |
| Timing and spatial scale | We collected the leaves of trees in each forest plots from July to August in 2021. The forest plots across at four elevational belts were established in each climatic region. Specifically, in the tropical region, four elevational belts (800 m, 1,000 m, 1,200 m and 1,400 m), In the sub-tropical region, four elevational belts (2,000 m, 2,200 m, 2,400 m and 2,600 m) and in the sub-alpine region, four elevational belts (3,200 m, 3,400 m, 3,600 m and 3,800 m). In each elevational belt, five 20 m × 20 m replicate plots were established approximately 200 meters apart from each other. |
| Data exclusions | N/A |

| | |
|---|---|
| Reproducibility | Our study was conducted at the permanent plot. Within each plot, all trees with diameter at breast height (DBH) > 5 cm were measured, tagged and identified. All our sample was saved at the center lab in XTBG, CAS. |
| Randomization | In each forest plot, we randomly sampled three individuals of each common species and collected all individuals of species with less than 3 individuals in the plot. For each individual, 2 to 10 of the expanded leaves were sampled. To assess whether the observed phytochemical alpha and beta diversities were higher, no different or lower than expectd given the observed phytochemical diversity within a forest type (i.e. phytochemical gamma diversity), we conducted null model analyses. The null model utilized randomized species names on the chemical dissimilarity matrix within a forest type. This was repeated 999 times and during each interation null phytochemical alpha and beta diversity values were calculated for each of the q exponent values. |
| Blinding | Blinding is not possible for our study because we collected data from field work. |

Did the study involve field work? ☒ Yes ☐ No

## Field work, collection and transport

| | |
|---|---|
| Field conditions | This study was conducted along an elevational gradient that encompasses three climatic regions: tropical, sub-tropical and sub-alpine, in Yunnan Province, a global biodiversity hotspot in southwestern China. During 2011 and 2012, four elevational belts were established in each climatic region. Specifically, in the tropical region, four elevational belts (800 m, 1,000 m, 1,200 m and 1,400 m) were established in Xishuangbanna National Nature Reserve, and are characterized by vegetation that includes tropical seasonal rain forest, tropical montane rain forest, tropical montane evergreen broad-leaved forest. In the sub-tropical region, four elevational belts (2,000 m, 2,200 m, 2,400 m and 2,600 m) were established in Ailaoshan National Nature Reserve, and are characterized with sub-tropical evergreen broad-leaved forest. In the sub-alpine region, four elevational belts (3,200 m, 3,400 m, 3,600 m and 3,800 m) were established in Yulong Snow Mountain Nature Reserve, and are characterized with vegetation of sub-alpine coniferous forest. In each elevational belt, five 20 m × 20 m replicate plots were established approximately 200 meters apart from each other. Locations impacted by anthropogenic and natural disturbances, such as large canopy gaps, were avoided. Within each plot, all trees with diameter at breast height > 5 cm were measured, tagged and identified in previous studies. |
| Location | Xishuangbanna: 21°37'N, 101°34'E; Ailao Mountain: 101°01’E, 24°32’N; and Yulong snow Mountain: 100°13’E, 27°08’N. |
| Access & import/export | We established this four elevation belts in Xishuangbanna, Ailao mountain and Yulong snow mountain. Our institute have sign the MOU with each national nature reserve to get the permission to do the research and collect the plant, soil sample. |
| Disturbance | N/A |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☐ | ☒ Plants |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |