Leaf metabolic traits reveal hidden dimensions of plant form and function

Tom W. N. Walker[1,2]*, Franziska Schrodt[3], Pierre-Marie Allard[4,5], Emmanuel Defossez[2], Vincent E. J. Jassey[6], Meredith C. Schuman[23], Jake M. Alexander[1], Oliver Baines[3], Virginie Baldy[7], Richard D. Bardgett[8], Pol Capdevila[9,10], Phyllis D. Coley[12], Nicole M. van Dam[13,14,15], Bruno David[16], Patrice Descombes[1,26,27], Maria-Jose Endara[12], Catherine Fernandez[7], Dale Forrister[12], Albert Gargallo-Garriga[17,18,19], Gaëtan Gauser[2], Sue Marr[13,20,21], Steffen Neumann[13,20], Loïc Pellissier[1,26], Kristian Peters[13,20,21], Sergio Rasmann[2], Ute Roessner[22], Roberto Salguero-Gómez[11], Jordi Sardans[17,18], Wolfram Weckwerth[24,25], Jean-Luc Wolfender[4,5] & Josep Peñuelas[17,18]

[1] Department of Environmental Systems Science, ETH Zürich, 8092 Zürich, Switzerland
[2] Institute of Biology, University of Neuchâtel, 2000 Neuchâtel, Switzerland.
[3] School of Geography, University of Nottingham, Nottingham NG7 2RD, UK
[4] School of Pharmaceutical Sciences, University of Geneva, 1211 Geneva, Switzerland
[5] Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, 1211 Geneva, Switzerland
[6] Laboratoire d'Ecologie Fonctionnelle et Environnement, Université de Toulouse, CNRS, 31062 Toulouse, France
[7] Aix Marseille Université, Avignon Université, CNRS, IRD, IMBE, Marseille, France
[8] Department of Earth & Environmental Sciences, The University of Manchester, Manchester M13 9PT, UK
[9] Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Universitat de Barcelona, 08028 Barcelona, Spain
[10] School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, UK
[11] Department of Zoology, Oxford University, Oxford OX1 3SZ, UK
[12] School of Biological Sciences, University of Utah, Salt Lake City, UT 84112, USA
[13] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany
[14] Institute of Biodiversity, Friedrich Schiller University Jena, 07743 Jena, Germany
[15] Leibniz Institute of Vegetable and Ornamental crops (IGZ), 14979 Großbeeren, Germany
[16] Green Mission Pierre Fabre, Institut de Recherche Pierre Fabre, 31562 Toulouse, France
[17] CSIC, Global Ecology Unit CREAF-CSIC-UAB, 08193 Bellaterra, Catalonia, Spain
[18] CREAF, 08193 Cerdanyola del Vallès, Catalonia, Spain
[19] Global Change Research Institute, Czech Academy of Sciences, 603 00 Brno, Czech Republic
[20] Leibniz Institute of Plant Biochemistry, Bioinformatics and Scientific Data, 06120 Halle, Germany
[21] Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle Wittenberg, 06108 Halle, Germany
[22] Research School of Biology, The Australian National University, 2601 Acton, Australia
[23] Departments of Geography and Chemistry, University of Zurich, 8057 Zurich, Switzerland
[24] Molecular Systems Biology, Department of Functional and Evolutionary Ecology, 1010 University of Vienna, Austria
[25] Vienna Metabolomics Center, 1010 University of Vienna, Austria
[26] Ecosystems & Landscape Evolution, Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), 8903 Birmensdorf, Switzerland
[27] Musée et Jardins botaniques cantonaux, 1007 Lausanne, Switzerland
* Correspondence to: thomas.walker@unine.ch

Summary

The plant metabolome encompasses the biochemical mechanisms through which evolutionary and ecological processes shape plant form and function[1,2]. However, while the metabolome should thus be an important component of plant life-history variation[3], we know little about how it varies across the plant kingdom. Here, we use the plant functional trait concept[4] – a powerful framework for describing plant form and function[5–7] – to interpret leaf metabolome variation among 457 tropical and 339 temperate plant species. Distilling metabolite chemistry into five discriminant metabolic functional traits reveals that plants vary along two major axes of leaf metabolic specialization – a leaf chemical defense spectrum and an expression of leaf longevity. These axes are qualitatively consistent for tropical and temperate species, with many trait combinations being viable. However, axes of leaf metabolic specialization vary orthogonally to life-history strategies described by widely used functional traits[5–7], while being at least equally important to them. Our findings question classical trait[6] and plant defense[8] theory that predicts relationships between the leaf chemical phenotype, plant productivity, and pace of life. Moreover, we show that metabolic functional traits describe unique dimensions of plant life-history variation that are complementary to, and independent from, those captured by existing plant functional traits.

Main Text

Plants produce a staggering diversity of metabolites – upwards of one million throughout the plant kingdom and many thousands within an individual[9] (the metabolome). The plant metabolome is well known to human society, being the source of most cosmetics[10], flavors[11], and medicines[12]. Yet, little is known about how the metabolome varies across the plant kingdom. The metabolome is the biochemical basis of physiology, comprising primary metabolites involved in cellular function (e.g., carbohydrates, amino acids) and specialized metabolites produced to confront stress (e.g., flavonoids, terpenoids)[9,13,14]. The metabolome thus encompasses the biochemical mechanisms through which evolutionary and ecological processes shape plant form and function (i.e., morphology and physiology [7])[1,2]. Plant form and function vary predictably among species because evolutionary and ecological processes generally act to maximize fitness[8,15–18], thus limiting the number of morphological or physiological strategies that are viable[5–7]. It follows that the metabolome should also be constrained to a limited number of strategies that maximize fitness[3]. However, although cross-species studies are now emerging[19–22], no attempt has been made to characterize how the metabolome varies systematically throughout the plant kingdom. We thus remain unable to place the vast diversity of the plant metabolome into a coherent ecological context.

The plant functional trait concept offers a powerful framework to contextualize the plant metabolome. Plant functional traits are a standardized set of morphological and physiological characteristics that provide universal metrics of plant form and function[4]. Plant functional traits have been pivotal in identifying the life-history strategies that govern plant fitness[5–7], understanding plant impacts on population[23], community[24], and ecosystem[25] processes, and predicting global change effects on ecosystems[26]. Plant functional traits thus provide a strong reference point from which to interpret variation in the plant metabolome. However, measurements of the metabolome remain absent from the functional trait concept, in part because they yield complex data describing thousands of metabolites from hundreds of biochemical pathways[3]. Advances in chemoinformatics make it now possible to translate metabolite identities into a targeted set of chemical properties[27,28]. Doing so resonates with the functional trait concept – i.e., that quantitative characteristics provide a bridge between identity and function[4] – and is an intuitive way to contextualize the metabolome against plant functional traits. Yet, while chemoinformatics is commonplace in natural products chemistry[29–31], it remains entirely absent from ecology.

Contextualizing the plant metabolome against existing plant functional traits raises two alternative hypotheses. If the metabolome varies colinearly with functional traits, then it offers a biochemical validation of the functional trait concept. Such validation is necessary because functional traits are emergent properties of multiple biochemical processes[4,32], making them proxies for physiology that can be poor predictors of ecological processes[33–35]. Alternatively, if the metabolome varies orthogonally to functional traits, then it describes dimensions of plant life-history variation missed by existing functional traits. Such orthogonal variation would enhance the explanatory power of the functional trait concept by providing additional axes of trait specialization to examine plant form and function. Either way, integrating measurements of the plant metabolome into the plant functional trait concept will yield insight into how the metabolome varies across the plant kingdom, while also offering a high-resolution mechanistic lens with which to advance the functional trait concept itself.

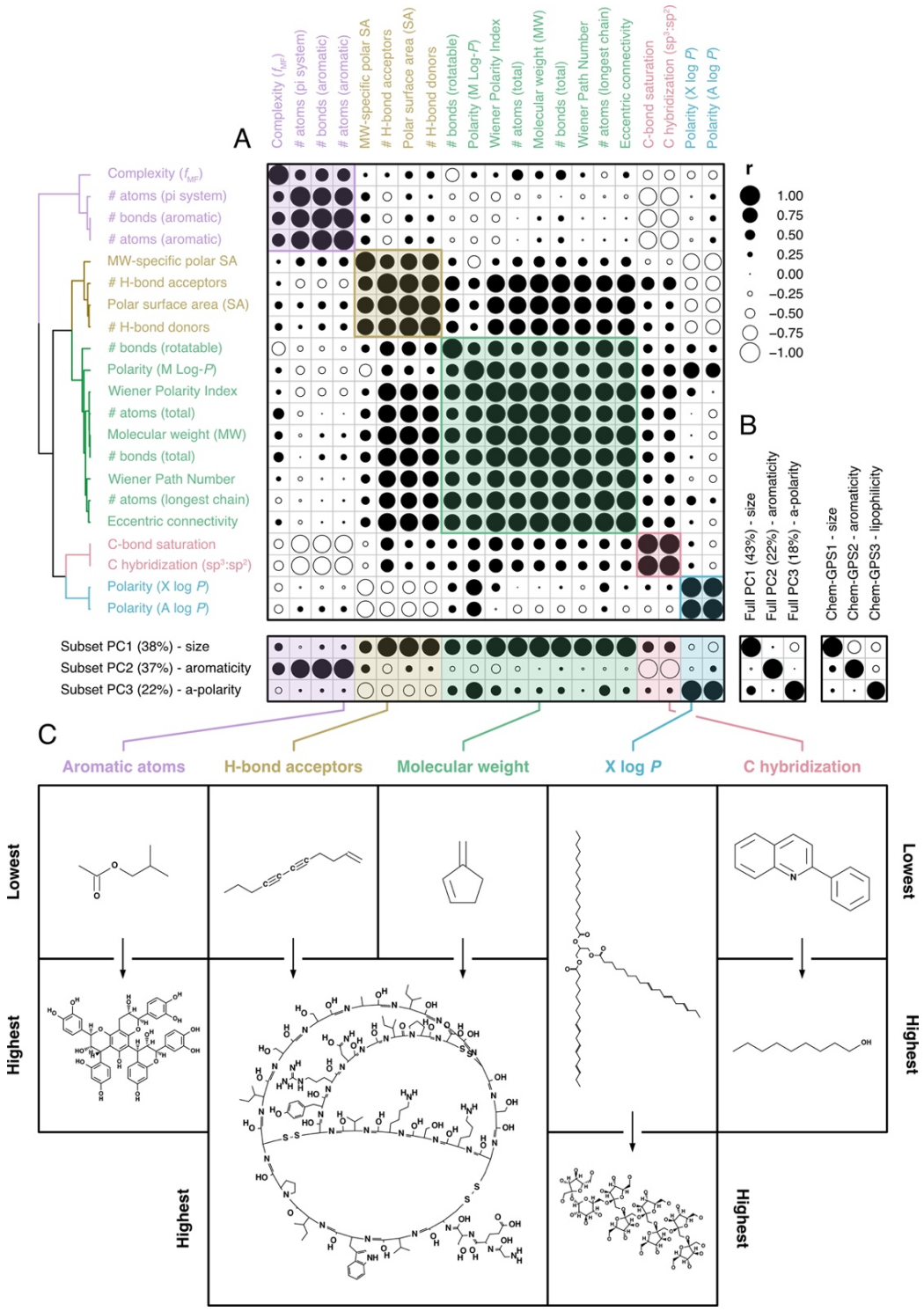


Figure 1 | Variation in leaf metabolite chemistry can be captured by five chemical properties. (A) Coefficients for Pearson correlations among 21 chemical properties (Table S1) of annotated metabolites detected in tropical and temperate datasets (N = 4292). Properties are ordered based on correlation similarity (hierarchical clustering; Ward), with colors separating distinct clusters of properties (maximal silhouette width: K = 5). The lower panel shows correlations between chemical properties and scores from the first three axes of a PCA performed on a subset of five properties, as well as (B) between subset PC scores and either scores from a PCA performed on all 21 properties (left panel) or metabolite positions on the first three axes of ChemGPS-NP[31] space (right panel). Circle sizes indicate coefficient strengths, while fills discriminate between positive (black) and negative (white) correlations. (C) The chemical properties selected to represent each cluster, alongside chemical graphs of metabolites with the lowest and highest values for each trait.

116 Here, we performed an assessment of the leaf metabolomes of 457 tropical and 339 temperate plant
117 species to characterize variation in the metabolome and test whether it validates, or expands upon,
118 existing plant functional traits. First, we examined the chemical properties of all identified metabolites
119 (N = 4292) to derive a minimum set of "metabolic functional traits" that capture chemical variation in
120 the leaf metabolome. We then characterized how metabolic traits vary among species to identify major
121 axes of leaf metabolic specialization in the plant kingdom. Finally, we combined metabolic traits with
122 eight widely used functional traits (plant height, seed mass, stem specific density, leaf area, specific
123 leaf area and leaf carbon, nitrogen and phosphorus concentrations; hereafter "classical traits")[7] to
124 determine whether axes of metabolic specialization are colinear with, or orthogonal to, plant life-history
125 strategies described by classical functional traits.
126
127 Five chemical properties capture chemical variation in the leaf metabolome
128 We expressed the leaf metabolomes of tropical[36] and temperate[37] plant species as the presence or
129 absence of annotated metabolic features (i.e., metabolites) detected through liquid chromatography
130 mass spectrometry (tropical N = 3356; temperate N = 2227). We then characterized the chemistry of
131 the leaf metabolome by examining variation in the chemical properties of all unique metabolites from
132 both datasets (N = 4292)[31]. We focused on 21 properties describing metabolite constitution, geometry,
133 or topology that have relationships with (bio)chemical activity (Table S1). Pairwise correlations among
134 chemical properties (Fig. 1A, upper panel) reveal that properties separate into five clusters, each
135 representing a distinct facet of leaf metabolite chemistry. The largest cluster (Fig. 1A, green) describes
136 overall size and structural complexity, comprising total numbers of atoms and bonds, molecular
137 weight, a predictor of polarity based on the proportion of non-carbon atoms (M log *P*)[38], and related
138 topological parameters (i.e., Wiener numbers[38], Eccentric Connectivity[39]). A second cluster (Fig. 1A,
139 purple) describes bond conjugation and reactivity, comprising numbers of aromatic atoms and bonds,
140 numbers of atoms in the largest conjugated (pi-) system, and a derived complexity index indicative of
141 promiscuous activity in drugs (f_{MF})[40]. A third cluster (Fig. 1A, yellow) describes polar intermolecular
142 (non-covalent) forces, comprising total and mass-specific polar surface area and numbers of hydrogen
143 (H) bond donors or acceptors[41,42]. Such forces are important for the selective binding of small
144 molecules to DNA[43], RNA[44], and proteins[45]. A fourth cluster (Fig. 1A, blue) comprises two predictors of
145 metabolite polarity based on summed properties of atoms and bonds (X log *P*[46], A log *P*[47]), where
146 lower values indicate higher polarity. A final cluster (Fig. 1A, red) describes carbon bond saturation *via*
147 the relative occurrence of sp^3 (unsaturated, three-dimensional) and sp^2 (saturated, planar)
148 hybridization[48]. Overall, these clusters indicate that leaf metabolite chemistry is highly structured and
149 can be described by five distinct facets: size/complexity, bond conjugation/reactivity, contributions to
150 polar intermolecular forces, general polarity, and carbon bond saturation.
151
152 The strong collinearity among chemical properties (Fig. 1A, upper panel) suggests that the five facets
153 of leaf chemistry can be represented by a small number of discriminant variables. Indeed, a PCA
154 performed on a subset of five chemical properties, one from each cluster (Fig. 1C), is equivalent to a
155 PCA performed on all 21 properties (Fig. 1B, left panel; PC1: $r_{4290} = 0.97$; PC2: $r_{4290} = 0.99$; PC3: r_{4290}
156 $= 0.93$; $P < 0.001$ in all cases). We selected molecular weight, H-bond acceptor count, aromatic atom
157 count, polarity (X log *P*), and carbon hybridization (sp^3:sp^2) as five properties that are intuitive for non-
158 specialists to interpret (Fig. 1C; Table S1). However, most combinations of properties covering all
159 facets of leaf chemistry also yield equivalent PCAs (Fig. S4). Pairwise correlations between all
160 properties and subset PC scores further confirm that selected properties capture observed leaf
161 chemical variation, in that correlations are consistent within each cluster (Fig. 1A, lower panel).

162 Selected properties also describe biologically relevant chemical space according to an existing
163 framework that classifies natural products based on size, then aromaticity, then lipophilicity (Fig. 1B,
164 right panel; PC1: $r_{4290} = 0.99$; PC2: $r_{4290} = 0.87$; $r_{4290} = 0.81$; $P < 0.001$ in all cases)[31]. The first axis of
165 the subset PCA (38% variance) separates heavy metabolites with many H-bond acceptors from light
166 metabolites with fewer H-bond acceptors. Alignment between metabolite size and numbers of H-bond
167 acceptors is expected because larger molecules usually engage in stronger intermolecular
168 interactions[49]. The second axis (37% variance) separates reactive metabolites with more aromatic
169 carbon bonds from unreactive metabolites with more saturated carbon bonds. This axis is also intuitive
170 because conjugated bond systems are unsaturated, reactive, and planar, whereas saturated bonds,
171 such as those abundant in fatty acids, are unreactive and three-dimensional[50]. The third axis (18%
172 variance) separates nonpolar metabolites from polar metabolites[29]. Our findings thus show that leaf
173 chemical variation can be represented by five chemical properties, one from each facet of leaf
174 chemistry.
175
176 Five chemical properties provide more information than metabolite family
177 The five selected chemical properties not only represent leaf metabolite chemistry (Fig. 1), but also
178 reflect known differences among metabolite families. Using the subset PCA to interpret variation
179 among metabolite families (Fig. 2), we observe that peptides are large, display average bond
180 saturation, and are moderately polar; carbohydrates are of average size, display high bond saturation,
181 and are highly polar; and fatty acids are smaller, display high bond saturation, and are of average
182 polarity. Lignans, alkaloids, coumarins and flavonoids, families of specialized metabolites, display low
183 bond saturation due to the presence of many aromatic bonds, with alkaloids and coumarins also being
184 relatively small. Terpenoids, a family of abundant specialized metabolites (N = 2306), vary in size,
185 bond saturation, and polarity, echoing their diversity across the plant kingdom[51]. Although chemical
186 properties generally differentiate among metabolite families, all families possess variable and
187 overlapping chemistries (Fig. 2, marginal boxplots). We thus suggest that chemical properties,
188 quantitative metrics of the chemistry that underpins metabolite composition and activity, provide more
189 precise information about metabolite function than metabolite family identity. Indeed, drug discovery
190 and medicinal chemistry rely on many of the same chemical properties to screen for bioactive
191 molecules[29–31]. Such a perspective resonates with the functional trait concept by reaching beyond
192 taxonomy to describe function[4] and provides a strong foundation for a novel set of metabolic functional
193 traits.

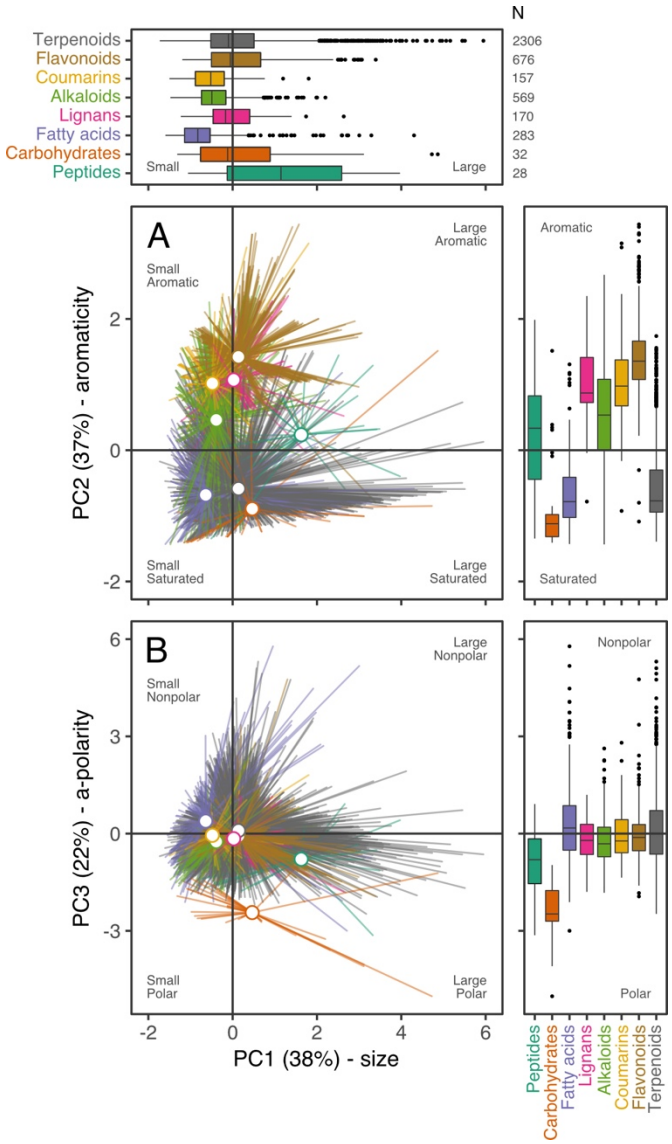


Figure 2 | Five chemical properties describe variation among and within metabolite families. Biplots of PC1 scores and (A) PC2 scores or (B) PC3 scores from a PCA describing variation in selected chemical properties (C hybridization, H-bond acceptors, molecular weight, polarity, aromatic atoms) among all annotated leaf metabolites (N = 4221) detected in 862 plant species. The PCA is functionally equivalent to a PCA containing all 21 chemical properties (see Main Text), with axes separating leaf metabolites based on size (PC1), aromaticity (PC2), and a-polarity (PC3; see Fig. 1A, lower panel). Colors separate metabolite families as defined by marginal boxplot fills/labels (center: median; box: 25% to 75% quantiles; whiskers: 1.5 × IQR; points: outliers beyond whiskers).

Five metabolic functional traits describe two axes of leaf metabolic specialisation

We converted selected properties into plant-level metabolic functional traits to examine how the leaf metabolome varies among tropical and temperate plant species. We first compared the volume and "lumpiness" (i.e., aggregation around values) of multidimensional space occupied by metabolic traits to four null models to test whether species converge on a restricted set of interrelated trait combinations[7]. Observed hypervolumes of metabolic traits are 98.1% to 99.8% smaller and 71.0% to 94.7% lumpier than those of three null models assuming no trait covariance (Table S2; Table S3; always $P < 0.001$ in all cases). However, observed hypervolumes are larger (tropical: $P = 0.002$;

temperate: $P = 0.047$) than that of the null model assuming trait covariance and selection against extreme values, while also being similarly lumpy (tropical: $P = 0.662$; temperate: $P = 0.858$). By comparison, observed hypervolumes of eight classical traits are smaller than null models assuming no trait covariance (Table S2; $P < 0.001$ in all cases), either the same size as (tropical species; $P = 0.154$) or smaller than (temperate species; $P = 0.018$) that of the null model assuming selection against extreme values, and lumpier than all null models (Table S3; $P < 0.001$ in all cases). Collectively, these findings yield three insights. First, metabolic functional traits are interrelated, so plants are constrained to a limited number of viable metabolic trait combinations. Second, while extreme metabolic trait values may be selected against in temperate species, this is not so for tropical species. Expressing extreme leaf metabolome chemistry is thus a viable life-history strategy in tropical environments, potentially due to more intense biotic interactions selecting for a diversity of defensive leaf metabolites and thus divergence of chemical strategies[52]. Third, metabolic traits are less aggregated around non-extreme values than classical traits. As such, although plants converge towards one of several viable classical trait strategies (e.g., woody, non-woody[7]), they can possess a wider array of leaf metabolome strategies.

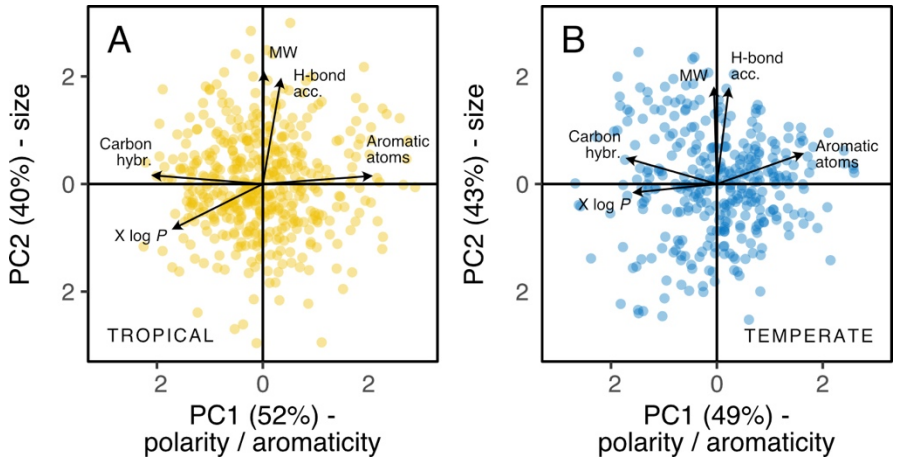


Figure 3 | Metabolic traits describe two axes of leaf metabolic specialization among plant species. Biplots of PC1 and PC2 scores for PCAs describing variation in five metabolic functional traits (arrows) between (A) tropical (yellow; N = 457) and (B) temperate (blue; N = 339) plant species. Axes separate species based on the mean aromaticity *versus* polarity (inverse X log *P*) and carbon bond saturation (PC1) and mean size/complexity (PC2) of all annotated leaf metabolites present (Methods).

We performed PCAs on metabolic traits to characterize major axes of leaf metabolic specialization among species. Metabolic trait variation is consistent among tropical and temperate species, with 92% variation being explained by the first two PCs (Fig. 3). The first axis (tropical: 52%; temperate 49%) separates species that produce more unsaturated, aromatic, nonpolar leaf metabolites from those that produce more saturated, polar, leaf metabolites. This axis likely describes a spectrum of leaf chemical defense strategies[8]. Unsaturated aromatic metabolites, such as alkaloids, coumarins, and flavonoids (Fig. 2), are reactive and serve as toxins or antioxidants in response to stress. For instance, conjugated bond structures can interfere with protein function by binding covalently to sidechains[53], generate or quench oxidative stress[54], or absorb damaging wavelengths of light[55]. By contrast,

248 saturated non-aromatic metabolites can play different roles in leaf defense, both directly as toxins[56]
249 and indirectly as signaling molecules[57]. An alternative explanation is that the first axis reflects an
250 investment (or not) into defense metabolites[8], which could arise due to the high cost of synthesizing
251 aromatic metabolites[17]. However, all plants must defend their leaves against stress, and we find that
252 plants can maximize fitness at either end of this axis (Fig. 3). It is thus more plausible that the first axis
253 represents a leaf chemical defense spectrum.
254
255 The second PCA axis (tropical: 40%; temperate: 43%) discriminates species that produce larger
256 metabolites with many H-bond acceptors from those that produce smaller metabolites with fewer H-
257 bond acceptors. This axis likely relates to leaf longevity, since large metabolites, such as lignans,
258 peptides, and waxes, form the basis of long-lived physical[58,59] and storage[60] structures. Many of these
259 structures also depend on strong intermolecular forces to form stable macromolecules or
260 complexes[61,62], which are enhanced by H-bond donors and acceptors[42]. As for the first axis, plants can
261 maximize fitness at either end of the second axis, with species at the positive end of the axis investing
262 in metabolites indicative of a longer leaf lifespan. As such, the second axis is probably a manifestation
263 of the fast-slow continuum[63] at the leaf level. Alternatively, the second axis may reflect an investment
264 (or not) into structurally complex metabolites for physical defense, such as lignans or waxes[58,59].
265 However, such metabolites act by creating rigid structures that lengthen leaf lifespan anyway[64], so
266 physical protection is one component of leaf longevity. Taken together, metabolic trait PCAs reveal
267 two major axes of leaf metabolic specialization. These axes relate to metabolite bond saturation and
268 polarity (i.e., a leaf chemical defense spectrum) and metabolite size and propensity for intermolecular
269 interactions (i.e., leaf longevity), with many combinations yielding successful life-history strategies.
270
271 <u>Axes of leaf metabolic specialisation reveal hidden dimensions of plant form and function</u>
272 The apparent consistency of metabolic trait variation with classical life-history theory[5–8,17,63,64] raises
273 the possibility that metabolic traits describe aspects of plant form and function already captured by
274 classical traits. We thus performed a PCA on metabolic traits plus the same widely used functional
275 traits used above to test whether axes of leaf metabolic specialization are colinear with, or orthogonal
276 to, trait-based life-history strategies. Considered alone, classical trait variation supports life-history
277 theory[7], with the first two PCs (tropical: 49%; temperate: 51%) describing the leaf economics spectrum
278 (SLA, leaf nitrogen, leaf phosphorus)[6] and variation indicative of plant longevity (plant height, seed
279 mass)[63] and hardiness (stem density, leaf carbon)[5] (Extended Data Fig. 1). However, when combining
280 classical and metabolic traits in the same PCA, the first four PCs (tropical: 66%; temperate: 67%) are
281 equivalent to axes described by individual PCAs (Fig. 4A, lower marginal matrices). The first and third
282 axes describe the first two axes of leaf metabolic specialization (Fig. 3), whereas the second and
283 fourth axes describe the first two axes of classical trait specialization (Extended Data Fig. 1). Pairwise
284 correlations among all traits further show that metabolic and classical traits do not covary (Fig. 4A,B),
285 but separate into six clusters describing unique aspects of either metabolic or classical trait variation.
286 As such, major axes of leaf metabolic specialization vary orthogonally to major axes of classical trait
287 variation (Fig. S4) and are at least equally relevant for describing variation in plant form and function
288 among tropical and temperate plant species.

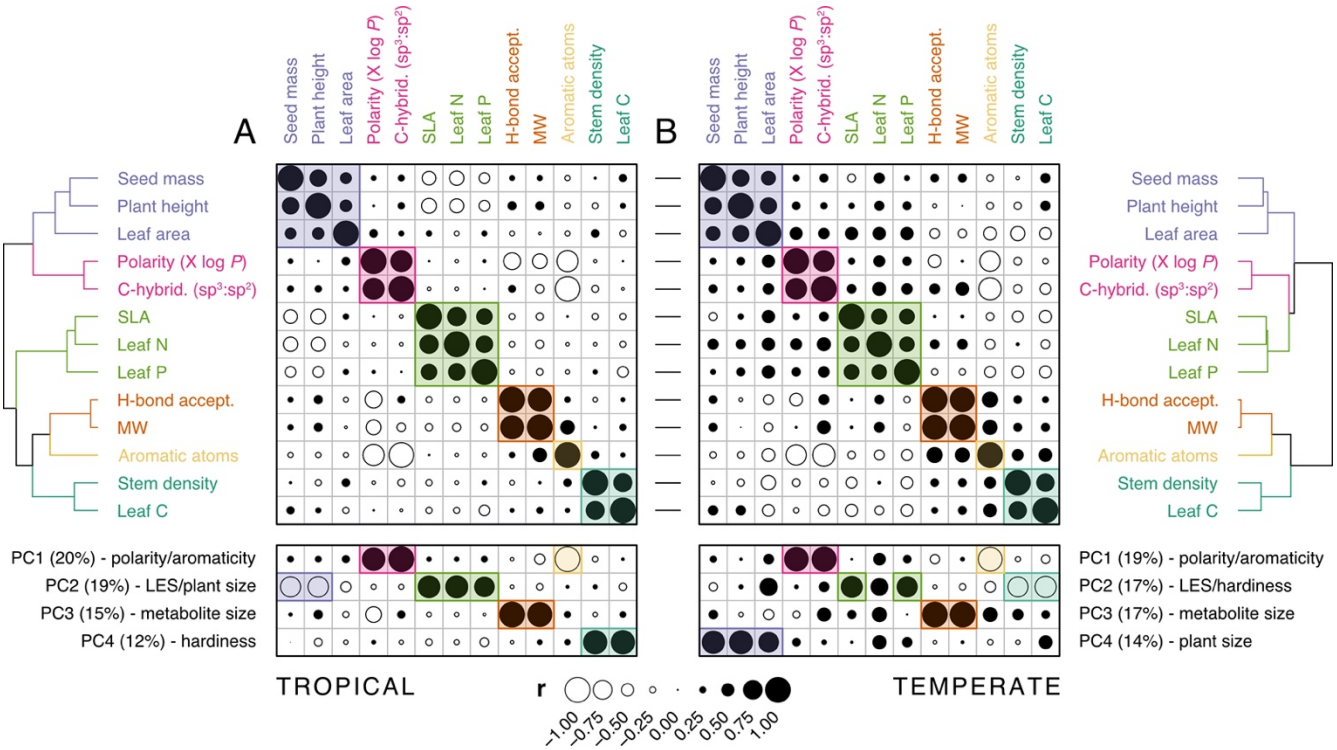


Figure 4 | Metabolic and classical traits describe unique dimensions of plant form and function. Correlation coefficients for pairwise Pearson correlations among metabolic (Fig. 3) and classical functional traits (Extended Data Fig. 1) in (A) tropical (N = 457) and (B) temperate (N = 339) species. Traits are ordered based on correlation similarity (hierarchical clustering; Ward), with colors separating distinct clusters (maximal silhouette width: K = 6) and horizontal lines between matrices showing correspondence between tropical and temperate traits. Lower panel shows correlations between traits and scores from the first four axes of a PCA of the same traits (tropical: 66%; temperate: 67%; LES = leaf economics spectrum). Circle sizes indicate coefficient strengths, while fills discriminate between positive (black) and negative (white) correlations.

Leaf metabolism underpins energy production *via* photosynthesis[65], but in doing so yields tissue that is sensitive to light[66] and temperature[67] and is of high nutritional value to herbivores[64]. Plants are thus under selective pressure to maximize photosynthetic rates while also protecting leaf tissue against damage. While this trade-off is consistent with life-history trade-offs derived from classical traits[5–7], it does not necessarily follow that individuals apply the same strategy in every organ to maintain fitness. Indeed, we observe that species with “productive” trait values (e.g., high SLA, high leaf nitrogen) can produce leaves containing many complex aromatic metabolites, species possessing “hardy” trait values (e.g., high stem density, high leaf carbon) can produce leaves containing many simple polar metabolites, and species possessing trait values typical of a fast pace of life (e.g., low stature, small seeds) can produce leaves with many large stable or structural metabolites. Our findings thus question classical trait[6] and plant defense[8] theory that predicts relationships between the leaf chemical phenotype, plant productivity, and pace of life. In short, metabolic functional traits describe unique dimensions of plant life-history variation that are complementary to, and independent from, those captured by classical functional traits.

Concluding remarks

In summary, we show that leaf metabolite chemistry is highly structured and can be described by five chemical properties. Using chemical properties as metabolic functional traits reveals that plants vary on two axes of leaf metabolic specialization – a leaf chemical defense spectrum (i.e., bond saturation, polarity) and leaf longevity (i.e., size, propensity for intermolecular interactions). Axes of metabolic specialization are similar in tropical and temperate species, with many combinations of trait values yielding successful life-history strategies. Axes are also consistent with expectations from life-history theory that plants invest strategically in tissue defense[6,8,17] and between a fast *versus* slow pace of life[7,63]. Nevertheless, we find that metabolic trait variation is orthogonal to, not colinear with, classical trait variation. Metabolic functional traits thus reveal dimensions of plant life-history variation missed by existing functional traits. It is important to note that metabolomics data were acquired differently for tropical and temperate species and were matched to open functional trait databases. Nevertheless, we find consistent variation in the leaf metabolomes of 457 tropical and 339 temperate species that possess distinct geographic distributions and evolutionary histories (Extended Data Fig. 1A,B), as well as that encompass most documented functional trait variation (Extended Data Fig. 1G, red ribbons). We thus propose that distilling the leaf metabolome into metabolic functional traits captures macroecological patterns of metabolic variation widely across the plant kingdom, enhances the explanatory power of the functional trait concept, and offers a new set of tools for the discovery of species or genotypes with trait combinations adapted to societal needs.

Methods

Sample collection & preparation. We used two existing datasets describing the leaf metabolomes of tropical and temperate plant species (as the unit of observation). The tropical dataset originated from a random subsample of tropical leaves from the Pierre Fabre Sample Library[68], which is an archive of ~17000 plant samples from *in situ* communities around the globe that was originally collected for drug discovery and is a registered collection of the European Union (registration code: 03-FR-2020)[69]. Samples were oven-dried (55 ºC for 3 d) and extracted using overnight maceration (8 g sample; 80 mL ethyl acetate). Extracts were dried once (Genevac, SP Industries, Warminster, PA, USA) and eluted (30 mg extract plus 200 mg silica; 1g Silica SPE cartridges; 6 mL DCM plus 85:15 DCM:MeOH), then dried again and resuspended in DMSO at a final concentration of 5 mg mL^{-1}. The temperate dataset originated from a study of landscape-scale ecological variation in phytochemical diversity[37], wherein leaves were sampled from *in situ* grassland communities across Switzerland. Briefly, leaves were dried (40 ºC for 5 d), milled (Retsch TissueLyser; Qiagen, Hilden, Germany) and extracted (20 mg tissue; 0.5 mL 80:19.5:0.5 MeOH:H_2O: H_2CO_2). Extracts were homogenized (glass beads; 30 Hz for 3 min) and centrifuged (14000 RPM for 3 min), and the supernatant was isolated for measurement.

Metabolomics measurements. Untargeted metabolomics was performed on all samples using Ultra High-Performance Liquid Chromatography Mass Spectrometry (UHPLC-MS). Tropical extracts were analyzed using a Waters Acquity UHPLC system coupled to a Q-Exactive Focus MS (Thermo Scientific, Bremen, Germany) with a heated electrospray ionization (HESI-II) source in positive mode. Extracts (2 μL) were separated on an Acquity C18 column (50 mm × 2.1 mm, 1.7 μm; Waters, Milford, MA, USA) at a flow rate of 600 μl min^{-1} using a binary solvent system (A: H_2O plus 0.1% H_2CO_2; B: acetonitrile plus 0.1% H_2CO_2) and a linear gradient elution (5%-100% B, 7 min) plus isocratic elution (100% B, 1 min). MS data were acquired in data-dependent acquisition mode, in which MS/MS scans were performed on the three most intense ions detected during repeated full MS scans (full scans = 35000 FWHM at m/z 200; MS/MS scans = 17000 FWHM; MS/MS isolation window width = 1 Da; normalized collision energy = 15, 30 and 45 units). The MS was calibrated using a quality control (QC) mix containing caffeine, methionine-arginine-phenylalanine-alanine-acetate (MRFA), sodium dodecyl sulphate, sodium taurocholate and Ultramark 1621 dissolved in acetonitrile/methanol/H_2O plus 1% H_2CO_2. Temperate extracts were analyzed using a Waters Acquity UHPLC system coupled to a Synapt G2 MS (Waters, Milford, MA, USA) with a heated electrospray ionization (HESI-II) source in positive mode. Extracts (2.5 μL) were separated on an Acquity C18 column (50 mm × 2.1 mm, 1.7 μm; Waters, Milford, MA, USA) at a flow rate of 600 μl min^{-1} using a binary solvent system (A: H_2O plus 0.05% H_2CO_2; B: acetonitrile plus 0.05% H_2CO_2) and a linear gradient elution (2%-100% B, 6 min). MS data were acquired in data-independent acquisition mode, in which all precursor ions across the full mass range (85-1200 Da) were fragmented to yield MS/MS spectra. The MS was calibrated using a QC mix derived from all sample extracts.

Metabolomics data pre-processing. Tropical MS data were treated using MZMine (version 2.53)[70]. Peak detection was performed using the centroid mass detector (MS noise = 1×10^4; MS/MS noise = 0) and the ADAP chromatogram builder (min. scan group size = 5, min. group intensity = 1×10^4; min. highest intensity = 5×10^5; m/z tolerance = 12 ppm)[71]. Peaks were deconvoluted using the wavelets algorithm in ADAP with an intensity window SN (S/N threshold = 10; min. feature height = 5×10^5;

380 coefficient/area threshold = 130; peak duration range = 0-5 min; retention time (RT) wavelet range =
381 0.01-0.03 min). Isotopes were detected using the isotopes peak grouper (m/z tolerance = 12 ppm;
382 absolute RT tolerance = 0.01 min absolute; max. charge = 2) and, where present, the most intense
383 isotope was chosen. Peaks were filtered to retain features possessing an MS/MS scan and a RT of
384 between 0.5 and 8 min, following which they were aligned using the join aligner method (m/z tolerance
385 = 40 ppm; absolute RT tolerance = 0.2 min; m/z weight = 2; RT weight = 1). Temperate MS data were
386 treated using MS-DIAL[72]. Peaks were detected with a minimum height of 300 and data were collected
387 with an MS1 tolerance of 0.05 Da and an MS2 tolerance of 0.01 Da from 100–1200 Da and 0.5 to 12
388 minutes. Peaks were deconvoluted with a sigma window value of 0.5 and an MS/MS abundance cut-
389 off of 50. Finally, raw features were aligned using the QC mix with a retention time tolerance of 0.1
390 minute. Following this, we separately clustered tropical and temperate metabolic features into sets of
391 spectrally similar "consensus" features using molecular networks constructed on the Global Natural
392 Products Social (GNPS; precursor and MS/MS fragment ion mass tolerances = 0.02 Da)[73,74]. We
393 retained edges with a cosine score of more than 0.7 and with more than six matched peaks, permitted
394 edges that connected two nodes only if each node appeared in the neighbour's "top 10" most similar
395 node list and, where necessary, limited the total size of a cluster to the 100 highest scoring edges.
396 Finally, we converted peak height/area data to presence-absence data by summing peak
397 heights/areas at the consensus metabolic feature level and expressing them as a one (present) or
398 zero (absent) (tropical N = 7343, temperate N = 6682).
399
400 **Metabolite annotation.** We annotated all metabolic features with chemical information by *in silico*
401 spectral matching and taxonomically-informed scoring using a theoretically fragmented version of the
402 Dictionary of Natural Products[75] following[76,77], retaining matches with a minimum score of 0.7 and at
403 least six matched peaks. We assigned consensus metabolic features with a single consensus
404 chemical classification, which we derived by taking the most common chemical classification across
405 individual features contributing to a consensus feature. We filtered data to retain those features that
406 were annotated with a metabolite identifier (i.e., SMILES notation; tropical N = 3356 (46%), temperate
407 N = 2227 (33%)). We then derived the chemistry of annotated metabolites by using SMILES identifiers
408 to query the Chemistry Development Kit[27] with the package "rcdk"[78]. The CDK comprises 51 chemical
409 descriptors, but we focused on 21 chemical properties that describe metabolite constitution, geometry,
410 and topology (Table S1). We also derived values for annotated metabolites from the first three
411 dimensions of the ChemGPS-NP tool[28], which is an existing framework for classifying natural products
412 that separates molecules based on size, then aromaticity, then lipophilicity[31].
413
414 **Classical functional traits.** All further data processing, analysis and visualization were undertaken in
415 R[79]. For core functionality, we used the R packages "cowplot"[80], "data.table"[81], "drake"[82], "furrr"[83],
416 "future"[84], "parallel"[79], "psych"[85] and "tidyverse"[86] (see below for packages pertaining to specific steps).
417 Plant species names were cleaned against the Catalogue of Life and Encyclopedia of Life using the
418 global name resolver in the package "taxize"[87], following which we added a taxonomic hierarchy using
419 "taxonlookup"[88]. We removed 14 tropical (ferns, cycads, conifers) and nine temperate (ferns, trees)
420 species that were phylogenetically distinct from others within their respective datasets to prevent a
421 small number of divergent species from overwhelming downstream analyses (Fig. S1). We obtained
422 data for classical functional traits with proven relevance to plant form and function from the open-
423 access version of the TRY database (version 5)[89] – namely plant height (m), seed mass (mg), stem

specific density (mg mm^{-3}), leaf area (mm^2), specific leaf area (mm^2 mg^{-1}) and leaf carbon, nitrogen and phosphorus concentrations (mg g^{-1})[6,7,90]. After removing clear outliers (reported error risk ≥ 4; values falling outside of documented limits), we calculated species-wise means and gap-filled missing trait values for all species present in the dataset (N = 46905 species) using a two-step process. First, we performed a Bayesian hierarchical probabilistic matrix factorization imputation (package "BHPMF"[91]), which is a taxonomically-constrained gap-filling approach that has been regularly applied to the TRY database[7,24]. The imputation was repeated 90 times, each time starting with a different combination of parameters (per-fold samples = 900-1000; cross-validation steps = 10-20; burn-in steps = 10% data length) and filtering out extreme (>1.5 times the maximum observed value for a trait) or highly uncertain (coefficient of variation >1) gap-filled values. We then calculated means of remaining gap-filled values across all iterations and used these to replace missing cases in original trait data. Second, we performed five iterations of multivariate imputation by chained equations (package "mice"[92]) on the partially gap-filled dataset and replaced remaining missing cases with the mean values from all iterations. Following this, we filtered fully gap-filled datasets for selected tropical and temperate species (see Table S4 for details about missing data), using genus-level mean trait values where species-level data were not available (215 tropical species, 5 temperate species). It is important to note that trait data coverage differed substantially between tropical and temperate species (Table S4). While this undoubtedly reduced the strength of observed trait variation in tropical relative to temperate data, it is unlikely that the reduced resolution of the tropical data changed the nature of functional trait variation, since artificially reducing the resolution of trait datasets to genus level did not change resulting PC axes (Fig. S2). Finally, we accounted for non-normal distributions in plant height, seed mass and leaf area data using log_{10}-transformations and summarized trait variation among tropical and temperate species separately using scores from the first two components of varimax-rotated PCAs performed on all traits (Extended Data Fig. 1).

Species distributions. We matched species names to Global Biodiversity Information Facility (GBIF) IDs using "rgbif"[93] and gathered all available species occurrences from the GBIF website (www.gbif.org; tropical N = 336632, temperate N = 1847095). Working with each species individually, we filtered occurrences to include human observations and living or preserved specimens recorded from the year 1945 to present day and accompanied by either no individual counts or counts with a reasonable value (>100)[94]. We used the packages "CoordinateCleaner"[95] to discard records falling in the sea, outside the stated country or within 10 km of capital cities, GBIF headquarters or known biodiversity institutions, records lying more than five times the interquartile range of the minimum distances to the nearest neighbor and records containing known issues with longitude/latitude information (i.e., zeros, rounding errors or conversion errors from other coordinate systems)[94,95]. Following this, we manually removed a further 20 tropical and four temperate records representing single occurrences within a continent, leaving a total of 239223 and 1251500 cleaned occurrences for tropical and temperate species, respectively. We plotted cleaned occurrences to visualize the geographical extents of tropical and temperate plant species (Extended Data Fig. 1AB), although it is important to emphasize that samples for metabolomics analyses were taken from within natural (not ornamental) distributions.

Leaf metabolite chemistry. We collated all unique annotated metabolites detected in any tropical or temperate species into a single dataset capturing overall variation in the chemical properties of the

468 leaf metabolome (N = 4292). We tested for interrelatedness between metabolite chemical properties
469 using pairwise Pearson correlations coupled with hierarchical clustering (Ward; Fig. 1, upper panel),
470 and estimated the number of discrete clusters of chemical properties using the maximal average
471 silhouette width[96] (Fig. 1A, dendrogram). We described major dimensions of leaf chemical variation
472 using a varimax-rotated PCA performed on a subset of five representative chemical properties (one
473 from each cluster: C-hybridization ratio, H-bond acceptor count, molecular weight, polarity (X log *P*),
474 aromatic atom count). The first three PC axes described 97% variation, which we anchored back to
475 individual chemical properties using pairwise Pearson correlations (Fig. 1A, lower panel). We tested
476 whether the subset PCA was functionally equivalent to a full PCA performed on all 21 properties using
477 two approaches. First, we performed pairwise Pearson correlations between subset *versus* full PC
478 scores to determine whether the two PCAs separated metabolic features similarly on the same
479 dimensions of chemical variation (Fig. 1B, left panel; see Main Text for statistics). Second, we used a
480 Procrustes Rotation test (package "vegan"[97]; 999 permutations) to determine whether maximally-
481 rotated configurations of full *versus* subset PCAs were significantly (i.e., non-randomly) associated
482 (m_{12} = 0.04, r^2 = 0.98, P < 0.001). We also confirmed that the similarity among subset and full PCAs
483 was not sensitive to which chemical properties were selected from each cluster. This was achieved by
484 repeating Procrustes Rotation tests and PC score correlations for PCAs performed on every
485 combination of five chemical properties that covered all five clusters and plotting distributions of
486 resulting coefficients (Fig. S4). We performed pairwise Pearson correlations between subset PC
487 scores and values from the first three dimensions of the ChemGPS-NP framework[28] to test for
488 correspondence between the subset PCA and existing descriptors of biologically relevant chemical
489 space (Fig. 1B, right panel; see Main Text for statistics). We visualized chemical variation among
490 metabolites using biplots of PC scores from the subset PCA, grouped by biochemical class (Fig. 2).
491 We also illustrated the range of chemical structures captured by the five selected chemical properties
492 by using the PubChem sketcher (https://pubchem.ncbi.nlm.nih.gov//edit3/index.html) to predict
493 chemical graphs of metabolites with the lowest and highest values for each property (Fig. 1C).
494
495 **Plant-level metabolic functional traits.** We selected C-hybridization ratio, H-bond acceptor count,
496 molecular weight, X log *P*, and aromatic atom count to convert into five plant-level metabolic functional
497 traits. For each species (i.e., unit of observation), we calculated the mean value of a chemical property
498 considering all the metabolites present. In doing so, we generated datasets akin those capturing
499 classical functional trait variation among tropical or temperate species, but instead quantifying
500 variation in the mean C-hybridization ratio, H-bond acceptor count, molecular weight, polarity, and
501 aromatic atom count of the annotated leaf metabolome among tropical or temperate species. We
502 considered tropical and temperate datasets as separate entities for data analysis of trait variation, for
503 four reasons. First, while *in silico* data treatment was identical for aligned MS data, sample collection
504 and LC-MS measurements were not (see above), making it challenging to combine data on species-
505 level metabolic variation without inflating differences between them. Second, unlike genomics and
506 proteomics analyses, where only one type of molecule is isolated, untargeted metabolomics aims to
507 simultaneously isolate thousands of molecules with a wide range of chemical properties[3,98]. Technical
508 decisions made during sample extraction and analysis discriminate for or against certain types of
509 metabolites, which filters the resulting view of the metabolome. There is still no single accepted
510 method for extracting samples or acquiring metabolomics data[99], so we chose here to use two
511 datasets generated using different approaches to confirm that findings are not sensitive to the
512 methods chosen. Third, classical functional trait and species occurrence data were sparser for tropical
513 than temperate species (see *Classical Functional Traits*, above), so we chose to keep datasets

separate rather than artificially reduce the resolution of the temperate dataset. Fourth, maintaining a separation between the two datasets increased our interpretive power by providing the opportunity to characterize coupling between the metabolome and functional traits that held true not only within, but also among, tropical and temperate plant species.

Plant multidimensional space. Following[7], we explored constraints on the multidimensional space occupied by metabolic and classical functional traits separately for tropical and temperate species using an observed *versus* simulated hypervolume approach. Briefly, we calculated an *n*-dimensional convex hull volume (package "geometry"[100]) representing the observed multidimensional trait space occupied by species, where *n* is the number of metabolic (n = 5) or classical (n = 8) functional traits. We compared the observed hypervolume to the same four simulated null hypervolumes used by ref. [7], which tested the following null hypotheses: (i) traits vary independently and have uniform distributions (i.e., each trait is a unique axis and no selection against extreme trait values); (ii) traits vary independently but have normal distributions (i.e., each trait is a unique axis and selection against extreme trait values); (iii) traits vary independently but distributions are as observed (i.e., each trait is a unique axis but no assumptions about selection on values); and (iv) traits covary as but have normal distributions (i.e., no assumptions about trait interdependence but selection against extreme values). We simulated null hypervolumes 999 times and used permutation tests (package "ade4"[100]) to test for differences between the overall size of observed *versus* simulated hypervolumes, and presented percent changes between observed and mean simulated hypervolume sizes (where positive = observed > simulated; Table S2). We also calculated the "lumpiness" (i.e., aggregation around certain values) of species in multidimensional space by splitting *n*-dimensional space into 10 bins (i.e., 10^n cells), counting the number of species present in each cell, and calculating the minimum number of cells required to capture 10% of species[7]. Again, we tested for differences between observed and simulated null hypervolumes using permutation tests and presented percent changes between observed and mean simulated hypervolumes (Table S3).

Axes of leaf metabolic specialization. We characterized major axes of leaf metabolic specialization separately among tropical or temperate species using varimax-rotated PCAs performed on metabolic functional traits. The first two PCs explained 92% variation for both tropical and temperate species, respectively, which we displayed using biplots of PC scores (Fig. 3). We explored whether axes of leaf metabolic specialization (Fig. 3) were orthogonal to, or colinear with, axes of classical functional trait variation (Fig. 4) by doing varimax-rotated PCAs containing metabolic plus classical functional traits, again separately among tropical and temperate species. Four PCs were needed to explain 66% and 67% variation among tropical and temperate species, respectively, which we displayed in two ways. First, we created a biplot matrix of all pairwise combinations of the first four axes (Fig. S3) to visualize the lack of overlap between metabolic and classical functional traits. Second, we performed Pearson correlations between PC scores and individual functional traits (Fig. 4, lower panels). Finally, we examined interdependence within and among metabolic and classical functional traits by performing pairwise Pearson correlations (Fig. 4, upper panels), clustering correlation coefficients (hierarchical clustering; Ward), and identifying discrete clusters of traits using the maximal silhouette width[96] (Fig. 4, dendrograms). Dendrograms are organized based on the outcome of a tanglegram that tested for correspondence between the clustering of tropical and temperate traits (Fig. 4, central connectors).

Data and materials availability. Raw tropical metabolomics data are available under the DOI 10.25345/C59J97[36]. Raw temperate species metabolomics data are available under the DOI 10.6084/m9.figshare.13032740[37]. Flat data and R code is available under the DOI 10.17605/OSF.IO/YZR4C.

References

1. Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3–19 (2003).
2. Schuman, M. C., van Dam, N. M., Beran, F. & Harpole, W. S. How does plant chemical diversity contribute to biodiversity at higher trophic levels? *Curr. Opin. Insect Sci.* **14**, 46–55 (2016).
3. Walker, T. W. N. *et al.* Functional Traits 2.0: The power of the metabolome for ecology. *Journal of Ecology* **110**, 4–20 (2022).
4. Violle, C. *et al.* Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).
5. Grime, J. P. *et al.* Integrated screening validates primary axes of specialisation in plants. *Oikos* **79**, 259 (1997).
6. Wright, I. J. *et al.* The worldwide leaf economics spectrum. *Nature* **428**, 821–827 (2004).
7. Díaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).
8. Feeny, P. Plant apparency and chemical defense. in *Biochemical interaction between plants and insects* (eds. Wallace, J. W. & Mansell, R. L.) 1–40 (Springer US, 1976). doi:10.1007/978-1-4684-2646-5_1.
9. Wang, S., Alseekh, S., Fernie, A. R. & Luo, J. The structure and function of major plant metabolite modifications. *Mol. Plant* **12**, 899–919 (2019).
10. Faccio, G. Plant complexity and cosmetic innovation. *iScience* **23**, 101358 (2020).
11. Klee, H. J. & Tieman, D. M. The genetics of fruit flavour preferences. *Nat. Rev. Genet.* **19**, 347–356 (2018).
12. Schmidt, B. M., Ribnicky, D. M., Lipsky, P. E. & Raskin, I. Revisiting the ancient concept of botanical therapeutics. *Nat. Chem. Biol.* **3**, 360–366 (2007).
13. Sardans, J. *et al.* Ecometabolomics for a better understanding of plant responses and acclimation to abiotic factors linked to global change. *Metabolites* **10**, (2020).
14. Sardans, J. *et al.* Ecometabolomics of plant–herbivore and plant–fungi interactions: a synthesis study. *Ecosphere* **12**, (2021).
15. Zanne, A. E. *et al.* Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92 (2014).
16. Hutchinson, G. E. *The Ecological Theater and the Evolutionary Play*. (Yale University Press, 1965).
17. Coley, P. D., Bryant, J. P. & Chapin, F. S. Resource availability and plant antiherbivore defense. *Science* **230**, 895–899 (1985).
18. Capdevila, P., Walker, T. W. N., Schrodt, F., Rodriguez Caro, R. C. & Salguero-Gomez, R. Global patterns of plant form and function are strongly determined by evolutionary relationships. *BioRxiv* (2023) doi:10.1101/2023.01.13.523963.
19. 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).
20. Forrister, D. L., Endara, M.-J., Younkin, G. C., Coley, P. D. & Kursar, T. A. Herbivores as drivers of negative density dependence in tropical forest saplings. *Science* **363**, 1213–1216 (2019).
21. Gargallo-Garriga, A. *et al.* Different “metabolomic niches” of the highly diverse tree species of the French Guiana rainforests. *Sci. Rep.* **10**, 6937 (2020).
22. Scherling, C., Roscher, C., Giavalisco, P., Schulze, E.-D. & Weckwerth, W. Metabolomics unravel contrasting effects of biodiversity on the performance of individual plant species. *PLoS ONE* **5**, e12569 (2010).
23. Adler, P. B. *et al.* Functional traits explain variation in plant life history strategies. *Proc Natl Acad Sci USA* **111**, 740–745 (2014).
24. Bruelheide, H. *et al.* Global trait-environment relationships of plant communities. *Nat. Ecol. Evol.* **2**, 1906–1917 (2018).
25. De Deyn, G. B., Cornelissen, J. H. C. & Bardgett, R. D. Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecol. Lett.* **11**, 516–531 (2008).
26. Lavorel, S. *et al.* A novel framework for linking functional diversity of plants with other trophic levels for the quantification of ecosystem services. *J. Veg. Sci.* **24**, 942–948 (2013).
27. Willighagen, E. L. *et al.* The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. *J. Cheminform.* **9**, 33 (2017).
28. Rosén, J. *et al.* ChemGPS-NP(Web): chemical space navigation online. *J. Comput. Aided Mol. Des.* **23**, 253–259 (2009).
29. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
30. Lovering, F., Bikker, J. & Humblet, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52**, 6752–6756 (2009).

31. Larsson, J., Gottfries, J., Muresan, S. & Backlund, A. ChemGPS-NP: tuned for navigation in biologically relevant chemical space. *J. Nat. Prod.* **70**, 789–794 (2007).
32. Funk, J. L. *et al.* Revisiting the Holy Grail: using plant functional traits to understand ecological processes. *Biol. Rev. Camb. Philos. Soc.* **92**, 1156–1173 (2017).
33. Firn, J. *et al.* Leaf nutrients, not specific leaf area, are consistent indicators of elevated nutrient inputs. *Nat. Ecol. Evol.* **3**, 400–406 (2019).
34. van der Plas, F. *et al.* Plant traits alone are poor predictors of ecosystem properties and long-term ecosystem functioning. *Nat. Ecol. Evol.* **4**, 1602–1611 (2020).
35. Walker, T. W. N. *et al.* Plastic and genetic responses of a common sedge to warming have contrasting effects on carbon cycle processes. *Ecol. Lett.* **22**, 159–169 (2019).
36. Allard, P.-M. *et al.* Open and reusable annotated mass spectrometry dataset of a chemodiverse collection of 1,600 plant extracts. *Gigascience* **12**, (2022).
37. Defossez, E. *et al.* Spatial and evolutionary predictability of phytochemical diversity. *Proc Natl Acad Sci USA* **118**, (2021).
38. Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **69**, 17–20 (1947).
39. Sharma, V., Goswami, R. & Madan, A. K. Eccentric connectivity index: A novel highly discriminating topological descriptor for structure–property and structure–activity studies. *J. Chem. Inf. Comput. Sci.* **37**, 273–282 (1997).
40. Yang, Y., Chen, H., Nilsson, I., Muresan, S. & Engkvist, O. Investigation of the relationship between topology and selectivity for druglike molecules. *J. Med. Chem.* **53**, 7709–7714 (2010).
41. Kuhn, B., Mohr, P. & Stahl, M. Intramolecular hydrogen bonding in medicinal chemistry. *J. Med. Chem.* **53**, 2601–2611 (2010).
42. Zeegers-Huyskens, Th. & Huyskens, P. Intermolecular Forces. in *Intermolecular Forces* (eds. Huyskens, P. L., Luck, W. A. P. & Zeegers-Huyskens, T.) 1–30 (Springer Berlin Heidelberg, 1991). doi:10.1007/978-3-642-76260-4_1.
43. Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **9**, 2215–2235 (2001).
44. Padroni, G., Patwardhan, N. N., Schapira, M. & Hargrove, A. E. Systematic analysis of the interactions driving small molecule-RNA recognition. *RSC Med. Chem.* **11**, 802–813 (2020).
45. Gao, M. & Skolnick, J. A comprehensive survey of small-molecule binding pockets in proteins. *PLoS Comput. Biol.* **9**, e1003302 (2013).
46. Wang, R., Fu, Y. & Lai, L. A New Atom-Additive Method for Calculating Partition Coefficients. *J. Chem. Inf. Comput. Sci.* **37**, 615–621 (1997).
47. Ghose, A. K. & Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **7**, 565–577 (1986).
48. Clemons, P. A. *et al.* Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. *Proc Natl Acad Sci USA* **107**, 18787–18792 (2010).
49. Israelachvili, J. N. The nature of van der waals forces. *Contemp. Phys.* **15**, 159–178 (1974).
50. Van Vranken, D. & Weiss, G. A. *Introduction to Bioorganic Chemistry and Chemical Biology*. 504 (Garland Science, 2012).
51. Pichersky, E. & Raguso, R. A. Why do plants produce so many terpenoid compounds? *New Phytol.* **220**, 692–702 (2018).
52. Forrister, D. L. *et al.* Diversity and Divergence: Evolution of secondary metabolism in the tropical tree genus Inga. *New Phytol.* (2022) doi:10.1111/nph.18554.
53. LoPachin, R. M., Barber, D. S. & Gavin, T. Molecular mechanisms of the conjugated alpha,beta-unsaturated carbonyl derivatives: relevance to neurotoxicity and neurodegenerative diseases. *Toxicol. Sci.* **104**, 235–249 (2008).
54. Widomska, J. & Subczynski, W. K. Mechanisms enhancing the protective functions of macular xanthophylls in the retina during oxidative stress. *Exp. Eye Res.* **178**, 238–246 (2019).
55. Tran, M. H., Phan, D.-P. & Lee, E. Y. Review on lignin modifications toward natural UV protection ingredient for lignin-based sunscreens. *Green Chem.* **23**, 4633–4646 (2021).
56. Tan, N.-H. & Zhou, J. Plant cyclopeptides. *Chem. Rev.* **106**, 840–895 (2006).
57. Lim, G.-H., Singhal, R., Kachroo, A. & Kachroo, P. Fatty Acid- and Lipid-Mediated Signaling in Plant Defense. *Annu. Rev. Phytopathol.* **55**, 505–536 (2017).
58. Yeats, T. H. & Rose, J. K. C. The formation and function of plant cuticles. *Plant Physiol.* **163**, 5–20 (2013).
59. Liu, Q., Luo, L. & Zheng, L. Lignins: biosynthesis and biological functions in plants. *Int. J. Mol. Sci.* **19**, (2018).

60. Staswick, P. E. Storage proteins of vegetative plant tissues. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 303–322 (1994).
61. Sandberg, T. O., Weinberger, C. & Smått, J.-H. Molecular Dynamics on Wood-Derived Lignans Analyzed by Intermolecular Network Theory. *Molecules* **23**, (2018).
62. Pollard, M., Beisson, F., Li, Y. & Ohlrogge, J. B. Building lipid barriers: biosynthesis of cutin and suberin. *Trends Plant Sci.* **13**, 236–246 (2008).
63. Salguero-Gómez, R. *et al.* Fast-slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proc Natl Acad Sci USA* **113**, 230–235 (2016).
64. Coley, P. D. Herbivory and defenses: a temperate/tropic comparison. in *Plant-Animal Interactions: Evolutionary Ecology in Tropical and Temperate Regions* (eds. Price, P. W., Lewison, T. M., Fernandes, G. W. & Benson, W. W.) 54–69 (John Wiley and Sons, 1991).
65. Sulpice, R. & McKeown, P. C. Moving toward a comprehensive map of central plant metabolism. *Annu. Rev. Plant Biol.* **66**, 187–210 (2015).
66. Powles, S. B. Photoinhibition of photosynthesis induced by visible light. *Annu. Rev. Plant Physiol.* **35**, 15–44 (1984).
67. Berry, J. & Bjorkman, O. Photosynthetic response and adaptation to temperature in higher plants. *Annu. Rev. Plant Physiol.* **31**, 491–543 (1980).
68. David, B. & Aussiel, F. High throughput screening of vegetal natural substances. in *Handbook of Chemical and Biological Plant Analytical Methods* (ed. Hostettmann, K.) 1176 (Wiley, 2014).
69. David, B. & Ausseil, F. High-Throughput Screening of Plant Chemodiversity. in *Encyclopedia of analytical chemistry: applications, theory and instrumentation* (ed. Meyers, R. A.) 1–24 (John Wiley & Sons, Ltd, 2006). doi:10.1002/9780470027318.a9944.
70. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
71. Myers, O. D., Sumner, S. J., Li, S., Barnes, S. & Du, X. One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. *Anal. Chem.* **89**, 8696–8703 (2017).
72. Tsugawa, H. *et al.* MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).
73. Frank, A. M. *et al.* Clustering millions of tandem mass spectra. *J. Proteome Res.* **7**, 113–122 (2008).
74. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
75. Chassagne, F., Cabanac, G., Hubert, G., David, B. & Marti, G. The landscape of natural product diversity and their pharmacological relevance from a focus on the Dictionary of Natural Products®. *Phytochem. Rev.* 1–22 (2019) doi:10.1007/s11101-019-09606-2.
76. Allard, P.-M. *et al.* Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. *Anal. Chem.* **88**, 3317–3323 (2016).
77. Rutz, A. *et al.* Taxonomically informed scoring enhances confidence in natural products annotation. *Front. Plant Sci.* **10**, 1329 (2019).
78. Guha, R. Chemical Informatics Functionality in *R. J. Stat. Softw.* **18**, (2007).
79. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2014).
80. Wilke, C. O. *cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2."* (R Project, 2019).
81. Dowle, M. & Srinivasan, A. *data.table: Extension of `data.frame`*. (R Project, 2021).
82. Michael Landau, W. The drake R package: a pipeline toolkit for reproducibility and high-performance computing. *JOSS* **3**, 550 (2018).
83. Vaughan, D. & Dancho, M. *furrr: Apply Mapping Functions in Parallel using Futures*. (R Project, 2021).
84. Bengtsson, H. A unifying framework for parallel and distributed processing in R using futures. *arXiv* (2021).
85. Revelle, W. R. *psych: Procedures for Personality and Psychological Research*. (Northwestern University, 2019).
86. Wickham, H. *et al.* Welcome to the tidyverse. *JOSS* **4**, 1686 (2019).
87. Chamberlain, S. A. & Szöcs, E. taxize: taxonomic search and retrieval in R. [version 2; peer review: 3 approved]. *F1000Res.* **2**, 191 (2013).
88. Cornwell, W., FitzJohn, R. & Pennell, M. *taxonlookup: A dynamically-updating versioned taxonomic resource for vascular plants*. (R Project, 2016).

89. Kattge, J. *et al.* TRY plant trait database - enhanced coverage and open access. *Glob. Chang. Biol.* **26**, 119–188 (2020).
90. Grime, J. P. Vegetation classification by reference to strategies. *Nature* **250**, 26–31 (1974).
91. Schrodt, F. *et al.* BHPMF - a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecology and Biogeography* **24**, 1510–1521 (2015).
92. van Buuren, S. & Groothuis-Oudshoorn, K. mice : Multivariate Imputation by Chained Equations in *R. J. Stat. Softw.* **45**, (2011).
93. Chamberlain, S. *et al. rgbif: Interface to the Global Biodiversity Information Facility API*. (R Project, 2021).
94. Maldonado, C. *et al.* Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Glob. Ecol. Biogeogr.* **24**, 973–984 (2015).
95. Zizka, A. *et al.* CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. *Methods Ecol. Evol.* (2019) doi:10.1111/2041-210X.13152.
96. Maynard, D. S. *et al.* Global relationships in tree functional traits. *Nat. Commun.* **13**, 3185 (2022).
97. Oksanen, J. *et al. vegan: Community Ecology Package*. (R Project, 2019).
98. Weckwerth, W. Metabolomics in systems biology. *Annu. Rev. Plant Biol.* **54**, 669–689 (2003).
99. Uthe, H. *et al.* A practical guide to implementing metabolomics in plant ecology and biodiversity research. in *Plant Metabolomics in full swing* vol. 98 163–203 (Elsevier, 2021).
100. Dray, S. & Dufour, A.-B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.* **22**, (2007).
101. Salguero-Gómez, R. Applications of the fast-slow continuum and reproductive strategy framework of plant life histories. *New Phytol.* **213**, 1618–1624 (2017).

Acknowledgements

Special thanks to Jeanne Clément and Ghislain Vieilledent for building bespoke functionality into the R package “jSDM” for exploratory analyses. This study was part funded by a grant from the Synthesis Centre (sDiv) of the German Centre for Integrative Biodiversity Research (iDiv), awarded to TWNW, FS and NMvD. Plant material collection and metabolomics analysis were financed by Swiss National Science Foundation grants awarded to LP (grant no. 31003A-162604), SR (grant no. 31003A-131956), and JLW (grant no. CRSII5-189921/1). JMA is funded by the Swiss National Science Foundation (grant no. 31003A-176044). FS acknowledges funding by the University of Nottingham Anne McLaren Fellowship. ED and SR are funded by the Swiss National Science Foundation (grant no. 31003A_179481). VEJJ is funded by the French National Research Agency (grant no. ANR-17-CE01-0007: Mixxopeat). VB is supported by the French Agence Nationale pour la Recherche (grant no. ANR-12-BSV7-0016-01: SecPriMe2), the BioDivMeX Mistrals program, the Aix-Marseille University ‘Investissements d’Avenir’ program and the Labex OT-Med programs. PC is supported by a Ramón Areces Foundation Postdoctoral Scholarship. PDC is funded by the US National Science Foundation (grant no. DEB-1135733). NMvD, SM and KP are supported by iDiv Halle-Jena-Leipzig, which is funded by the German Research Foundation (grant nos DFG–FZT 118, 202548816). SN and KP are additionally funded by the German Network for Bioinformatics Infrastructure (de.NBI) and acknowledge BMBF funding (grant no. 031L0107). DLF is supported by the US National Science Foundation Graduate Research Fellowships Program. RSG is funded by a UK Natural Environment Research Council Independent Research Fellowship (grant no. NE/M018458/1). JP and JS are funded by the Spanish Government (grant no. CGL2016-79835), Catalan Government (grant no. SGR 2017-1005) and a European Research Council Synergy Grant (grant no. ERC-SyG-2013-610028: IMBALANCE-P). MCS is supported by a NOMIS Foundation project (“Remotely Sensing Ecological Genomics”) and by the University of Zürich Research Priority Program on Global Change and Biodiversity.

Author contributions

Conceptualization: TWNW in collaboration with PMA, RDB, PC, BD, ED, AG-G, VEJJ, JP, SR, RSG, JS, MCS, FS, NMvD and WW, with support from all co-authors. Funding acquisition: FS, NMvD, and TWNW. Data collection: PMA, ED, and BD. Data curation: TWNW, with support from PMA, OB, PC, and ED. Formal analysis: TWNW, with support from PC, ED, MJE, DF, and VEJJ. Writing: TWNW, with support from MCS. Reviewing and editing: all co-authors.

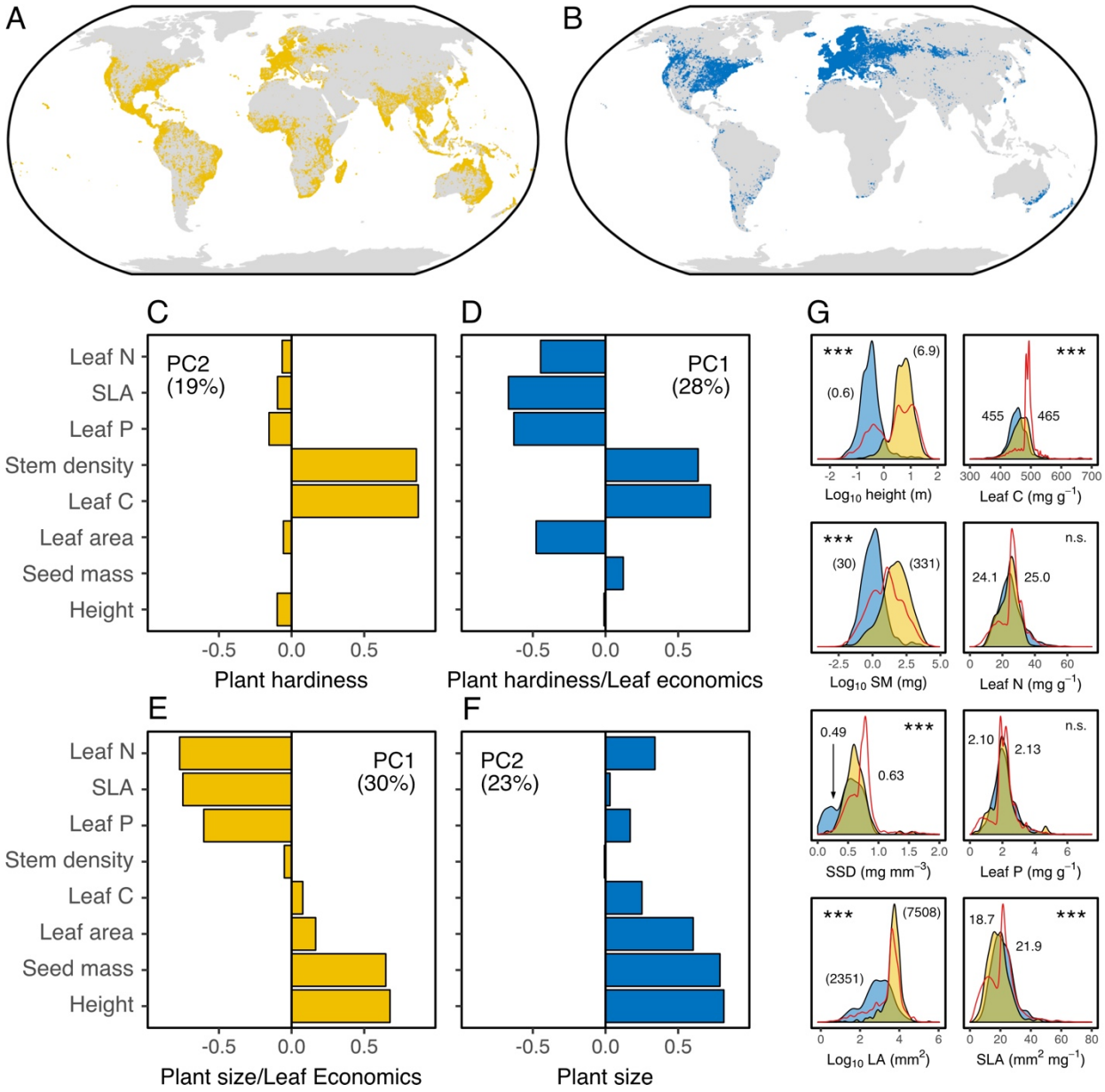
Competing interests

The authors declare no competing interests.

Materials & correspondence

Requests for materials and correspondence should be sent to the corresponding author.

Extended Data Figures



Extended Data Figure 1 | Tropical and temperate species capture major trade-offs in plant functional traits. (A,B) The global geographic extents of (A) tropical (yellow; N = 457) and (B) temperate (blue; N = 405) plant species used in this study. Points are processed GBIF records representing native and non-native occurrences, with sample collection being restricted to *in situ* native populations. (C-F) Loadings for the first two PCs of PCAs performed on eight classical functional traits separately for (C,E) tropical and (D,F) temperate species. Axes collectively capture well-established life-history trade-offs (i.e., leaf economics spectrum[6], plant size/longevity[7,101], plant hardiness[15]; see Main Text), although the coupling of these trade-offs and the ordering of PC axes differs between tropical and temperate species. (G) Kernel density estimates showing the distributions of classical traits used in PCAs (height, seed mass (SM), stem specific density (SSD), leaf area (LA) and leaf carbon, nitrogen and phosphorus) for tropical (blue) and temperate (yellow) species, as well as for all species in the TRY database (red ribbon[89]). Text shows group means, with values in brackets being back-transformed to original units ($\log_{10}$-transformed traits; see Methods). Asterisks illustrate significant differences ($P < 0.001$; linear models) between tropical and temperate species means.