npj Biodiversity

https://doi.org/10.1038/s44185-026-00150-9

Article in Press

Plant chemistry for biodiversity conservation

Received: 13 July 2025

Accepted: 23 July 2026

Published online: 08 August 2026

Cite this article as: Saldanha, L.L., Bigler, L., Nyffeler, R. *et al.* Plant chemistry for biodiversity conservation. *npj biodivers* (2026). https://doi.org/10.1038/s44185-026-00150-9

Luiz L. Saldanha, Laurent Bigler, Reto Nyffeler, Ingrid Olivares & Rodrigo Cámara-Leret

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

ARTICLE IN PRESS

Plant Chemistry for Biodiversity Conservation

Luiz L. Saldanha[1], Laurent Bigler[2], Reto Nyffeler[1], Ingrid Olivares[1] and Rodrigo Cámara-Leret[1]*

[1]Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland
[2]Department of Chemistry, University of Zurich, Zurich, Switzerland

*Corresponding author. Email: rodrigo.camaraleret@uzh.ch

ARTICLE IN PRESS

Plant chemical compounds are crucial for ecosystem functioning and resilience[1], yet the extent to which plant chemical diversity may disappear as threatened plants go extinct[2] remains largely unknown. Here, we show that only 5% of threatened plants are chemically studied and that 35% of these contain 3,037 unique compounds not reported in other species. The remaining 95% of chemically neglected threatened species are concentrated in the Neotropical, Afrotropical, and Indomalayan phytogeographic realms. To accelerate research of chemically neglected threatened plants, we evaluate whether untargeted metabolomics on herbarium specimens can detect a significant fraction of compounds found in living botanical garden plants. We find that herbarium specimens contain 35-84% of compounds of living plants, including from all major plant biosynthetic pathways (e.g., alkaloids, terpenoids, flavonoids), regardless of collection date. Together, our findings introduce chemistry into conservation planning of species and ecosystems and position botanical gardens and herbaria as invaluable resources for unravelling plant chemical diversity at scale.

ARTICLE IN PRESS

Introduction

Over the last decades, significant advances have been made to catalogue the world's biological diversity[3-5]. So far, however, no study has assessed the chemical diversity of the world's flora or its vulnerability to plant extinction. Notwithstanding growing research by chemists on individual species' functions[6,7], ecologists on the chemistry of entire clades to inform speciation (e.g., *Inga*[8]), or evolutionary biologists on the role of secondary metabolites in plant-animal interactions[9], we still do not know how much of the world's chemical diversity has been studied and how much may disappear as threatened plants vanish. Crucially, the potential to advance our understanding of the unique chemistry of threatened plants through the 400 million specimens curated in global herbaria[10] and ex situ botanical garden collections (conserving over 41% of known threatened species[11]), remains unassessed. And yet, identifying and mapping this threatened chemical space at taxonomic, national, and macroecological scales is critical for assessing to what degree chemical arguments can inform science and policy for preserving species and ecosystems.

Results

The known chemistry of threatened plants

To start assessing what is known about the chemistry of threatened plants, we examine how much chemical data exists for 27,359 threatened species assessed by the International Union for Conservation of Nature (IUCN) [2] (see *Methods*). The plant kingdom contains approximately 125,000 known secondary metabolites[12] which chemists have recently started to make available in natural product databases[13,14]. Linking natural product databases with IUCN assessments, we find three striking results about the status of chemical knowledge of the world's threatened flora. First, chemical information (i.e., at least one compound reported in the literature) exists for only 5% of the world's threatened plant species. Of these, 41% have five or fewer reported compounds. Second, gaps in chemical research on threatened plants are spread throughout the plant phylogeny, abound in both old and young plant families, and are greatest in Fabaceae, Rubiaceae, and Lauraceae (Fig. 1). Third, the proportion of chemically surveyed threatened species in a family is uncorrelated to family richness ($r = -0.05$, $P = 0.549$). Notably, this indicates that research effort is not simply a function of a family's species richness and suggests that chemical neglect is pervasive and affects both large and small families alike. In short, the chemistry of threatened plants remains vastly understudied across the phylogeny, irrespective of family richness.

ARTICLE IN PRESS

To better understand the chemistry of threatened plant species, we now turn to examine the identity of the compounds reported in literature. Overall, we find 14,500 compounds reported for the world's threatened plant species. Of these, 3,037 are unique compounds (i.e., compounds associated to a single species and not shared with any threatened or non-threatened species). These unique compounds are associated to 475 species and account for 21% of the compound reports for threatened plants (Fig. 2a). Approximately 15% of the unique compounds are found in critically endangered, 32% in endangered, and 53% in vulnerable species. Notably, a handful of families account for 30% of the unique compounds of threatened species (e.g., Fabaceae, Meliaceae, Taxaceae)(Table S1). Whether these unique compounds from threatened plants exhibit functional roles in ecological processes remains, however, unknown.

Knowing how many unique compounds occur in threatened plants says little of how much structural uniqueness may be at risk. Thus, we next explore the chemical structural diversity of threatened unique compounds using molecular descriptors[20], which we summarize into three principal component axes of molecular size, lipophilicity, and aromaticity. These axes jointly account for 89% of the variability of structural diversity (Fig. 2b) and offer complementary information to compound richness (see *Methods*). We find that families with higher chemical structural diversity tend to have more unique compounds (Fig. S1, r = 0.65, P =8.04e^{-11}). This is the case of Cupressaceae, Fabaceae, or Dipterocarpaceae (Table S1). However, several families depart from this trend and have a relatively low structural diversity but many unique compounds (e.g., Taxaceae and Meliaceae). Considering the main biosynthetic pathways (through which organic compounds of primary and secondary metabolism are synthesized in plants), we find that chemical structural diversity is correlated to the number of unique compounds (r = 0.756; P = 0.049), (Fig. S2). Despite this, shikimates and phenylpropanoids have a structural diversity that is disproportionately high relative to their compound richness when compared to other biosynthetic pathways (e.g., terpenoids). Across plant families, comparative analyses using proportional hypervolume revealed that the chemical signatures of threatened species differ from those of non-threatened species within families. These differences are pathway-specific, with some biosynthetic classes being disproportionately enriched: as for polyketides and fatty acids in Proteaceae, alkaloids in Lauraceae, terpenoids in Jungladaceae, shikimates and phenylpropanoids in Rutaceae, and others reduced, indicating non-random restructuring of chemical space (Fig. 2c, Fig. S3, Table S2-3). These results demonstrate that the threatened structural diversity of species is not simply a

ARTICLE IN PRESS

scaled-down version of that of non-threatened taxa, but rather reflects selective threatened biosynthetic pathways. In summary, our findings reveal that the unique chemical structures of threatened species vary across clades. Previous studies, in turn, have shown that chemical structural classes can co-occur in distantly related families[16] and may not be clustered within families[17] or among closely related species[8]. Collectively, this growing body of evidence suggests that chemical traits can provide a complementary metric for biodiversity and enhance conservation planning frameworks.

Hotspots of chemically neglected threatened species

Having examined what chemistry is at risk across clades, we now turn to identify "chemical hotspots" having a high number of threatened species with unknown chemistry (see *Methods*). We find contrasting patterns depending on whether we assess absolute numbers vs. proportion of threatened species (Fig. 3). On the one hand, the Neotropical, Afrotropical, and Indomalayan phytogeographic realms have the highest absolute number of threatened species with unknown chemistry. This is striking since these regions have the highest richness of plants used by people[18]. On the other hand, the proportion of chemically neglected species that are critically endangered peaks in the Oceanian (55%), Palearctic (28%), and Indomalayan realms (25%). The countries with the highest number of threatened species with unknown chemistry are Brazil, Colombia, Ecuador, Mexico, USA, and Madagascar. Other works have identified these countries as biodiversity research priorities[19,20], and our results suggest that chemical studies in these regions are urgently needed too. We note, however, that our biodiversity data are affected by the "Linnean shortfall"—i.e., most species on Earth have not been described—which results in biodiversity being better documented near research institutions and historical collection centers. Moreover, IUCN extinction risk assessments are geographically biased (i.e., 509 threatened species in Oceania vs. 7,809 in the Neotropical realm) and are affected by the "Wallacean shortfall"—i.e., knowledge on the geographic distribution of most species remains incomplete. This and other large-scale biodiversity knowledge shortfalls[21] likely shape our current chemical knowledge (e.g., bias toward widespread, conspicuous, human utilized species). Despite these unavoidable shortfalls, our maps linking extinction risk and chemical neglect introduce a novel framework for guiding chemical conservation research at national and phytogeographical scales.

ARTICLE IN PRESS

Herbariomics accelerates the study of unknown chemistry of threatened plants

Because 95% of threatened plant species are chemically neglected and plant extinction is a growing concern[22], accelerating phytochemical surveys is urgently needed. While metabolomics of living plant tissues yields a comprehensive picture of the metabolome, finding and sampling thousands of species in the wild is financially and logistically challenging . By contrast, herbaria and botanical gardens preserve species from the world's chemical hotspots and are thus promising, but overlooked, resources for large-scale phytochemical screening[11,23]. Still, it is unclear whether compounds from the main classes of plant metabolites can be detected in herbarium specimens vs. living plants and to what degree this may vary across the Tree of Life. To address this, we implemented an untargeted metabolomics workflow to compare the metabolite profile of leaf tissues of living botanical garden species vs. leaf tissues of herbarium specimens, covering 45 plant species across 27 families and all seven phytogeographic realms (Fig. 4, Supplementary Data 1-5). We found four main results. First, herbarium specimens had 35–84% (mean = 72%, SD =10.3%) of the chemical features found in living plants (Fig. 4). This confirms that the chemical composition of herbarium specimens reflects that of living plants. Second, herbarium specimen's collecting date was not correlated with the total number of features (r = -0.16, P = 0.27) or with the number of features shared with living plants (r = 0.0, P = 0.98; Fig. S4). This shows that even old collections can be tapped for chemical research and lends a chemical argument for the long-term preservation of historical herbaria. Third, herbarium samples had unexpectedly more features than living botanical garden samples (Fig. S5). The cause of these differences is hard to ascertain but may stem from a combination of different environmental conditions among samples[24], timing of collection[25], compound degradation, and to a lesser extent, to specimen preservation methods in the field[26]. To minimize sampling bias and simulate the drying and storage process, we recommend collecting paired samples from the same individual at the same day and location—immediately preserving one portion as a control and pressing the other for the herbarium. Finally, we found no significant differences between herbarium and living plants in the detection of compounds from different biosynthetic pathways across the plant phylogeny (EBT-BH, adjusted P > 0.05) (Fig. S6). Thus, compounds from the major biosynthetic pathways—beyond alkaloids[27]—can be detected in herbarium specimens. Taken together, our results indicate that herbaria can be leveraged to accelerate research on chemically neglected threatened species.

ARTICLE IN PRESS

Discussion

To our knowledge, this is the first study to globally quantify the chemical diversity of threatened plants—integrating phylogeny, extinction risk, chemoinformatics, and phytochemistry. We find that 95% of the world's threatened plants are chemically neglected and that chemically studied threatened species harbour unique compounds not found in non-threatened plants. Threatened plants with unknown chemistry are unevenly distributed across the globe and peak in the Neotropics. Beyond the well-recognized importance of species richness in conservation science, we highlight how the presence of unique threatened compounds in specific clades and phytogeographic realms lends additional chemical arguments for conserving species and ecosystems. Importantly, we demonstrate that herbaria and botanical gardens can significantly accelerate our chemical understanding of threatened plants— revealing yet another important contribution from well-curated natural history collections to science and society.

We identify at least four ways to expand and accelerate our chemical knowledge of threatened plants. First, it will be important to strengthen multi-institutional efforts and address data integration challenges—both of which will benefit from transparent reporting of specimen metadata, analytical procedures, and raw spectra acquisition and processing[28-30]. Second, while our study focused on leaves, developing extraction protocols for other plant parts (e.g., bark, fruits) will be essential. To analyse fresh plant materials, standard procedures usually involve lyophilization as an initial step, followed by grinding and systematic use of polar and apolar extraction solvents such as MeOH/H2O 8:2 and tert-butyl methyl ether. There is not really any significant margin for optimizing extraction protocols, as their effectiveness depends heavily on the specific plant parts (e.g., roots, leaves, flowers) and their physical properties (e.g., thickness, hardness). A potential approach would be to adhere to the procedures outlined in the pharmacopoeia for plants belonging to the same or closely related families. Third, we recommend improving data quality, consistency, and interpretation by identifying compound classes susceptible to degradation[28]. Unlike fresh collections, herbarium specimens are not expected to retain volatile compounds, as these typically small apolar or relatively apolar structures (e.g., terpenes, sesquiterpenes) tend to evaporate rapidly. Susceptibility to degradation is not determined solely by compound class, but also by the functional groups present in each structure. For instance, compounds susceptible to oxidation —such as unsaturated fatty acids, thiols, and amines— undergo partial or complete decomposition due to the presence of double bonds and reactive

ARTICLE IN PRESS

functional groups. Decomposition rates are also significantly influenced by storage conditions, including humidity, temperature, exposure to air (oxygen), and light. The inadvertent presence of fungi and other microorganisms can further accelerate transformation processes, and analyses of such samples often include metabolites derived from these organisms. Finally, we propose capturing a broader spectrum of compounds from natural history collections by using additional instrumentation (e.g., GC-MS[24]) and conducting analyses in positive and negative ionization modes[31] to detect basic and acidic compounds.

Achieving a globally comprehensive chemical knowledge of the world's threatened flora to inform biodiversity conservation will require close collaboration between botanists, chemists, and curators of herbaria and botanical gardens. In this study, we collaborated across these disciplines and institutions to implement a best-practice protocol that recognizes the irreplaceable value of natural history collections and ensures that specimens are sampled sustainably, equitably, and ethically (see *Methods*). Ultimately, sustained institutional support and long-term commitments to maintain, expand, and connect natural history collections will be essential to advance our understanding of global plant chemical diversity, informing biodiversity conservation, and maximizing the societal benefits provided by botanical gardens and herbaria.

ARTICLE IN PRESS

Methods

Data gathering and harmonization. The plant species considered to be threatened were sourced from the International Union for Conservation of Nature (IUCN) List of Threatened Species[2] (https://www.iucnredlist.org, accessed February 2025), the most comprehensive dataset available. We considered 'threatened' terrestrial plant species to be those with a critically endangered, endangered, or vulnerable status. To gather chemical information from the literature, we used two of the largest and most comprehensive datasets of plant metabolites: LOTUS[13] and COCONUT 2.0[14]. To avoid duplications of the same species having different names on the lists, we analysed the list of taxa against the two most comprehensive list of plant names, the World Checklist of Vascular Plants (WCVP, https://powo.science.kew.org) and NCBI taxonomy[32], which resulted in the removal of invalid taxon names, sub-species, horticultural cultivars, and synonyms. The classification of botanical families follows the Angiosperm Phylogeny Group (APG IV)[33]. For COCONUT and LOTUS, we filtered out only data on plant species classified under Viridiplantae

ARTICLE IN PRESS

kingdom (NCBITaxon:33090). After this resolution step, the united chemical data set contained 30,967 unique plant species names, and the threatened dataset contained 27,359 unique names. Lastly, we removed invalid and parsed SMILES from the united chemical data set with RDKit[34] and converted them to InChIKeys (first 14 characters). This resulted in 117,046 unique chemical structures. The united list containing pairs of taxon names and structures was used to search for chemical data from the list containing the names of threatened species, level of threat, and geographic location. Species distribution was sourced from the WCVP using the R package rWCVP version 1.2.6[35]. Chemical structural diversity is expressed here as the spread of metabolite structures in chemical property space[15,36]. Briefly, molecular descriptors data for each compound were calculated from SMILES (Simplified Molecular Input Line Entry Specifications) using RDKit[35] and charted using a model based on principal component analysis. Next, the volume occupied in the chemical space (smallest convex hull for a set of compounds) was computed with the "quickhull" algorithm[37] using SciPy 1.0[38]. With this, a measure (volume) of the spread for sets of compounds was obtained for each class of compounds to obtain a value for structural diversity. To minimize the influence of outliers, only classes of compounds having at least three compounds were considered for the hypervolume calculations. To obtain an overview of threatened compounds, we depicted them against the known chemical space by means of a TMAP[39] to visualize large structure datasets. Chemical structures were organized by similarity to calculate the nearest neighbours using the MAP4[40] (MinHashed Atom-Pair fingerprint up to 4 bonds). Compounds exclusively found in threatened species were coloured to differentiate them from compounds shared between threatened and non-threatened species. The threatened chemical space was calculated as the hypervolume occupied by threatened unique compounds relative to the total hypervolume occupied by all compounds reported in the literature for each botanical family. For each family, the proportional contribution of each biosynthetic class to total hypervolume was calculated for threatened and non-threatened species. Log_2 ratios of these proportions were used to assess pathway-specific shifts, and overall compositional differences were quantified as the Euclidean distance between proportional hypervolume.

Sample collection and preparation. To investigate chemical differences between herbarium and living tissues of the same species, leaf samples were collected from the Zurich Herbarium (Z) and the Botanical Garden of the University of Zürich, Switzerland. At the time of collection, no

ARTICLE IN PRESS

national Access and Benefit-Sharing legislation was in force in the countries of origin of the sampled material (i.e., before 2014). Still, the procedure ensures that all materials, metadata, and sampling events are publicly documented and accessible as suggested by the current Open Research Data (ORD) and FAIR Guiding principles for scientific data management and stewardship.

Leaf samples were collected from specimens at the Zurich Herbarium, prioritizing taxonomically updated collections. Whenever possible, we sampled leaf fragments stored in specimen envelopes. A detailed list of requested specimens, specifying the leaf material and the required quantity, was created. An internal project was set up in the herbarium's "digitalis" database to manage the specimen list, sampling details, and search criteria. Each specimen was digitized and photographed as part of the project. Sampling involved carefully extracting a 1–2 cm^2 leaf section from each specimen, which was placed in a paper envelope labelled with the collector's name, collection number, and specimen barcode. A small pencil mark was made at the sampling site on the specimen, with a sticker indicating the consecutive sample number (e.g., 001 for first-time sampling). Following the sampling event, metadata including the project name, sampler's name, and sampling date were recorded, and a unique identifier linked to the specimen barcode was generated to ensure traceability of all sampling events. For each sampling event, all relevant metadata, along with images, are packed into individual ZIP files, named according to the sampling ID, and are publicly available through the Cámara-Leret lab Zenodo account (https://doi.org/10.5281/zenodo.19691261). In addition to herbarium samples, our ongoing fieldwork in tropical regions continues to generate new materials for local herbaria and duplicate specimens for the Zurich Herbarium (Z).

For living samples, ca. 1 g of fresh leaves were removed. The procedure to document sampled plant individuals and collected leaf part followed the same protocol as for the herbarium specimens detailed above. To minimize intraspecific variation in secondary metabolite composition due to genetic, environmental, and developmental factors, more than one specimen per species was collected when available. Voucher specimens of each accession with preserved diagnostic plant parts and images were prepared and integrated in the Herbarium of the University of Zürich. All documentation of plant material was recorded in the botanical garden's "datisca" database, and then extracted and packed into ZIP files that are are deposited on Zenodo

ARTICLE IN PRESS

(https://doi.org/10.5281/zenodo.19691261) as part of the full research project material documentation.

In total, 151 leaf samples were obtained from 96 herbarium and 53 living specimens. Samples from living specimens were frozen after collection under liquid nitrogen and dried by freeze-drying for 72h. Aliquots of 20 or 50 mg of the dried material from herbarium and living specimens were grounded and extracted using 80:20 MeOH:H_2O with 0.125% HCOOH in a ball mill Retsch® MM400 (30 Hz; 5 min) with 4mm glass beads. Extraction was performed at a 1:30 (*w*/*v*) mass-to-solvent ratio. Ampicillin ($C_{16}H_{19}N_3O_4S$, 0.25 mg/l) was added to the solvent and used as internal standard. Subsequently the samples were centrifuged (Eppendorf Centrifuge 5417R) at 20 °C and 14'000 rpm for 5 min. The supernatant was removed and added to HPLC vials and stored at 5 °C prior to analysis. The extraction solvent was used as blank samples. The quality control (QC) samples consisted of 15 mg aliquots of *Urtica dicoica* L. leaves, *Matricaria chamomilla* L. flowers, and *Lavandula angustifolia* Mill. flowers extracted with 80% methanol. The QC samples were injected every 10 samples.

Metabolomics analysis. All samples (5 µl injection) were analyzed with a *Vanquish™ Horizon UHPLC* System (*Thermo Fisher Scientific*, Waltham, USA) connected to a *Vanquish* detector and a *timsTOF Pro* TIMS-QTOF high-resolution mass spectrometer (*Bruker Daltonics*, Bremen, Germany). Separation was performed with an *Acquity BEH C18* HPLC column (1.7 µm particle size, 2x100 mm, *Waters*) kept at 30 °C. The mobile phase consisted of A: H_2O + 0.1% HCOOH and B: CH_3CN + 0.1% HCOOH. A linear gradient was run from 5 to 98% B within 10 min followed by flushing with 98% B for 1 min at a flow rate of 400 µl min^{-1}. The mass spectrometer was operated in the positive electrospray ionization mode at 4'000 V capillary voltage and 500 V endplate offset with a N_2 nebulizer pressure of 2.8 bar and dry gas flow of 8.1 min^{-1} at 220°C. Spectra were acquired in the mass range from *m/z* 50 to 1'300 at 20'000 resolution (*m/z* 500 full width at half maximum) and at a 1.5 Hz rate. The mass analyzer was calibrated at the beginning of each LC run between *m/z* 158 and 1'450 using a 10 mM solution of sodium formate that was injected using a 6- port-valve with a 20 µL loop giving a mass accuracy below 2 ppm. MS/MS spectra were acquired at a collision energy of 35/70 eV with a 3 *m/z* isolation width in a mass range from *m/z* 50 to 600 at a 2.0 Hz rate. N_2 was used as a collision gas.

ARTICLE IN PRESS

Metabolomics data processing. Metaboscape (Bruker, 2022b) was used for LC-MS data processing following the T-ReX 3D workflow. The .raw files were processed for peak detection, peak filtering, chromatogram construction, chromatogram deconvolution, isotopic peak grouping, chromatogram alignment and gap filling. For peak picking, an intensity threshold of 8000 was used. This resulted in an aligned feature list with 7,895 features (*m/z* x retention time x mass signal intensity). In a second step, features that were only detected in QC or blank samples and without MS/MS spectra were removed. The final data set contained 7,128 features aligned was exported using the export function for GNPS/SIRIUS module. To analyse the spectral data, a molecular network (MN) was built using the Feature-Based Molecular Networking (FBMN) workflow in the Global Natural Products Social Network[42] (GNPS2). The .mgf file containing MS/MS data of all features was organized in a single MN. Ion intensities mean for each feature were color-coded in the nodes of the MN.

Metabolite annotation and statistical analyses. We identified the compounds (features) using experimental and theoretical publicly available spectra libraries. The full MS^2 spectral dataset acquired was searched against Metaboscape and SIRIUS 6.1.0[42] libraries. To obtain a biosynthetic understanding of the annotated features, we further predicted the NP class, superclass and pathway using NPClassifier[43] annotation using CANOPUS[42]. For each species and feature category, the ratio of detections in herbarium samples to living samples was calculated: Ratio=Herbarium detections/living detections. For statistical analysis, only compounds with a COSMIC confidence score above 0.7[44] and biosynthetic class (molecular formulas) with confidence level above 0.9 were considered.

Data availability. All .d (Bruker) and .mzML LC-MS datafiles in positive ionization mode, metadata, and metabolite annotation tables will be made available on the MassIVE repository under accession number MSVXXX. Molecular networking results will be made available upon publication on the UCSD Computational Mass Spectrometry Website: https://gnps2.org/status?task=a18bd6c8ec2241ef85499a97bc5b2b1d&resultdisplay_type=latest. Specimen metadata will be available in Zenodo: https://doi.org/10.5281/zenodo.19691261

ARTICLE IN PRESS

Code availability. All code used in the manuscript will be made available in Github: https://github.com/saldanhall/Endangered-chemistry.

Acknowledgements. We thank René Stalder and Manfred Knabe at the Botanical Garden of the University of Zürich for support sampling living plants and Franziska Schmid at the United Herbaria Z+ZT of the University and ETH Zürich for support sampling herbarium specimens. We are grateful to the members of the Cámara-Leret lab for feedback on a previous version of the manuscript and to two anonymous reviewers. This study was funded by a Swiss National Science Foundation Starting Grant (TMSGI3_211659 to R.C.-L.). We acknowledge the support by the swissuniversities grant (P-5 B Open Science - Phase B (2021 - 2024) / 66. E-Specimens to R.N.) for implementing the research sampling documentation at the herbarium and the botanical garden.

Author Contributions. L.L.S. and R.C.-L. conceived the idea. L.L.S compiled and analysed data with contributions from R.C.-L. and L.B. L.L.S., L.B., R.N., I.O., and R.C.-L. wrote the manuscript. R.N. & I.O. facilitated herbarium sampling and documented sampling of vouchers. All authors approved the final draft.

Competing interests. The authors declare no competing interests.

ARTICLE IN PRESS

Reference list

1 Walker, T. W. et al. Functional Traits 2.0: The power of the metabolome for ecology. *J. Ecol.* **110**(1), 4-20, doi:10.1111/1365-2745.13826 (2022).

2 *The IUCN Red List of Threatened Species*, https://www.iucnredlist.org (2024).

3 Ulloa Ulloa, C. *et al.* An integrated assessment of the vascular plant species of the Americas. *Science* **358**, 1614-1617, doi:doi:10.1126/science.aao0398 (2017).

4 Cardoso, D. *et al.* Amazon plant diversity revealed by a taxonomically verified species list. *Proc Natl Acad Sci USA* **114**, 10695-10700, doi:10.1073/pnas.1706756114 (2017).

5 Cámara-Leret, R. *et al.* New Guinea has the world's richest island flora. *Nature* **584**, 579-583, doi:10.1038/s41586-020-2549-5 (2020).

6 Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. & McPhail, A. T. Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent

ARTICLE IN PRESS

from Taxus brevifolia. *J. Am. Chem. Soc.* **93**, 2325-2327, doi:10.1021/ja00738a045 (1971).

7 Saldanha, L. L. *et al.* Hypoglycemic active principles from the leaves of Bauhinia holophylla: Comprehensive phytochemical characterization and in vivo activity profile. *PLOS ONE* **16**, e0258016, doi:10.1371/journal.pone.0258016 (2021).

8 Forrister, D. L. *et al*. Diversity and divergence: evolution of secondary metabolism in the tropical tree genus Inga. *New Phytol.* **237**(2), 631-642, doi:10.1111/nph.18554 (2023).

9 Schiestl, F. P. *et al.* Orchid pollination by sexual swindle. *Nature* **399**, 421-421, doi:10.1038/20829 (1999).

10 Thiers, B. M. (2025). The World's Herbaria 2025: Report on activity for April–June 2025 (Issue 8.2). New York Botanical Garden. https://sweetgum.nybg.org/science/wp-content/uploads/2025/07/2025_Second_Quarter_IndexHerbariorum_Newsletter.pdf

11 Mounce, R., Smith, P. & Brockington, S. Ex situ conservation of plant diversity in the world's botanic gardens. *Nat. Plants* **3**, 795-802, doi:10.1038/s41477-017-0019-3 (2017).

12 Hart, C. E. *et al.* Defining the limits of plant chemical space: challenges and estimations. *bioRxiv*, doi:10.1101/2025.01.08.631938 (2025).

13 Rutz, A. *et al.* The LOTUS initiative for open knowledge management in natural products research. *eLife* **11**, e70780, doi:10.7554/eLife.70780 (2022).

14 Chandrasekhar, V. *et al.* COCONUT 2.0: a comprehensive overhaul and curation of the collection of open natural products database. *Nucleic Acids Res.* **53**, D634-d643, doi:10.1093/nar/gkae1063 (2025).

15 Larsson, J., Gottfries, J., Muresan, S. & Backlund, A. ChemGPS-NP: tuned for navigation in biologically relevant chemical space. *J. Nat. Prod.* **70**, 789-794, doi:10.1021/np070002y (2007).

16 Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochem.* **64**(1), 3-19, doi:10.1016/S0031-9422(03)00300-5 (2003).

17 Pérez-Mesa, P. A. & Roda, F. Alkaloid evolution in the Solanaceae. *Curr. Opin. Plant Biol.* **85**, 102727, doi:10.1016/j.pbi.2025.102727 (2025).

18 Pironon, S. *et al.* The global distribution of plants used by humans. *Science* **383**, 293-297 doi: 10.1126/science.adg8028 (2024).

ARTICLE IN PRESS

ARTICLE IN PRESS

19 Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858, doi:10.1038/35002501 (2000).

20 Ondo, I. *et al.* Plant diversity darkspots for global collection priorities. *New Phytol.* **244**, 719-733, doi:10.1111/nph.20024 (2024).

21 Hortal, J. *et al.* Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **46**(1), 523-549, doi:10.1146/annurev-ecolsys-112414-054400 (2015).

22 Díaz, S. *et al.* Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* **366**, eaax3100, doi:10.1126/science.aax3100 (2019).

23 Burbano, H. A. & Gutaker, R. M. Ancient DNA genomics and the renaissance of herbaria. *Science* **382**, 59-63, doi:doi:10.1126/science.adi1180 (2023).

24 Foutami, I. J., Mariager, T., Rinnan, R., Barnes, C. J. & Rønsted, N. Hundred fifty years of herbarium collections provide a reliable resource of volatile terpenoid profiles showing strong species effect in four medicinal species of *Salvia* across the Mediterranean. *Front. Plant Sci.* **9**, 1877, doi:10.3389/fpls.2018.01877 (2018).

25 Yilmaz, A., Nyberg, N. T. & Jaroszewski, J. W. Extraction of alkaloids for NMR-based profiling: exploratory analysis of an archaic *Cinchona* bark collection. *Planta Med.* **78**, 1885-1890, doi:10.1055/s-0032-1315396 (2012).

26 Tasca, J. A. *et al.* HPLC-MS detection of pyrrolizidine alkaloids and their N-oxides in herbarium specimens dating back to the 1850s. *Appl. Plant. Sci.* **6**, e1143, doi:10.1002/aps3.1143 (2018).

27 Canales, N. A. *et al.* Historical chemical annotations of *Cinchona* bark collections are comparable to results from current day high-pressure liquid chromatography technologies. *J. Ethnopharmacol.* **249**, 112375, doi:https://doi.org/10.1016/j.jep.2019.112375 (2020).

28 Alseekh, S. *et al.* Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat. Methods* **18**, 747-756, doi:10.1038/s41592-021-01197-1 (2021).

ARTICLE IN PRESS

ARTICLE IN PRESS

29 Clark, T. N. *et al.* Interlaboratory comparison of untargeted mass spectrometry data uncovers underlying causes for variability. *J. Nat. Prod.* **84**, 824-835, doi:10.1021/acs.jnatprod.0c01376 (2021).

30 Heuckeroth, S. *et al.* Reproducible mass spectrometry data processing and compound annotation in MZmine 3. *Nat. Protoc.* **19**, 2597-2641, doi:10.1038/s41596-024-00996-y (2024).

31 Mannochio-Russo, H. *et al*. Untargeted metabolomics sheds light on the diversity of major classes of secondary metabolites in the Malpighiaceae botanical family. *Front. Plant Sci.* **13**, 854842, doi: 10.3389/fpls.2022.854842 (2022).

32 Cox, E. *et al.* NCBI Taxonomy: enhanced access via NCBI Datasets. *Nucleic Acids Res.* **53**, D1711-D1715, doi:10.1093/nar/gkae967 (2024).

33 Angiosperm Phylogeny Group (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181** (1), 1-20, doi: 10.1111/boj.12385 (2016).

34 Landrum, G. *et al*. RDKit: Open-source cheminformatics. *Zenodo*, 10.5281/zenodo.591637 (2025).

35 Brown, M. J. M. *et al.* rWCVP: a companion R package for the World Checklist of Vascular Plants. *New Phytol.* **240**, 1355-1365, doi:10.1111/nph.18919 (2023).

36 Henz Ryen, A. & Backlund, A. Charting Angiosperm chemistry: evolutionary perspective on specialized metabolites reflected in chemical property space. *J. Nat. Prod.* **82**, 798-812, doi:10.1021/acs.jnatprod.8b00767 (2019).

37 Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **22**, 469–483, doi:10.1145/235815.235821 (1996).

38 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261-272, doi:10.1038/s41592-019-0686-2 (2020).

39 Probst, D. & Reymond, J.-L. Visualization of very large high-dimensional data sets as minimum spanning trees. *J. Cheminform.* **12**, 12, doi:10.1186/s13321-020-0416-x (2020).

40 Capecchi, A., Probst, D. & Reymond, J. L. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. *J. Cheminform.* **12**, 43, doi:10.1186/s13321-020-00445-4 (2020).

ARTICLE IN PRESS

ARTICLE IN PRESS

41 Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828-837, doi:10.1038/nbt.3597 (2016).

42 Dührkop, K. *et al.* Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **39**, 462-471, doi:10.1038/s41587-020-0740-8 (2021).

43 Kim, H. W. *et al.* NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. *J. Nat. Prod.* **84**, 2795-2807, doi:10.1021/acs.jnatprod.1c00399 (2021).

44 Hoffmann, M. A. *et al.* High-confidence structural annotation of metabolites absent from spectral libraries. *Nat. Biotechnol.* **40**, 411-421, doi:10.1038/s41587-021-01045-9 (2022).

ARTICLE IN PRESS

ARTICLE IN PRESS

Figure legends

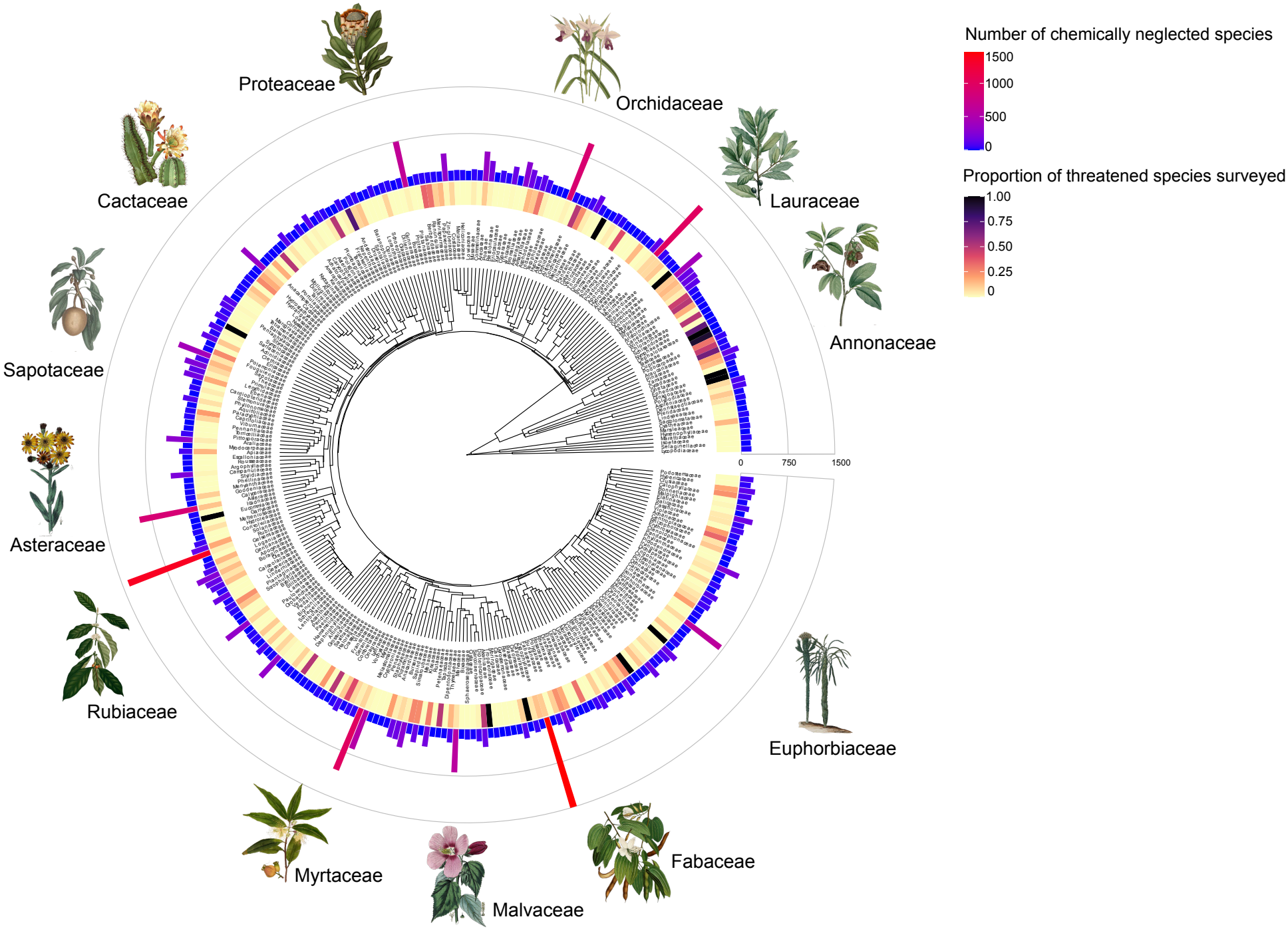
Fig. 1 | Only 5% of threatened plant species have been chemically studied. Inner heatmap shows the proportion of threatened species per family that have been chemically surveyed. Outer bars show the number of threatened species per family that lack chemical studies. Illustrated species highlight families with many threatened species that remain chemically neglected. All illustrations from www.plantillustrations.org belong to the public domain.

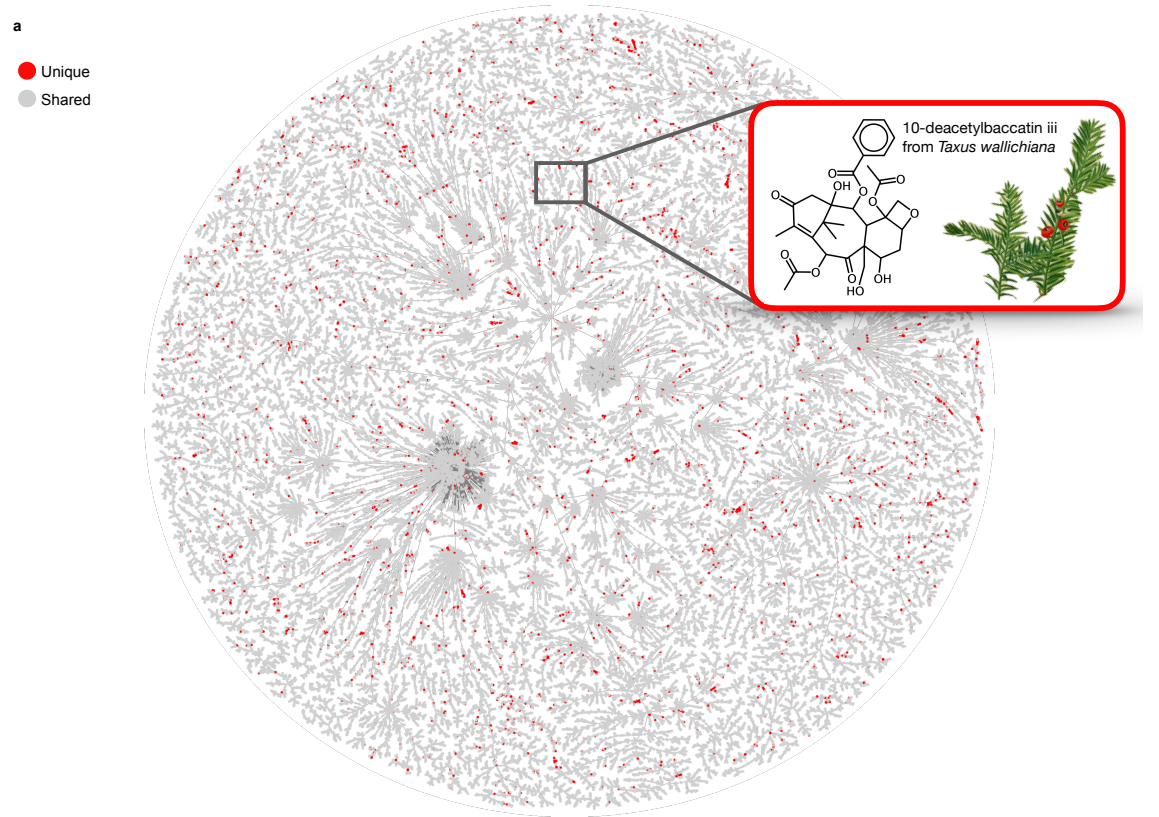
Fig. 2 | The unique chemistry of threatened plants. **a**, The unique compounds of threatened plants (red) within the known chemical space of plants (grey), with one unique compound from a threatened species illustrated. **b**, Chemical structural diversity of threatened species, summarized into three principal component (PC) axes of metabolite molecular size, lipophilicity, and aromaticity. **c**, Proportion of family-level structural space occupied by threatened compounds, grouped by seven biosynthetic pathways.

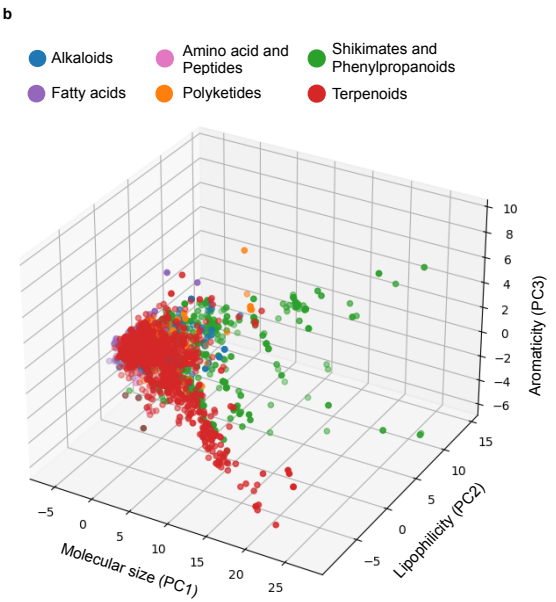
Fig. 3 | Global distribution across phytogeographic realms of threatened plants with unknown chemistry. Known number of threatened species with neglected chemistry per country (purple gradient). Waffle charts display the percentage of threatened species with neglected chemistry per IUCN threat category (in parenthesis: the total number per phytogeographic realm).

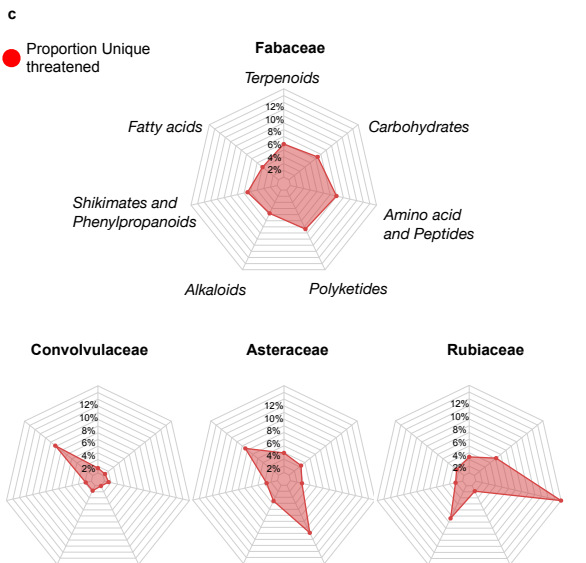
Fig. 4 | Comparison of metabolomic data from living plants and herbarium specimens. a, Number of chemical features detected in living vs. herbarium samples in 45 plant species from 27 families. **b**, Chromatographic profile by Liquid Chromatography-Mass Spectrometry (LC-MS) comparing living vs. herbarium samples of the nutmeg tree (*Myristica fragrans* Houtt., Myristicaceae). **c**, Selected clusters of typical cinnamic acid derivatives and their relative mass signal intensity displayed in the molecular network obtained from the nutmeg tree in herbarium and living samples.

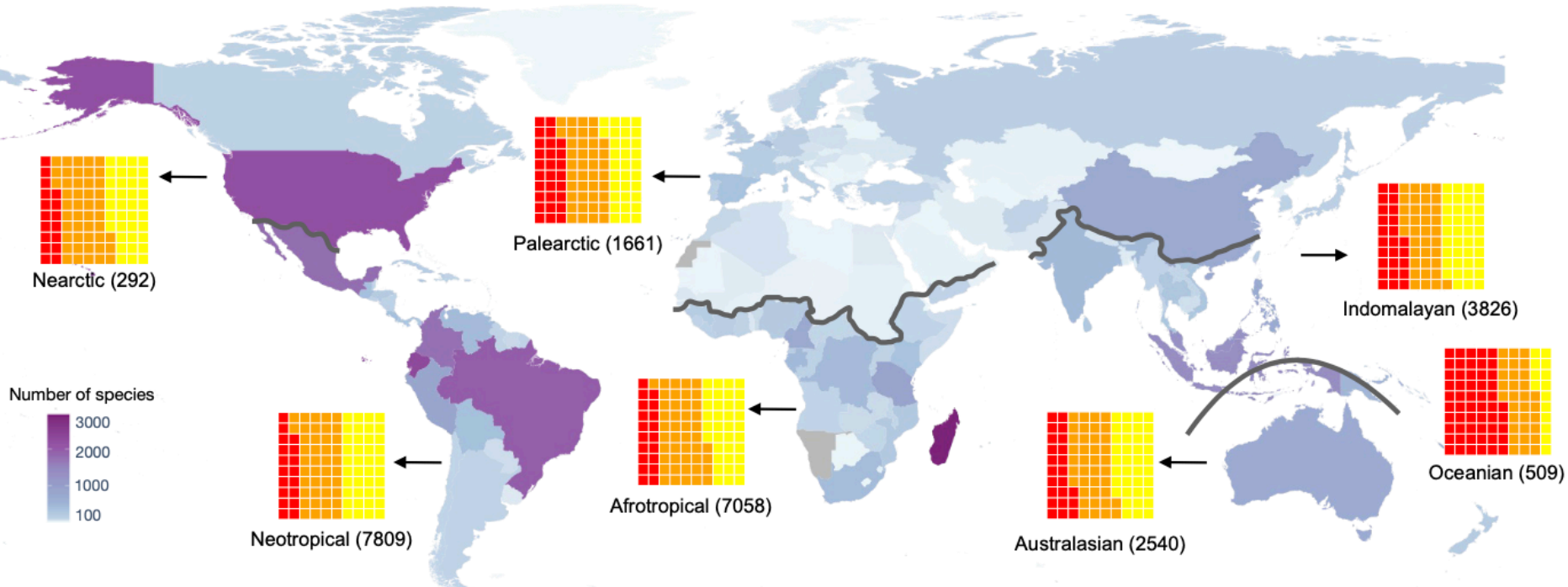
ARTICLE IN PRESS

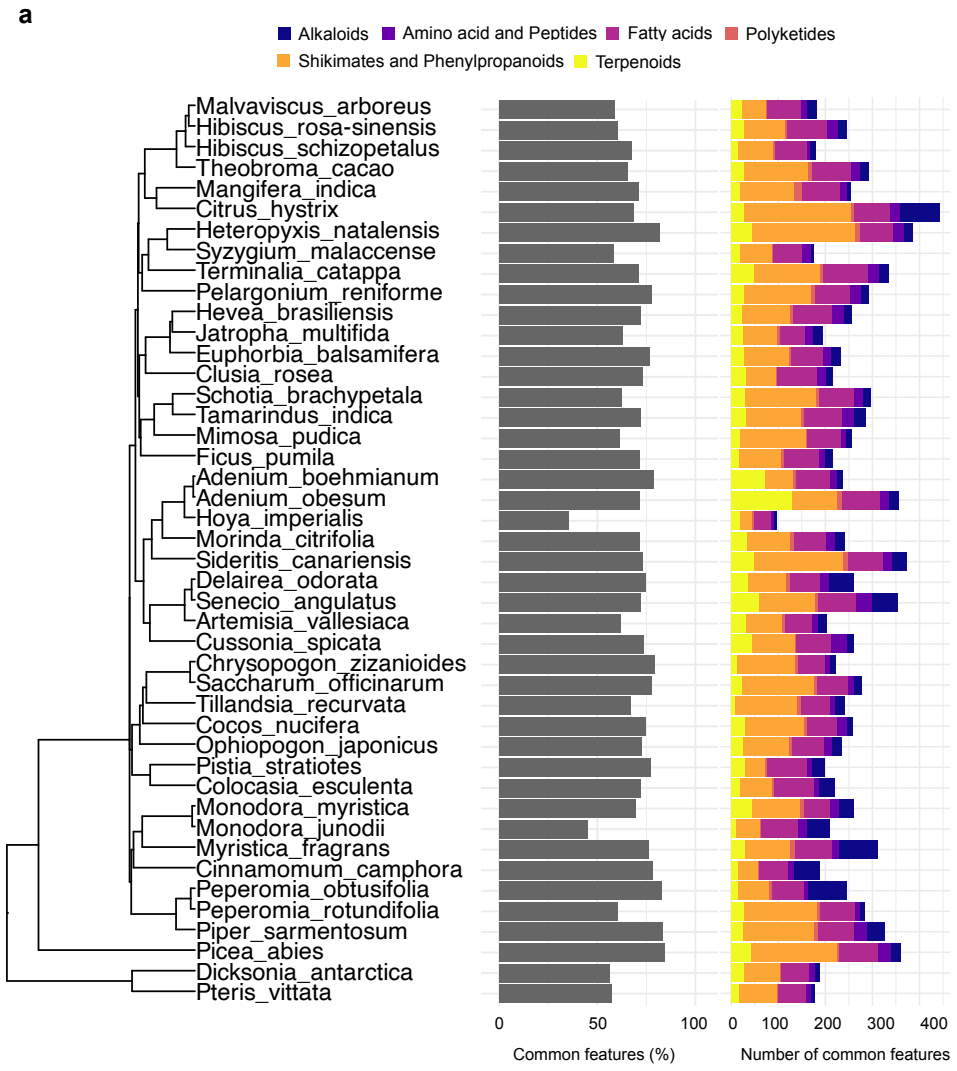

Proteaceae
Orchidaceae
Cactaceae
Lauraceae
Annonaceae
Sapotaceae
Asteraceae
Rubiaceae
Euphorbiaceae
Myrtaceae
Malvaceae
Fabaceae
0
750
1500
Number of chemically neglected species
1500
1000
500
0
Proportion of threatened species surveyed
1.00
0.75
0.50
0.25
0

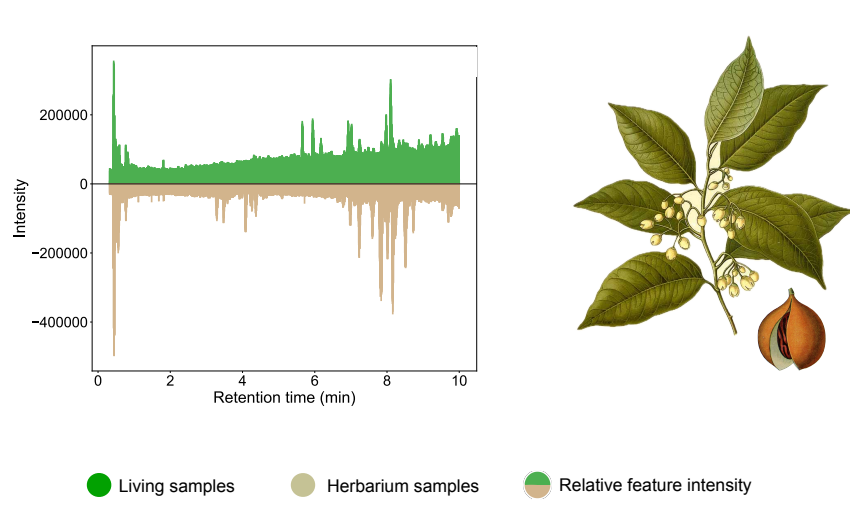

a
Unique
Shared
10-deacetylbaccatin iii
from Taxus wallichiana
OH
O
HO

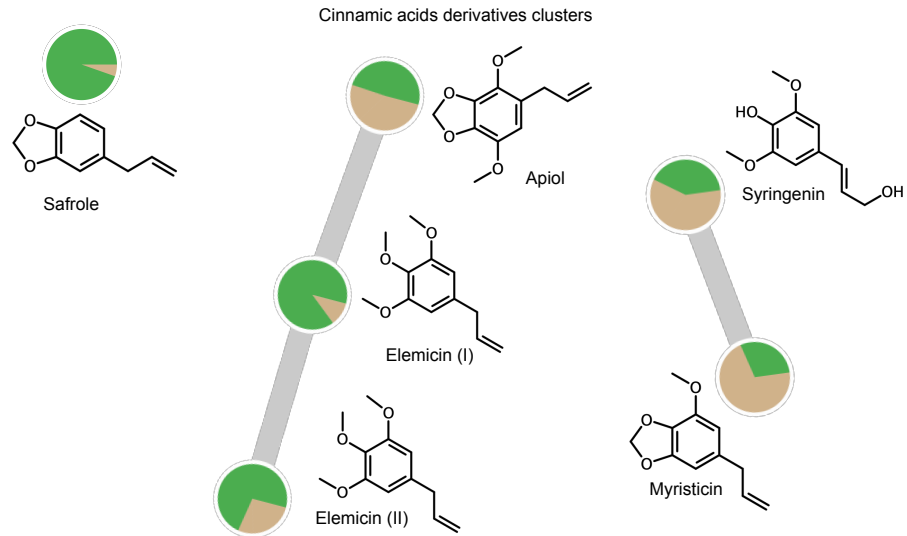

b
Alkaloids
Amino acid and Peptides
Shikimates and Phenylpropanoids
Fatty acids
Polyketides
Terpenoids
Molecular size (PC1)
Lipophilicity (PC2)
Aromaticity (PC3)


c
Proportion Unique threatened
Fabaceae
Terpenoids
Carbohydrates
Amino acid and Peptides
Polyketides
Alkaloids
Shikimates and Phenylpropanoids
Fatty acids
Convolvulaceae
Asteraceae
Rubiaceae


Nearctic (292)
Palearctic (1661)
Indomalayan (3826)
Oceanian (509)
Afrotropical (7058)
Australasian (2540)
Neotropical (7809)
Number of species
3000
2000
1000
100
Threat category of chemically neglected species
Critically Endangered
Endangered
Vulnerable


a
Alkaloids
Amino acid and Peptides
Fatty acids
Polyketides
Shikimates and Phenylpropanoids
Terpenoids
Malvaviscus_arboreus
Hibiscus_rosa-sinensis
Hibiscus_schizopetalus
Theobroma_cacao
Mangifera_indica
Citrus_hystrix
Heteropyxis_natalensis
Syzygium_malaccense
Terminalia_catappa
Pelargonium_reniforme
Hevea_brasiliensis
Jatropha_multifida
Euphorbia_balsamifera
Clusia_rosea
Schotia_brachypetala
Tamarindus_indica
Mimosa_pudica
Ficus_pumila
Adenium_boehmianum
Adenium_obesum
Hoya_imperialis
Morinda_citrifolia
Sideritis_canariensis
Delairea_odorata
Senecio_angulatus
Artemisia_vallesiaca
Cussonia_spicata
Chrysopogon_zizanioides
Saccharum_officinarum
Tillandsia_recurvata
Cocos_nucifera
Ophiopogon_japonicus
Pistia_stratiotes
Colocasia_esculenta
Monodora_myristica
Monodora_junodii
Myristica_fragrans
Cinnamomum_camphora
Peperomia_obtusifolia
Peperomia_rotundifolia
Piper_sarmentosum
Picea_abies
Dicksonia_antarctica
Pteris_vittata
0
50
100
Common features (%)
0
100
200
300
400
Number of common features


b
200000
0
−200000
−400000
Intensity
0
2
4
6
8
10
Retention time (min)
Living samples
Herbarium samples
Relative feature intensity


c
Cinnamic acids derivatives clusters
Safrole
Apiol
Elemicin (I)
Elemicin (II)
Syringenin
HO
OH
Myristicin