Received: 7 August 2023 | Accepted: 24 January 2024

DOI: 10.1002/ajb2.16308

BRIEF COMMUNICATION

American Journal of Botany
BOTANICAL SOCIETY OF AMERICA

Traditional medicinal use is linked with apparency, not specialized metabolite profiles in the order Caryophyllales

Alex H. Crum[1] | Lisa Philander[2] | Lucas Busta[3] | Ya Yang[1]

[1]Department of Plant and Microbial Biology, University of Minnesota, 1445 Gortner Avenue, St. Paul, MN 55108, USA

[2]Como Park Zoo and Conservatory, 1225 Estabrook Drive, St. Paul, MN 55103, USA

[3]University of Minnesota Duluth, 1038 University Drive, Duluth, MN 55812, USA

Correspondence
Alex H. Crum, Department of Plant and Microbial Biology, University of Minnesota, 1445 Gortner Avenue, St. Paul, MN 55108, USA.
Email: a-crum@live.com

Abstract

Premise: Better understanding of the relationship between plant specialized metabolism and traditional medicine has the potential to aid in bioprospecting and untangling of cross-cultural use patterns. However, given the limited information available for metabolites in most plant species, understanding medicinal use–metabolite relationships can be difficult. The order Caryophyllales has a unique pattern of lineages of tyrosine- or phenylalanine-dominated specialized metabolism, represented by mutually exclusive anthocyanin and betalain pigments, making Caryophyllales a compelling system to explore the relationship between medicine and metabolites by using pigment as a proxy for dominant metabolism.

Methods: We compiled a list of medicinal species in select tyrosine- or phenylalanine-dominant families of Caryophyllales (Nepenthaceae, Polygonaceae, Simmondsiaceae, Microteaceae, Caryophyllaceae, Amaranthaceae, Limeaceae, Molluginaceae, Portulacaceae, Cactaceae, and Nyctaginaceae) by searching scientific literature until no new uses were recovered. We then tested for phylogenetic clustering of uses using a "hot nodes" approach. To test potential non-metabolite drivers of medicinal use, like how often humans encounter a species (apparency), we repeated the analysis using only North American species across the entire order and performed phylogenetic generalized least squares regression (PGLS) with occurrence data from the Global Biodiversity Information Facility (GBIF).

Results: We hypothesized families with tyrosine-enriched metabolism would show clustering of different types of medicinal use compared to phenylalanine-enriched metabolism. Instead, wide-ranging, apparent clades in Polygonaceae and Amaranthaceae are overrepresented across nearly all types of medicinal use.

Conclusions: Our results suggest that apparency is a better predictor of medicinal use than metabolism, although metabolism type may still be a contributing factor.

KEYWORDS

apparency, Caryophyllales, hot nodes, medicinal plants, phylogenetic comparative methods, phylogenetics, specialized metabolism

Nearly 80% of the population in low- and middle-income countries relies on medicinal plants as their primary form of healthcare (Hamilton, 2004; Gaoue et al., 2021), and globally, many will access pharmaceuticals developed from plant-based natural product extracts. Natural products in pharmacological studies are also known as specialized metabolites (SMs), reflecting their function in planta rather than their service to people. Specialized metabolites are important for plant survival, helping mediate biotic and abiotic stress responses and playing a role in plant communications (e.g., attracting pollinators). These compounds are extremely diverse, and it is estimated that the plant kingdom produces over one million unique SMs (Afendi et al., 2012). The specialized metabolite profile of a species tends to be influenced by phylogeny—while some SMs are relatively widespread, such as anthocyanins, others are more lineage specific, such as the glucosinolates, which outside Brassicales are only known in the family Putranjivaceae (Rodman et al., 1998). Overall, closely related plants tend to produce similar classes and abundances of SMs, being constrained by the SM biosynthetic pathways of their common ancestor (Wink, 2003; Holeski et al., 2021; Youssef et al., 2023).

Species used in traditional medicine are also clustered into certain phylogenetic lineages, with a species tending to be related to plants with similar use (Saslis-Lagoudakis et al., 2011, 2012; Gaoue et al., 2021). Because medicinal effect is determined by specialized metabolism and closely related plants tend to produce similar SMs, this phylogenetic pattern of traditional use likely reflects phylogenetic clustering of medicinal SMs (Pellicer et al., 2018). If this hypothesis is correct, then traditional medicinal use in a clade should reflect the classes of SMs to be found in that clade. Though this correspondence can be clouded by instances of independent evolution, such as that observed for glucosinolates (Rodman et al., 1998; Huang et al., 2016; Yang et al., 2020), this approach has the potential to guide bioprospecting efforts, specifically as the gaps in reported knowledge in both ethnobotanical uses and metabolomics/natural products studies meet renewed interest in natural products research, in part to combat antibiotic-resistant pathogens (Atanasov et al., 2021).

Although most of the challenges of associating SM and traditional medicinal use across a phylogeny cannot be dealt with in a single study, the order Caryophyllales is a compelling system to examine the relationship between SMs and traditional medicine due to its unique evolution of tyrosine- and phenylalanine-dominated specialized metabolisms. With approximately 12,500 species across 39 families, Caryophyllales includes many well-known species used in traditional medicine including jojoba [*Simmondsia chinensis* (Link) C.K. Schneid; *Simmondsiaceae*], peyote [*Lophophora williamsii* (Lem. ex. J.F. Cels) J.M. Coult.; Cactaceae], and dock (*Rumex* spp.; Polygonaceae). Although the order suffers from the same lack of characterization of metabolites as in all other plant groups, the active constituents of some of the order's more charismatic plants have been identified. For example, mescaline from peyote is synthesized from tyrosine (Tyr) and is responsible for the plant's hallucinogenic effect.

Mescaline is also emblematic of a unique evolution of specialized metabolism in the order, which is also responsible for the Caryophyllales-specific betalain pigments. Mescaline, dopamine, and norepinephrine, together with the betalain pigments, which range from yellow to pink, are derived from the amino acid Tyr, and are part of a Tyr-derived SM diversification in the order (Chen et al., 2003; Lopez-Nieves et al., 2018). This diversification is the result of the duplication of the gene encoding arogenate dehydrogenase (ADH), which is responsible for converting Tyr from its precursor, arogenate, early in Caryophyllales (Lopez-Nieves et al., 2018). While the canonical version (ADH-β) is feedback-inhibited by excess tyrosine, the new copy (ADH-α) has relaxed that regulation. Species with ADH-α therefore produce an excess of tyrosine, which in Caryophyllales has been incorporated into specialized metabolism in a variety of ways, including isoquinoline alkaloids such as mescaline, betalains, and catecholamines (Lopez-Nieves et al., 2018). These SMs, in turn, help plants mediate both abiotic and biotic stressors such as UV light and herbivory. In the case of the betalains, there is some evidence that these SMs have some advantages over their non-tyrosine derived functional homologs in other plants, such as being more pH stable and therefore providing more protection for CAM plants than the more common anthocyanin pigments (Jain and Gould, 2015).

Conversely, arogenate is also the precursor to the amino acid phenylalanine (Phe), which is itself a major precursor to phenylpropanoid SMs, such as flavonoids, lignins, and anthocyanins. These SMs are canonically a huge component of specialized metabolism, aiding in structure, pollinator attraction, and defense. For example, anthocyanin pigments, which range from red to blue, play many roles, including photoprotection, pollinator attraction, and antioxidant action (Landis et al., 2015). However, because Tyr and Phe compete for the same precursor, the duplication and relaxed regulation of ADH in Caryophyllales means less arogenate is fed into Phe synthesis and the downstream pathways.

One apparent consequence of this competition between Tyr- and Phe-dominant SM pathways is the mutual exclusivity of betalains and anthocyanins in the order. In plants, betalains are exclusively found in Caryophyllales. However, the anthocyanins, derived from Phe, are ubiquitous throughout the plant kingdom. In lineages that have betalain, anthocyanin is never found, and vice versa (Clement and Mabry, 1996). Although the gain of ADH-α appears to have occurred only once, betalain itself likely had multiple origins in Caryophyllales, leading to a pattern of anthocyanin-producing families sister to betalain-producing ones (Sheehan et al., 2020; Figure 1). These later-diverging anthocyanin-producing lineages (the most notable being the order's namesake, Caryophyllaceae) tend to lose the function of their ADH-α or lose it entirely, making the type of pigment a species produces a good proxy for presence/absence of ADH-α and therefore also a shorthand for Phe- vs. Tyr-dominant metabolism types (Lopez-Nieves et al., 2018).

This unusual tension between SM pathways makes Caryophyllales a compelling system to study the relationship between specialized metabolism and human medicinal use. We predicted that within Caryophyllales, lineages with different dominant metabolic pathways would be associated with different categories of medicinal uses. For example, several Tyr-derived SMs such as salidroside, mescaline, and catecholamines affect the nervous system, while there is some evidence supporting anticancer, cardiovascular, and dermatological benefits of Phe-derived SMs (Berman et al., 2017; Zhong et al., 2018; Casarini et al., 2020; Vamvakopoulou et al., 2023). Tyrosine- and phenylalanine-derived metabolites in general typically make up a large part of SM in plants, with the phenylalanine-based phenylpropanoid pathway playing a large role in plant stress management and in potential medicinal effect for humans (Sun and Shahrajabian, 2023).

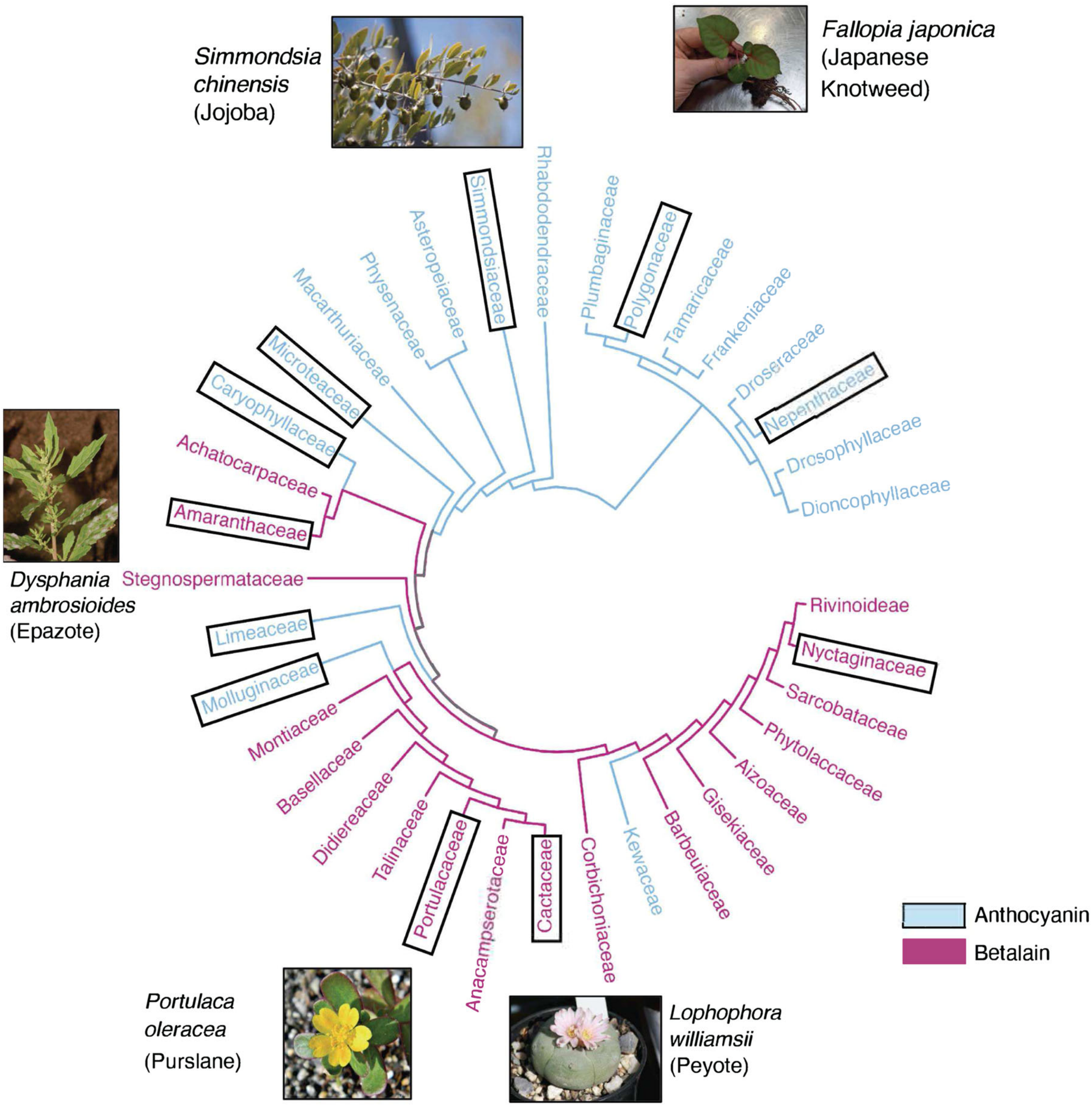

FIGURE 1 A family-level phylogeny of Caryophyllales. Blue branches: anthocyanin-producing lineages, magenta: betalain-producing, gray: ancestral pigment type uncertain. Families surveyed in our literature review of global medicinal uses are outlined in black boxes. Representatives of well-known medicinal species are shown next to the family, including *Simmondsia chinensis* (Simmondsiaceae), *Fallopia japonica* (Polygonaceae), *Dysphania ambrosioides* (Amaranthaceae), *Portulaca oleracea* (Portulacaceae), and *Lophophora williamsii* (Cactaceae). Photo credits: Ken Bosma (Jojoba); Barry Hammel (Epazote); Bob Peterson (Purslane); CC-BY-2.0.

To test the hypothesis that lineages with Tyr- and Phe-dominant specialized metabolism have different categories of use in traditional medicine, we used a phylogenetic "hot nodes" approach to explore over- and underrepresentation of different categories of medicinal use in selected lineages of the order, with representatives from both betalain- and anthocyanin-producing families (Saslis-Lagoudakis et al., 2012; Pellicer et al., 2018; Zaman et al., 2021). To test whether a factor other than metabolism, such as proximity to humans, may be driving medicinal plant selection, we used Global Biodiversity Information Facility (GBIF) occurrence data on all North American members of the Caryophyllales to explore the relationship between medicinal use and how apparent or frequently encountered and noticed a species is to humans via phylogenetic generalized least squares regression (PGLS).

MATERIALS AND METHODS

Literature search

Assembling all known medicinal uses in the ~12,500 species of Caryophyllales is prohibitive. Therefore, we selected families that represent the phylogenetic and metabolic diversity of the Caryophyllales (Figure 1). These include ancestrally Phe-dominant, anthocyanin-producing families (Polygonaceae, Nepenthaceae, and Simmondsiaceae), Tyr-dominant, betalain-producing families (Cactaceae, Portulacaceae, and Amaranthaceae), and anthocyanin-producing families in later-diverging Caryophyllales representing loss of ADH and reversals to Phe-enriched metabolism (Caryophyllaceae, Limeaceae, Molluginaceae, and Microteaceae; Lopez-Nieves et al., 2018). Apart from representing either Phe- or Tyr-dominant specialized metabolism using contrasting pigment types as a proxy, these families either have well-known species with medicinal use (i.e., San Pedro cactus, rhubarb, and chickweed), or are of manageable sizes. The selection of families also ensures that the tree is roughly balanced between anthocyanin- and betalain-producing lineages. Additionally, to mitigate the potential of a specific climate selecting for metabolites of a given effect, these families represent a range of climatic conditions, with desert, tropical, and temperate species represented in selected anthocyanin- and betalain-producing families with an overall global distribution.

The medicinal uses in these families across the globe were reviewed using Google Scholar, the Native American Ethnobotany Database (NAEB), and the PubMed database via the R package rentrez (Moerman, 2003; Winter, 2017). When taxonomy in the reported medicinal use differed from the species name on the guide tree (see methods below), it was checked using Plants of the World and Tropicos and reconciled (POWO, 2023; Tropicos). Specific medicinal use and culture/region were recorded for each species. Medicinal uses were then categorized into broader categories based on physiological system (e.g., respiratory, cardiovascular, or digestive) according to Level 2 of the Economic Botany Data Collection Standard (Cook, 1995). The region of reported use for a plant species was categorized using the M49 standard of the United Nations Statistics Division (1999).

Testing for phylogenetic clustering

Phylogenetic clustering of medicinal use was tested using a "hot nodes" approach, a method originally designed to test for phylogenetic clustering in an ecological community (Saslis-Lagoudakis et al., 2012; Ernst et al., 2016). In this method, nodes on a phylogeny with statistical overrepresentation of species with a reported medicinal use in a given category are "hot". Additionally, nodes with statistical underrepresentation are "cold" nodes. The Caryophyllales phylogeny as published by Smith et al. (2018) was trimmed to the selected families using the ape package in R, resulting in a tree with 3629 tips (Paradis and Schliep, 2019). Over- and underrepresentation of each category of medicinal use across this guide tree was tested using the nodesig command in Phylocom v4.2 (Webb et al., 2008). This function takes the number of species in a medicinal category and simulates a category with the same number of species randomized across the tree. For each medicinal category, we used the default number of 999 simulated runs. If a node includes more medicinal species of that category than that node in 97.5% of simulated runs, it is classified as a hot node. If it has fewer medicinal species than 97.5% of the simulations, it is classified as a cold node. Therefore, each node was classified as hot, cold, or insignificant for each medicinal use category with no associated confidence interval. A hot nodes analysis was run for each medicinal category in the global analysis of selected families as well as all Caryophyllales families in North America. At the global level, due to the low number of reported uses in the Nervous System and Mental Disorders categories and their action on the same body system, the two categories were combined for hot nodes analysis. All other categories were kept separate. Phylocom output, region of use, and reported species were visualized on the phylogeny in R using the ggtree and ggtreeExtra packages (Yu et al., 2017; Xu et al., 2021).

Phylogenetic regression of North American medicinal use

To explore the relationship between medicinal use and apparency to humans, we collected medicinal uses exclusively from North America and tested them for hot nodes as described above for the entirety of Caryophyllales. Additionally, occurrence records of all Caryophyllales species from the M49 standard of North America (Bermuda, Greenland, the USA, Canada, and St. Pierre and Miquelon) were downloaded from GBIF, cleaned, and quality checked using the R package CoordinateCleaner (United Nations Statistics Division, 1999; Zizka et al., 2019). Data set cleaning included removing data points with zero or without latitude and longitude, without a country code corresponding to the M49 standard definition of North American countries, and those with locations that matched a country centroid, administrative capital, or biodiversity institution (United Nations Statistics Division, 1999). Data points without a species name were also dropped. For filtering by institution or capital, data points within a radius of 1000 m around the site's latitude and longitude was removed. Occurrences belonging to species not on the guide tree were discarded. Reciprocally, species not in the occurrence data set were trimmed from the guide tree. Each remaining species on the guide tree was classified as "medicinal" or "not medicinal". Occurrence records for each species were counted and used directly in PGLS as a proxy for apparency, which was tested for correlation with medicinal species. The PGLS was conducted in R using the package caper (Orme et al., 2023).

Analysis was constrained to one region to mitigate any potential bias due to some regions having a more complete ethnobotanical record in literature than others. North America is an ideal choice because, for one, it has a relatively comprehensive ethnobotanical record, thanks in part to NAEB. However, the general population is, for the most part, not reliant on traditional plant medicine, and there is less emphasis on native medicinal plant research here than in regions such as southern and southeastern Asia (Moerman, 2003). This comparative lack of interest in medicinal plants means that in North America, species are less likely to be collected or observed specifically because they are medicinal and that the number of occurrence records can be a good proxy for apparency.

RESULTS

Global analysis of selected Caryophyllales families

A total of 1781 categorized medicinal uses matching 465 species across 11 Caryophyllales families were retrieved from the literature search (data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.6q573n64z; Crum, 2024). Of those, 319 species (69%) with 1365 uses overlapped with the guide tree. Any species not on the guide tree were discarded from further analysis.

The category with the highest number uses was Infections & Infestations (Appendix S1: Table S1). North America had the most species with reported uses (108 species); however, the most reported uses across species and categories come from Southern Asia (392 uses; Appendix S1: Table S2). Eastern Asia had the highest average number of different use categories per species (4.28; Appendix S1: Table S2).

Clades with hot nodes were largely consistent among categories, although categories with fewer reported uses predictably had fewer hot nodes (Figure 2; Appendix S2). Cactaceae, a betalain-producing family with 1108 species represented on the guide tree (out of approximately 1851 total species in the family; 60%), consistently had scattered cold nodes and only a few, mostly deep, hot nodes across the categories, with the notable exception of the Endocrine System Disorder category, which had a hot clade in the genus *Opuntia* (Korotkova et al., 2021; Appendix S1: Table S3; Appendix S2). Outside of Cactaceae, cold nodes were relatively uncommon. On the other end of the spectrum, the anthocyanin-producing family Polygonaceae, with only 565 of about 1200 species (47%) represented in the guide tree, was fairly consistently overrepresented across medicinal use categories (Burke and Sanchez, 2011; Figure 2; Appendix S2). Hot nodes in Polygonaceae mostly occurred in the clades representing the tribes Rumiceae and Persicarieae. Hot nodes in Caryophyllaceae were few and less consistent across categories, with medicinal uses scattered across the family (Figure 2). Amaranthaceae hot nodes mostly occurred in the Amaranthoideae and Gomphrenoideae subfamilies, with less consistent hot nodes in the Chenopodioideae subfamily occurring in some categories (Figure 2; Appendix S2).

North American analysis of all Caryophyllales families

Based on the overlap with North American Caryophyllales GBIF data, the guide tree was trimmed to 914 species, 103 of which had medicinal use in at least one category, for a total of 300 medicinal uses. Similar patterns of use were observed, although hot clades were less consistently present across categories due to the reduced sample size, and cold nodes overall were less common compared to the global analysis. When present in Amaranthaceae and Polygonaceae, hot nodes were in the same clades highlighted in the global analysis. Nyctaginaceae, now a larger percentage of the tree, had hot nodes in several additional categories on the North American tree. Cactaceae still had few, inconsistent hot nodes across categories, again with the exception of the "Endocrine System Disorders" category (Figure 3; Appendix S3). Medicinal use significantly correlated with a higher level of apparency ($F = 171.2$, df = 1, $P < 2.2\text{e-}16$, $\lambda = 0.005$; Figure 4).

DISCUSSION

Despite the dichotomy of Tyr- or Phe-dominant metabolism in Caryophyllales, overall, reported traditional medicinal uses did not differ by metabolism type. Instead, the same clades were largely selected for medicinal use repeatedly across all categories and geographic regions. In the global analysis, the two main hot clades in Polygonaceae and Amaranthaceae were also notable for having a pattern of use across many regions in all categories.

Species in these hot clades also tend to be common or even weedy and have a wider geographic range of distribution. For example, the Polygonaceae genus *Rumex*, which is consistently overrepresented across medicinal use categories, is worldwide in distribution, with several species considered invasive in the United States. In the larger categories, such as Digestive System Disorders, most *Rumex* species are also used in multiple regions. *Rumex* species used in multiple categories, such as the ubiquitous *Rumex crispus* L., are also usually used in multiple regions depending on category. On the other hand, Cactaceae, which had few hot nodes, has many slow-growing species largely restricted to North and South America. Although a few *Opuntia* species have been introduced worldwide by humans and are used outside their native range, for the most part, medicinal use of Cactaceae is restricted to the North America and Latin America/Caribbean regions. Common species being disproportionately used medicinally has been reported in several cultures and locations in the Highland Maya of Mexico and South Africa (Stepp and Moerman, 2001; Lewu and Afolayan, 2009).

The North American analysis supports apparency associated with medicinal plant selection

The pattern of increased medicinal use in common, widespread species from our global analysis suggests that plant species that have more contact with humans and a

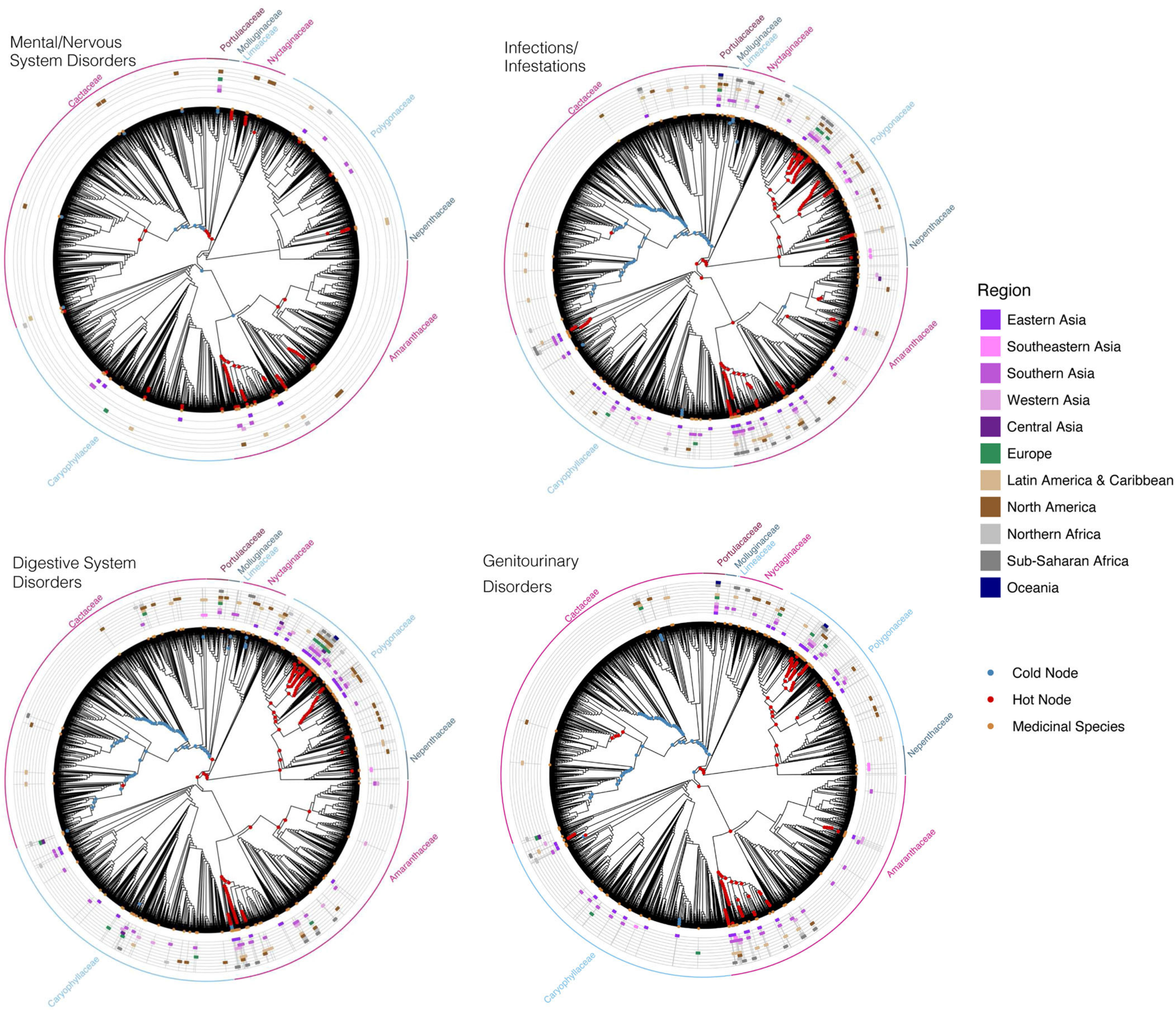

FIGURE 2 Medicinal use in four major categories in representative families of Caryophyllales across the globe. Red nodes indicate clades where medicinal use is phylogenetically overrepresented in that category. Gray nodes indicate clades where medicinal use is underrepresented. The geographical region(s) where a medicinal species is used in that category is given in concentric rings outside the phylogeny. The outermost ring of magenta and blue indicates clades that are either anthocyanin or betalain-producing. The top three medicinal categories are shown here, as well as the Mental/Nervous System Category, to illustrate its pattern of little recorded use as discussed in the text and potential tyrosine-specific pattern. The remaining, less abundant medicinal categories show a similar pattern of hot nodes in Polygonaceae and Amaranthaceae and cold nodes in Cactaceae (with the exception of Endocrine System Disorders; Appendix S2).

variety of human cultures are more likely to be selected for medicinal use across categories than the species less apparent to humans. Here, the apparency of a plant species is a measure of how common a plant species is and how easy it is for humans to notice and interact with it. Higher apparency plants being more likely used medicinally is supported by our analysis restricted to North America but including all Caryophyllales families.

Because North America has less of an emphasis on medicinal plant research than countries like India and China, GBIF data from North America has less bias toward medicinal plants being observed specifically because they are medicinal. Therefore, the North American analysis allows us to use occurrence data as a proxy for apparency without introducing a bias of species being observed specifically because they are medicinal. Additionally, with its simultaneous expansion of families considered and restriction of species to a specific region, the North American analysis supports the patterns of medicinal use observed in the global analysis being grounded in actual patterns of human use, rather than geographic or taxonomic selection bias. While hot nodes are less consistent across medicinal use categories in North America, which is not surprising given the reduced number of species included on the tree and used medicinally,

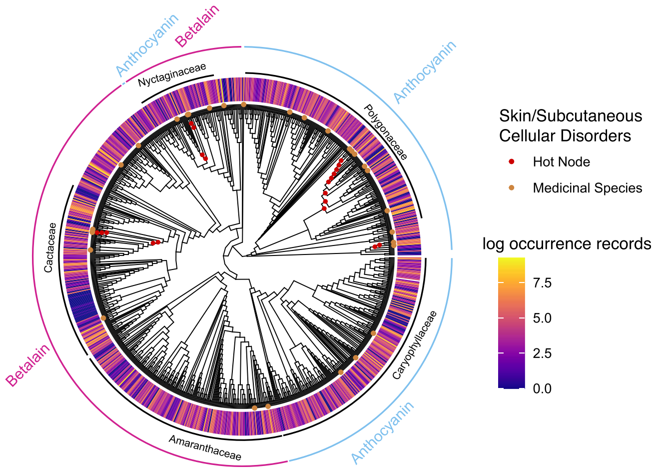

FIGURE 3 A hot nodes analysis of medicinal use of Caryophyllales species in North America for skin/subcutaneous cellular tissue disorders. The ring outside the phylogeny represents log-transformed occurrences from GBIF.

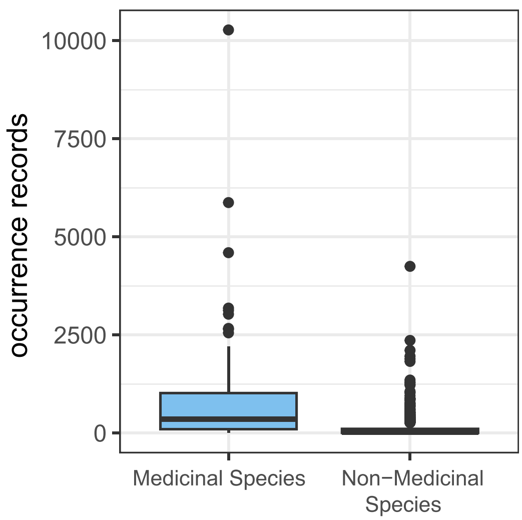

FIGURE 4 Boxplot of number of occurrence data from GBIF of North American Caryophyllales species with any type of medicinal use vs. those with no reported medicinal use. The numbers of occurrence records on GBIF are plotted on the *y*-axis, with the assumption that a greater number of records represents greater apparency. Species were categorized as either medicinal or non-medicinal.

the same clades in Polygonaceae and Amaranthaceae tend to either have hot nodes or be one of the few clades with medicinal use of any kind. Nyctaginaceae, now a larger proportion of the tree than in the global analysis, still has relatively few and scattered hot nodes. The preference of certain clades across North American cultures, the attribution of many medicinal uses across categories to the same clades, and the significant correlation between occurrence data and whether a plant is used medicinally supports our hypothesis that in Caryophyllales, apparency is strongly associated with medicinal plant selection.

The availability and resource availability hypotheses

Apparency being strongly associated with medicinal use is consistent with the availability hypothesis that accessible species, often growing in disturbed areas, are more likely to be selected for medicinal use due to frequent human contact (Stepp and Moerman, 2001; Voeks, 2004; Gaoue et al., 2017). Some of these species spread with humans as they move across the globe, increasing the likelihood of contact with different cultural groups, and helping them acquire uses in new categories (Chapman et al., 2017). The availability hypothesis also explains the lack of use for Cactaceae, which is mostly restricted to North and South America, with the exception of several widely introduced species (i.e., *Opuntia* spp.). Tellingly, the species of Cactaceae used in the most categories (15) is *Opuntia ficus-indica*, which also is used in the most regions (6). Additionally, as Stepp and Moerman (2001) point out, humans are more likely to seek out medicines that are abundant and easily acquired when necessary, rather than relying on uncommon or difficult-to-access plants. While the availability hypothesis argues that underlying metabolism is not the primary factor underlying human selection for medicinal use, there may be other factors to consider.

There is prior evidence to suggest that faster-growing and shorter-lived plant species tend to have more quickly made, bioactive, and therefore potentially medicinal, compounds rather than investing in longer-lasting, but slower and costly to produce defenses such as lignins and tannins (Endara and Coley, 2011). This pattern of metabolite accumulation is in line with the resource availability hypothesis, which posits that plants adapted to high-resource areas, such as human-disturbed areas, tend to be quick-growing and therefore put more resources into "quick" herbivory defenses, like bioactive specialized metabolites, rather than "slow" structural defenses that reduce the plant's digestibility (Endara and Coley, 2011; Gaoue et al., 2017). In the context of this study, the resource availability hypothesis suggests the more apparent species get used because they truly do tend to produce more medicinal compounds than their relatives growing in more resource-scarce areas. The resource availability hypothesis would also explain the relative lack of medicinal uses in Cactaceae, a comparatively slow-growing group that may invest more in structural defenses, such as spines and glochids, than bioactive SMs. The patterns of medicinal use may also be due to a combination of both apparency and SM, where apparent species have evolved to have more SMs to deter herbivory, and also are more available to be selected for medicinal use. It should also be noted that there are examples outside of Caryophyllales of plants with limited apparency or a slow growth rate having significant medicinal effects, such as the tropical montane trees of the *Cinchona* genus (González-Orozco et al., 2023). Therefore, further exploration of the relationship among medicinal use, apparency, and herbivory defense in Caryophyllales is needed.

While statistical power is limited in medicinal categories with few uses, and it is difficult to separate the two potential drivers of selection, there may still be some signal of metabolite-based selection under the overwhelming signal of apparency-driven selection. It is compelling to note that global analysis of the Mental/Nervous System Disorders category has few uses in Polygonaceae (an anthocyanin-producing family), and the only two hot nodes in the family occur toward the tips (Figure 2). Although a few of the other small categories (i.e., Sensory System Disorders and Pregnancy/Birth/Puerperium Disorders) also lack hot nodes in Polygonaceae, they typically also show little clustering in Amaranthaceae (Appendix S3). In the Mental/Nervous System Disorders Category, the Amaranthaceae (a betalain lineage) hot nodes still show clustering. Further, in the North American analysis of Mental and Nervous System Disorders, there is not a single medicinal use in anthocyanin lineages, with most occurring in either Nyctaginaceae, Amaranthaceae, or scattered through Cactaceae. Mental/Nervous System disorders may be generally underrepresented in the medicinal plant tool chest, as Halse-Gramkow et al. (2016) found just over 500 psychoactive species in a review of the entire plant kingdom. Although their literature search was based primarily on a few encyclopedias of known psychoactive plants, and like our literature search here, is not fully comprehensive, 501 species across the plant kingdom is notably small when considering that in just the families selected for our global analysis, there are 165 species used in the Infections/Infestations category. Despite the limited uses reported for these categories, the even more limited use in anthocyanin lineages supports the predicted pattern of more use in lineages with Tyr-dominant metabolism. Halse-Gramkow et al. (2016) also recovered hot nodes in Cactaceae in their analysis, although they did not discuss the rest of Caryophyllales. A more targeted exploration of use of Caryophyllales species in this category is needed to definitively distinguish this pattern from statistical noise and the potential bias of there being more betalain species than anthocyanin species on the global and the North American trees in this study.

Caveats

The ethnobotanical record available in the literature is likely incomplete, and a substantial body of work does not exist on PubMed, Google Scholar, and our queried databases, such as non-digitized books. as evidenced from the somewhat biased geographic patterns observed in our data. The most reported uses we recovered come from East Asia, where traditional Chinese medicine has been actively studied and documented; whereas the most medicinal species come from North America, which has an extensive database of traditional use (Moerman, 2003). It is possible that we might start to see differences emerge in medicinal categories if more ethnobotanical research were available using our search approach in less-studied areas such as Oceania, which has native and introduced species from several of the families reviewed here, but only three medicinal species (*Rumex maderensis* Lowe, *Portulaca oleracea* L., and *Amaranthus viridis* L.) recovered in our literature search. However, the fact that the hot nodes we recovered are often at the base of clades with wide geographic distributions suggests that, with additional documentation of medicinal use in some regions, these nodes would still be consistently hot across categories. Additionally, the North American analysis, while having hot nodes that are less consistent across medicinal use categories, did not show a radically different placement pattern for hot nodes when present compared to the global analysis.

The correlation between a SM and medicinal use is also likely not a linear one. Potential interactions of multiple plants in a medicinal recipe or the use of one plant to boost another's effect cannot be discounted, as interactions could contribute to a medicinal effect in a way individual species would not be reported in literature. For example, interactions between *Psychotria viridis* Ruiz & Pav. (Rubiaceae) and *Banisteriopsis caapi* (Spruce ex Griseb.) C.V. Morton (Malpighiaceae) in a common ayahuasca recipe led to its psychoactive effect. The active compound from *P. viridis* would not be able to reach the central nervous system without SMs from *B. caapi* protecting it from degradation, a relationship that has only been understood after extensive investigation of these plants' SMs (Gambelunghe et al., 2008). It is also possible that not all plant species in the indigenous pharmacopeia have a pharmacological effect matching their reported uses. Evidence of the efficacy of some common herbal medicines can vary by study, with some studies finding little to no effect for their reported uses (Bent and Ko, 2004; Lin et al., 2022).

Another caveat is phylogenetic uncertainty in the guide tree used, especially deep relationships in Cactaceae (Appendix S4). Phylocom's nodesig command is unable to account for either branch length or phylogenetic uncertainty, so hot node estimation may be biased by a poorly supported tree. However, given the overall limited medicinal use reported in Cactaceae and that 38.7% of the family's medicinal species occur in one genus, *Opuntia*, the lack of hot nodes in the family is unlikely to be an artifact from phylogenetic uncertainty in the family.

Additionally, occurrence data from GBIF is an imperfect measure of plant apparency. The likelihood of humans encountering and noticing a plant is difficult to measure directly, but occurrence data by its very nature serves as a proxy because it requires a human to interact with the plant in question. In fact, one of the well-known biases in GBIF and occurrence data in general is a bias toward disturbed or urban areas, while remote locations remain undercollected (Petersen et al., 2021; Bowler et al., 2022). While being an issue in some other uses of occurrence data, in our approach, this bias works in our favor to indicate apparency. However, geographical, temporal, and taxonomic biases also exist in GBIF, and while also probably correlated with apparency, they can also be driven by other factors, such as the research interests of data contributors (Meyer et al., 2016). Additionally, GBIF data, as a massive data set with many contributors, is known to have curation issues (Zizka et al., 2020). While we did take steps to clean questionable datapoints from our analysis, data points that were misidentified or incorrectly georeferenced may remain. However, as our approach for estimating appearance relies on counts of records within a chosen region, we do not require as high-quality georeferencing as some other uses for GBIF data, such as species distribution modeling, which requires precise location data.

Apart from methodological caveats, the weak association between medicinal use category and the dominant metabolism type may alternatively be driven by overlap between the bioactivity of the metabolites derived from the two related amino acids. Phenylalanine-derived specialized metabolism includes phenolics like flavonoids, coumarins, and lignans, which have been ascribed a wide range of potential health benefits. While tyrosine-derived metabolites tend to be less ubiquitous, some of the best-known include neurologically active compounds like mescaline, dopamine, and morphine (Xu et al., 2019). However, other Tyr-derived compounds may have as broad action as Phe metabolism. For example, apart from neurological effect, some Tyr-derived catecholamines like epinephrine act on the cardiovascular and respiratory system (Abul-Ainine and Luyt, 2002; Bao et al., 2007). If Tyr- and Phe-derived metabolism overall have enough overlap in medicinal activity, human selection of apparent plants would be mainly driven by apparency and less by the dominant metabolism type.

Alternatively, metabolites derived from precursors other than Tyr or Phe may drive broad patterns of medicinal activity. At least one species in Caryophyllales is best known for specialized metabolites from a different pathway: *Simmondsia chinensis*, used largely in the Skin/Subcutaneous Cellular Tissue Disorders category is famous for the lipids in its oil. However, given the overall importance of the phenylalanine-derived phenylpropanoid pathway in plant specialized metabolism across the plant kingdom, broad-scale differences in medicinal use between lineages would still be expected to be affected by the switches between Tyr- and Phe-dominant metabolism.

CONCLUSIONS

Given the renewed interest in natural products for drug discovery, understanding the relationships among specialized metabolism, phylogeny, and medicinal use is important in prioritizing plants to investigate and narrow down the metabolites responsible for medicinal function. In addition, an approach combining phylogenetic, metabolic, and global use information can elucidate cross-cultural patterns of medicinal use and what is driving them. Here, we show that, at least in Caryophyllales, the apparency of a plant to humans is a dominant factor in its being selected for medicinal use across cultures, as opposed to the plant's broad SM profile. Although there are several possible explanations for this, a combination of humans selecting readily apparent plants (availability hypothesis) and those same plants having greater bioactive activity (resource availability hypothesis) appear to be the most likely. Further investigation of the Tyr- and Phe-derived metabolic diversity in both medicinal and non-medicinal clades of the order would help determine which or both of these hypotheses is driving the selection of medicinal plants in Caryophyllales and help untangle the effects of apparency from other potential drivers of medicinal plant selection.

AUTHOR CONTRIBUTIONS

A.C. conceived the research, collected and analyzed data, and wrote the manuscript. L.P. contributed to the design of the research and manuscript revision. L.B. wrote scripts for data collection and contributed to manuscript revision. Y.Y. aided in conception, design, and manuscript revision.

ACKNOWLEDGMENTS

We thank M. Mohamed and L. Magpali for their assistance in data collection and A. Lee, R.A. Mohn, A. Smeenk, and two anonymous reviewers for their feedback on the work. Funding was provided by the National Science Foundation (NSFDEB-NERC 1939226).

DATA AVAILABILITY STATEMENT

Data and scripts used in this analysis are deposited on Dryad: https://doi.org/10.5061/dryad.6q573n64z (Crum, 2024).

ORCID

Alex H. Crum http://orcid.org/0009-0004-5506-9491
Lucas Busta http://orcid.org/0000-0002-0102-9986
Ya Yang http://orcid.org/0000-0001-6221-0984

REFERENCES

Abul-Ainine, A., and D. Luyt. 2002. Short term effects of adrenaline in bronchiolitis: a randomized controlled trial. *Archives of Disease in Childhood* 86: 276–279.

Afendi, F. M, T. Okada, M. Yamakazi, A. Hirai-Morita, Y. Nakamura, K, Nakamura, S. Ikeda, et al. 2012. KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology* 53: e1.

Atanasov, A. G., S. B. Zotchev, V. M. Dirsch, the International Natural Product Sciences Taskforce, and C. T. Supuran. 2021. Natural products in drug discovery: advances and opportunities. *Nature Revies Drug Discovery* 20: 200–216.

Bao, X., C. M. Lu, F. Liu, Y. Gu, N. D. Dalton, B. Q. Zhu, E. Foster, et al. 2007. Epinephrine is required for normal cardiovascular responses to stress in the phenylethanolamine N-methyltransferase knockout mouse. *Circulation* 116: 1024–1031.

Bent, S., and R. Ko. 2004. Commonly used herbal medicines in the United States: a review. *American Journal of Medicine* 116: 478–485.

Berman, A. Y., R. A. Motechin, M. Y. Wiesenfeld, and M. K. Holz. 2017. The therapeutic potential of resveratrol: a review of clinical trials. *NPJ Precision Oncology* 1: 35.

Bowler, D. E., C. T. Callaghan, N. Bhandari, K. Henle, M. B. Barth, C. Koppitz, R. Klenke, et al. 2022. Temporal trends in the spatial bias of species occurrence records. *Ecography* 2022: e06219.

Burke, J. M., and A. Sanchez. 2011. Revised subfamily classification for Polygonaceae, with a tribal classification for Eriogonoideae. *Brittonia* 63: 510–520.

Casarini, T. P. A., L. A. Frank, A. R. Pohlmann, and S. S. Guterres. 2020. Dermatological applications of the flavonoid phloretin. *European Journal of Pharmacology* 889: 173593.

Chapman, D., B. V. Purse, H. E. Roy, and J. M. Bullock. 2017. Global trade networks determine the distribution of invasive non-native species. *Global Ecology and Biogeography* 26: 907–917.

Chen, J., Y.-P. Shi, and J.-Y. Liu. 2003. Determination of noradrenaline and dopamine in Chinese herbal extracts from *Portulaca oleracea* L. by high-performance liquid chromatography. *Journal of Chromatography A* 1003: 127–132.

Clement, J. S., and T. J. Mabry. 1996. Pigment evolution in the Caryophyllales: a systematic overview. *Botanica Acta* 109: 360–367.

Cook, F. E. M. 1995. Economic botany data collection standard. Royal Botanic Gardens, Kew, UK.

Crum, A. 2024. Traditional medicinal use is linked with apparency, not specialized metabolite profiles in the order Caryophyllales (Version 6) [Data set]. Dryad. https://doi.org/10.5061/DRYAD.6Q573N64Z

Endara, M. J., and P. D. Coley. 2011. The resource availability hypothesis revisited: A meta-analysis. *Functional Ecology* 25: 389–398.

Ernst, M., C. H. Saslis-Lagoudakis, O. M. Grace, N. Nilsson, H. T. Simonsen, J. W. Horn, and N. Rønsted. 2016. Evolutionary prediction of medicinal properties in the genus *Euphorbia* L. *Scientific Reports* 6: 30531.

Gambelunghe, C., K. Aroni, R. Rossi, L. Moretti, and M. Bacci. 2008. Identification of *N,N*-dimethyltryptamine and β-carbolines in psychotropic ayahuasca beverage. *Biomedical Chromatography* 22: 1056–1059.

Gaoue, O. G., M. A. Coe, M. Bond, G. Hart, B. C. Seyler, H. McMillen. 2017. Theories and Major Hypotheses in Ethnobotany. *Economic Botany* 71: 269–287.

Gaoue, O. G., K. Yessoufou, L. Mankga, and F. Vodouhe. 2021. Phylogeny reveals non-random medicinal plant organ selection by local people in Benin. *Plants People Planet* 3: 710–720.

González-Orozco, C. E., E. G. Guillén, and N. Cuvi. 2023. Changes of *Cinchona* distribution over the past two centuries in the northern Andes. *Royal Society Open Science* 10: 230229.

Halse-Gramkow, M., M. Ernst, N. Rønsted, R. R. Dunn, and C. H. Saslis-Lagoudakis. 2016. Using evolutionary tools to search for novel psychoactive plants. *Plant Genetic Resources: Characterization and Utilization* 14: 246–255.

Hamilton, A. C. 2004. Medicinal plants, conservation and livelihoods. *Biodiversity and Conservation* 13:1477–1517.

Holeski, L. M., K. Keefover-Ring, J. M. Sobel, and N. J. Kooyers. 2021. Evolutionary history and ecology shape the diversity and abundance of phytochemical arsenals across monkeyflowers. *Journal of Evolutionary Biology* 34: 571–583.

Huang R., A. J. O'Donnell, J. J. Barboline, and T. J. Barkman. 2016. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proceedings of the National Academy of Sciences, USA* 113: 10613–10618.

Jain, G., and K. S. Gould. 2015. Are betalain pigments the functional homologues of anthocyanins in plants? *Environmental and Experimental Botany* 119: 48–53.

Korotkova, N., D. Aquino, S. Arias, U. Eggli, A. Franck, C. Gómez-Hinostrosa, P. C. Guerrero, et al. 2021. Cactaceae at Caryophyllales.org – a dynamic online species-level taxonomic backbone for the family. *Willdenowia* 51: 251–270.

Landi, M., M. Tattini, and K. S. Gould. 2015. Multiple functional roles of anthocyanins in plant–environment interactions. *Environmental and Experimental Botany* 119: 4–17.

Lewu, F. B., and A. J. Afolayan. 2009. Ethnomedicine in South Africa: the role of weedy species. *African Journal of Biotechnology* 8: 929–934.

Lin, C.-H., Y.-A. Lin, S.-L. Chen, M.-C. Hsu, and C.-C. Hsu. 2022. American ginseng attenuates eccentric exercise-induced muscle damage via the modulation of lipid peroxidation and inflammatory adaptation in males. *Nutrients* 14: 78.

Lopez-Nieves, S., Y. Yang, A. Timoneda, M. Wang, T. Feng, S. A. Smith, S. F. Brockington, and H. A. Maeda. 2018. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. *New Phytologist* 217: 896–908.

Meyer, C., P. Weigelt, and H. Kreft. 2016. Multidimensional biases, gaps, and uncertainties in global plant occurrence information. *Ecology Letters* 19: 992–1006.

Moerman, D. E. 2003. Native American Ethnobotany: A database of foods, drugs, dyes and fibers of Native American peoples derived plants [online]. University of Michigan, Dearborn, MI, USA. Website: http://naeb.brit.org/

Orme, D., R. P. Freckleton, G. H. Thomas, T. Petzoldt, S. Fritz, N. Isaac, and W. Pearse. 2023. caper: Comparative analyses of phylogenetics and evolution in R, version 1.0.2. Website: https://cran.r-project.org/web/packages/caper/index.html

Paradis, E., and K. Schliep. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.

Pellicer, J., C. H. Saslis-Lagoudakis, E. Carrió, M. Ernst, T. Garnatje, O. M. Grace, A. Gras, et al. 2018. A phylogenetic roadmap to antimalarial *Artemisia* species. *Journal of Ethnopharmacology* 225: 1–9.

Petersen, T. K., J. D. M. Speed, V. Grøtan, and G. Austrheim. 2021. Species data for understanding biodiversity dynamics: The what, where, and when of species occurrence data collection. *Ecological Solutions and Evidence* 2: e12048.

POWO. 2023. Plants of the world [online]. Royal Botanical Gardens, Kew, UK. Website: http://www.plantsoftheworldonline.org/

Rodman, J. E., P. S. Soltis, D. E. Soltis, K. J. Sytsma, and K. G. Karol. 1998. Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *American Journal of Botany* 85: 997–1006.

Saslis-Lagoudakis, C. H., B. B. Klitgaard, F. Forest, L. Francis, V. Savolainen, E. M. Williamson, and J. A. Hawkins. 2011. The use of phylogeny to interpret cross-cultural patterns in plant use and guide medicinal plant discovery: an example from *Pterocarpus* (Leguminosae). *PLoS One* 6: e22275.

Saslis-Lagoudakis, C. H., V. Savolainen, E. M. Williamson, F. Forest, S. J. Wagstaff, S. R. Baral, M. F. Watson, et al. 2012. Phylogenies reveal predictive power of traditional medicine in bioprospecting. *Proceedings of the National Academy of Sciences, USA* 109: 15835–15840.

Sheehan, H., T. Feng, N. Walker-Hale, S. Lopez-Nieves, B. Pucker, R. Guo, W. C. Yim, et al. 2020. Evolution of L-DOPA 4,5-dioxygenase activity allows for recurrent specialisation to betalain pigmentation in Caryophyllales. *New Phytologist* 227: 914–929.

Smith, S. A., J. W. Brown, Y. Yang, R. Bruenn, S. F. Brockington, J. F. Walker, N. Last, et al. 2018. Disparity, diversity, and duplications in the Caryophyllales. *New Phytologist* 217: 836–854.

Stepp, J. R., and D. E. Moerman. 2001. The importance of weeds in ethnopharmacology. *Journal of Ethnopharmacology* 75: 19–23.

Sun, W., and M. H. Shahrajabian. 2023. Therapeutic potential of phenolic compounds in medicinal plants – Natural health products for human health. *Molecules* 28: 1845.

United Nations Statistics Division. 1999. Standard country or area codes for statistics use [online version]. Revision 4. Statistics Division, United Nations, NY, NY, USA.

Vamvakopoulou, I. A., K. A. D. Narine, I. Campbell, J. R. B. Dyck, and D. J. Nutt. 2023. Mescaline: The forgotten psychedelic. *Neuropharmacology* 222: 109294.

Voeks, R. A. 2004. Disturbance pharmacopeias: medicine and myth from the humid tropics. *Annals of the Association of American Geographers* 94: 868–888.

Webb, C. O., D. D. Ackerly, and S. W. Kembel. 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* 24: 2098–2100.

Wink, M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* 64: 3–19.

Winter, D. J. 2017. Rentrez: an R package for the NCBI eUtils API. *R Journal* 9: 520–526.

Xu, S., Z. Dai, P. Guo, X. Fu, S. Liu, L. Zhou, W. Tang, et al. 2021. ggtreeExtra: Compact visualization of richly annotated phylogenetic data. *Molecular Biology and Evolution* 38: 4039–4042.

Xu, J. J., X. Fang, C. Y. Li, L. Yang, and X. Y. Chen. 2019. General and specialized tyrosine metabolism pathways in plants. *aBIOTECH* 1: 97–105.

Yang, J., Z. Li, J. Lian, G. Qi, P. Shi, J. He, Z. Hu, and M. Zhang. 2020. Brassicaceae transcriptomes reveal convergent evolution of super-accumulation of sinigrin. *Communications Biology* 3: 779.

Youssef, D., R. El-Bakatoushi, A. Elframawy, L. El-Sadek, and G. El Badan. 2023. Molecular phylogenetic study of flavonoids in medicinal plants: a case study of family Apiaceae. *Journal of Plant Research* 136: 305–322.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T. Lam. 2017. GGTree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8: 28–36.

Zaman, W., J. Ye, S. Saqib, Y. Liu, Z. Shan, D. Hao, Z. Chen, and P. Xiao. 2021. Predicting potential medicinal plants with phylogenetic topology: Inspiration from the research of traditional Chinese medicine. *Journal of Ethnopharmacology* 281: 114515.

Zhong, Z., J. Han, J. Zhang, Q. Xiao, J. Hu, and L. Chen. 2018. Pharmacological activities, mechanisms of action, and safety of salidroside in the central nervous system. *Drug Design, Development, and Therapy* 12: 1479–1489.

Zizka, A., F. A. Carvalho, A. Calvente, M. R. Baez-Lizarazo, A. Cabral, J. F. R. Coelho, M. Colli-Silva, et al. 2020. No one-size-fits-all solution to clean GBIF. *PeerJ* 8: e9916.

Zizka, A., D. Silvestro, T. Andermann, J. Azevedo, C. D. Ritter, D. Edler, H. Farooq, et al. 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* 10: 744–751.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Tables of number of records for medicinal uses by category of use, geographic region, and taxonomic family. Records from a global literature search of selected families in Caryophyllales.

Table S1. Medicinal uses by category.

Table S2. Medicinal uses by region.

Table S3. Medicinal uses by family.

Appendix S2. Hot and cold nodes mapped out on a phylogeny of selected Caryophyllales families using medicinal use data across the globe. All medicinal categories not part of Figure 2 are shown. The color of family names follows Figure 1.

Appendix S3. Hot and cold nodes mapped out on a phylogeny of North American Caryophyllales species. A selection of medicinal categories is included.

Appendix S4. Guide tree used in the global analysis with bootstrap support plotted at each node.

How to cite this article: Crum, A. H., L. Philander, L. Busta, and Y. Yang. 2024. Traditional medicinal use is linked with apparency, not specialized metabolite profiles in the order Caryophyllales. *American Journal of Botany* 111(4): e16308. https://doi.org/10.1002/ajb2.16308