

METHODS AND TECHNIQUES

Hot node limitations and impact of taxonomic resolution on phylogenetic divergence patterns: A case study on Ecuadorian ethnomedicinal flora

Darío Atienza-Barthelemy¹  | Manuel J. Macía^{1,2}  | Rafael Molina-Venegas^{2,3} 

¹Departamento de Biología, Área de Botánica, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, Spain

²Centro de Investigación en Biodiversidad y Cambio Global (CIBC-UAM), Universidad Autónoma de Madrid, Madrid, Spain

³Departamento de Ecología, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, Spain

Correspondence

Rafael Molina-Venegas, Centro de Investigación en Biodiversidad y Cambio Global (CIBC-UAM), Universidad Autónoma de Madrid, Madrid, 28049, Spain.
Email: rafael.molina@uam.es

Funding information

Spanish Ministry of Science, Innovation, and Universities, Grant/Award Number: PID2022-138776NA-I00; European Union (NextGenerationEU); Universidad Autónoma de Madrid, Grant/Award Number: contract No. 39

Summary**Societal Impact Statement**

Novel phylogenetic approaches have emerged in recent years to study traditional medicinal plants, aiming to identify potential sources of new drugs. This line of research holds considerable promise yet remains in its early stages. Here, we examine the prevalent methods employed in the field, revealing the impact of methodological choices that have often been made arbitrarily. We also highlight a widespread misconception regarding ‘hot node analysis’, a tool gaining popularity for identifying plants with high bioactive potential. Our findings should advance future research aimed at guiding the selection of promising candidates for bioprospection.

Summary

- Advances in phylogenetics offer a ground-breaking approach to analysing ethnobotanical data. This line of research typically involves calculating the degree of phylogenetic clustering (i.e. phylogenetic divergence) of a set of medicinal plants and identifying clades with a significant overabundance of these plants, known as hot nodes, which are purportedly responsible for the clustering patterns. However, despite showing great promise, the adequacy of this procedure remains to be tested, and the results have so far been inconclusive and, at times, contradictory.
- Here, we examine two key elements through a case study of Ecuadorian medicinal flora: the impact of taxonomic resolution (species- and genus-level data) on phylogenetic divergence metrics (Mean Pairwise Distance; MPD, and Mean Nearest Taxon Distance; MNTD), and the efficacy of the hot node analysis in identifying clades that significantly influence these metrics. To identify clades with a significant impact on MPD and MNTD, we implemented a jackknifing procedure and compared outcomes with the hot nodes.
- Phylogenetic divergence was strongly dependent on the taxonomic resolution, with clustering mainly revealed by MPD at the genus level. Further, the hot nodes incompletely matched those with significant influence on MPD and MNTD according to the jackknifing analysis.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Plants, People, Planet* published by John Wiley & Sons Ltd on behalf of New Phytologist Foundation.

- We highlight the impact of taxonomic resolution on commonly used phylogenetic divergence metrics and the limitations of hot nodes in identifying influential clades, stressing the potential of alternative jackknifing techniques. We recommend the use of the most resolved tree possible and combining jackknifing with hot node analyses as complementary sources to pinpoint clades for bioprospection.

KEYWORDS

bioprospecting, eco-phylogenetics, ethnobotany, medicinal plants, phylogenetic methods, phylogenetic structure, traditional use

1 | INTRODUCTION

Over the past few centuries, traditional knowledge of medicinal plants has played a major role in the discovery of globally commercialised drugs, such as artemisinin, acetylsalicylic acid, ephedrine and camptothecin, among others (Khaiwa et al., 2021; Miao et al., 2020; Montinari et al., 2019; Tu, 2011). Despite this historical success, researchers still face challenges to prove the benefits of exploring traditional medicine for phytochemical potential (Atanasov et al., 2021; Choudhury et al., 2020; Gyllenhaal et al., 2012; Kingston, 2011; Wangkheirakpam, 2018).

Recent advances in phylogenetics have provided a revolutionary way to exploit ethnomedicinal data. This innovative approach typically involves exploring the degree of phylogenetic divergence of a set of medicinal plants, with the premise that phylogenetic clustering may indicate underlying medicinal properties (Rønsted et al., 2012; Salsis-Lagoudakis et al., 2012). To calculate phylogenetic divergence, two distance-based metrics borrowed from community phylogenetics are most frequently used: Mean Pairwise Distance (MPD) and Mean Nearest Taxon Distance (MNTD). These metrics measure phylogenetic distances among ethnomedicinal taxa, indicating significant clustering or overdispersion when the distances are shorter or longer than expected by chance. Despite showing great promise, results have thus far proved inconclusive, with medicinal plants revealing highly variable levels of phylogenetic clustering and, at times, even contradictory, as studies on the same ethnoflora have produced different outcomes (Salsis-Lagoudakis et al., 2012; Yessoufou et al., 2015). These inconsistencies have been attributed to various factors, including the ethnomedicinal classification system (Ernst et al., 2016; Yessoufou et al., 2015) and the size of the study pool (Yessoufou et al., 2015; Zaman et al., 2021). Nonetheless, the potential impact of taxonomic resolution has been largely overlooked, with previous studies inconsistently employing either species- (Ernst et al., 2016; Salsis-Lagoudakis et al., 2011; Zaman et al., 2021; Zaman et al., 2022) or genus-level data (Halse-Gramkow et al., 2016; Lei et al., 2020; Salsis-Lagoudakis et al., 2012; Souza et al., 2018; Thompson & Hawkins, 2024; Yessoufou et al., 2015).

Phylogenetic divergence analyses are often coupled with the identification of those clades showing a significant overabundance of medicinal plants, commonly referred to as 'hot nodes' (Ernst et al., 2016; Halse-Gramkow et al., 2016; Salsis-Lagoudakis et al., 2012; Souza et al., 2018; Thompson & Hawkins, 2024; Zaman

et al., 2021; Zaman et al., 2022). Whenever phylogenetic divergence metrics indicate that a given medicinal use is significantly clustered in the phylogeny, the hot node technique is applied to locate the specific clades responsible for such clustering (Salsis-Lagoudakis et al., 2011), as species within these clades, including those with no recorded use, are purportedly likely to share active compounds. In words, hot nodes are assumed to be the main contributors to phylogenetic clustering (Letcher, 2010; Pellicer et al., 2018; Salsis-Lagoudakis et al., 2012; Souza-Neto et al., 2016; Villalobos et al., 2019; Webb et al., 2008). However, these methods rely on different principles: phylogenetic divergence metrics are based on phylogenetic distances (i.e. branch lengths), whereas the hot nodes depend exclusively on tree topology. Therefore, their presumed association lacks quantitative support.

Ecuador is home to a remarkable biological and cultural diversity that thrives on a strong interdependent relationship established between humans and nature (de la Torre et al., 2012). Along with the absence of an effective healthcare system, particularly in impoverished rural areas, the extent of ancestral knowledge of medicinal plants explains their widespread use (de la Torre et al., 2008). The country hosts over 18,000 species of vascular plants, nearly one-third of which are endemic (POWO, 2022). Among this outstanding floristic richness, more than 5,000 species have documented uses, with almost 2,000 being used for medicinal purposes (de la Torre et al., 2008). This vast reservoir of ethnomedicinal knowledge presents an exceptional opportunity for the exploration of potentially bioactive plants.

Here, we use an extensive database of Ecuadorian traditional medicine to assess two key elements of applied phylogenetics. Firstly, we examine the impact of taxonomic resolution (species- and genus-level data) on phylogenetic divergence metrics. Subsequently, we perform a jackknifing analysis to identify clades that significantly influence phylogenetic divergence patterns (Davies et al., 2016), and test whether hot nodes overlap with influential clades.

2 | MATERIALS AND METHODS

2.1 | Ethnomedicinal data and flora of Ecuador

We compiled data from the *Encyclopaedia of the Useful Plants of Ecuador* (de la Torre et al., 2008), which includes 16,219 records of medicinal uses. Several modifications were applied to this data. Firstly, we lumped infraspecific records ($n = 184$) into corresponding

species and removed those limited to genus or family level ($n = 3,315$). This left us with 12,904 species-level records. Secondly, non-native medicinal species ($n = 336$) were excluded to focus on plants prospected by Ecuadorian communities throughout history. Native status was determined following POWO (2022) and, for omission errors in this source, the *Catalogue of the Vascular Plants of Ecuador* (Jørgensen & León-Yáñez, 1999) and GBIF (2022) were consulted. Finally, given that phylogenetic divergence can substantially change with phylogenetic extent (i.e. taxonomic breadth of a phylogenetic analysis; Graham et al., 2018), we focused on the angiosperm clade since it included most medicinal species, hence excluding gymnosperms ($n = 5$) and pteridophytes ($n = 132$).

Medicinal uses were classified into 22 categories (Notes S1) following the *Economic Botany Data Collection Standard* (Cook, 1995). This categorisation resulted in a presence-absence ‘community matrix’ with medicinal categories in rows and species in columns. In addition, we created an extra category by merging all medicinal categories together (hereafter ‘whole medicinal flora’). For the genus-level analysis, the data were combined into a single column per genus.

Analysing the phylogenetic divergence of a group of species requires a reference pool, which in our case included all angiosperms native to Ecuador. We extracted this checklist from POWO (2022) and harmonised taxonomic names with those of our medicinal dataset using WFO (2022) criteria through the ‘WorldFlora’ R package (Kindt, 2020). This resulted in a total of 15,757 species and 2,032 genera, of which 11% ($n = 1,803$) and 36% ($n = 736$) have medicinal applications, respectively.

2.2 | Phylogenetic reconstruction

Species-level phylogenies of Ecuador's native angiosperms were generated using ‘V.PhyloMaker’ (Jin & Qian, 2019), an R package that provides a time-calibrated molecular mega-tree of DNA encompassing 74,533 species of vascular plants. To address missing species (phylogenetic uncertain taxa; PUTs), we opted to randomly insert them at and below the crown node representing the most recent common ancestor of their genus, or family in cases where the former was unavailable (scenario 2 in the *phylo.maker* function). For polyphyletic groups, PUTs were placed as described above but in the largest monophyletic subgroup of congeners (*build.nodes1* in the *phylo.maker* function). Once PUTs were bound, the mega-tree was pruned to the angiosperm native flora of Ecuador. PUTs represented 74% of the species pool and 54% of the medicinal plants, which was to be expected, considering that the Neotropical Region is underrepresented in DNA repositories due to its inaccessibility and high biodiversity. To address this phylogenetic uncertainty, we iteratively replicated the procedure described above to generate $n = 100$ phylogenies, conducting analyses and averaging results across all trees (Molina-Venegas et al., 2022; Rangel et al., 2015). Genus-level phylogenies were obtained by randomly selecting one species per genus from the species-level trees and pruning the remaining tips.

2.3 | Phylogenetic divergence analysis

We used the MPD and MNTD metrics to compute the phylogenetic divergence of each medicinal category. MPD represents the average distance among all taxa pairs in a sample, while MNTD measures the average distance from each taxon to its closest relative (Webb et al., 2002). Here, non-parametric p -values were calculated to test for significant clustering or overdispersion (two-tailed test) using the formula:

$$p\text{-value} = \text{obs.rank}/\text{runs} + 1, \quad (1)$$

where *obs.rank* is the observed value of the metric within a null distribution of values obtained by shuffling taxa labels in the community matrix $n = 999$ times (runs). All computations were conducted using the functions *ses.mpd* and *ses.mntd* of ‘Picante’ R package (Kembel et al., 2010). For a 5% nominal alpha, p -values below 0.025 and above 0.975 indicate significant clustering and overdispersion, respectively. Calculations were performed for each species- and genus-level phylogenetic tree, with results averaged for each metric and taxonomic resolution. Additionally, 95% confidence intervals were computed to assess the impact of phylogenetic uncertainty in the analyses (Molina-Venegas et al., 2022).

2.4 | Identifying clades that significantly influence phylogenetic divergence metrics

We implemented and expanded the jack-knifing method originally described by Davies et al. (2016) to identify clades significantly influencing the phylogenetic divergence metrics MPD and MNTD. Our goal was strictly limited to describing and comparing this method against the hot node technique, so replicating the analyses across all trees and medicinal categories was unnecessary. The procedure was consequently illustrated using a single genus-level tree and the whole medicinal flora, primarily owing to the minimal phylogenetic uncertainty at this level and for ease of visualisation.

The jackknifing analysis was conducted as follows. We systematically pruned all branches descending from each internal node i (excluding the root node), one node at a time and calculated MPD and MNTD among medicinal genera for the remaining tree (hereafter $\text{MPD}_{\text{pruned}}$ and $\text{MNTD}_{\text{pruned}}$). As nodes towards the root tend to subtend more taxa, greater shifts in the metrics are expected for deeper nodes. Therefore, we computed MPD and MNTD after randomly pruning the same number of tips representing medicinal genera as in the previous step, repeating this procedure $n = 999$ times to draw a distribution of $\text{MPD}_{\text{pruned-null}}$ and $\text{MNTD}_{\text{pruned-null}}$ values. Non-parametric p -values were computed following equation 1, where *obs.rank* represented the rank of $\text{MPD}_{\text{pruned}}$ and $\text{MNTD}_{\text{pruned}}$ within their respective null distributions. For a given node i and a 5% nominal alpha, $p > 0.975$ indicates that the medicinal genera descending from i have a significant influence on the corresponding metric by significantly increasing clustering levels. Conversely, $p < 0.025$ indicates that

the medicinal genera descending from i contribute significantly to overdispersion. Hereafter, we will collectively refer to these nodes as ‘jackknifing nodes’, distinguishing between ‘clustering nodes’ ($p > 0.975$) and ‘overdispersion nodes’ ($p < 0.025$). Note that any node must include at least one medicinal genus to contribute minimally to phylogenetic divergence (see Code S1 for R code and further details).

2.5 | Hot node analysis

Using the same genus-level tree as above and the whole medicinal flora, we conducted a hot node analysis to identify clades encompassing a significantly high number of medicinal genera. Following equation 1, we computed a non-parametric p -value for each internal node i (excluding the root node), where $obs.rank$ was the rank of the observed value of the hot node metric (i.e. number of medicinal genera descending from node i) within a null distribution of values obtained by shuffling the binary trait ‘medicinal’ (1) or ‘not medicinal’ (0) across phylogenetic tips $n = 999$ times (runs). Hot nodes were defined as those showing p -values above 0.95 (5% nominal alpha, one-tailed test). The R code provided by Molina-Venegas et al. (2020) was used for this analysis, with modifications made to accommodate the computation of p -values (see Code S2 for R code).

2.6 | Crossover of hot nodes with jackknifing nodes

We conducted Chi-square tests to assess if the proportion of hot nodes overlapping jackknifing nodes deviated significantly from random expectation. For instance, if 20% of all nodes potentially significant in both hot node and jackknifing analyses (i.e. internal nodes including at least one medicinal genus) are identified as clustering nodes, the expectation is that 20% of hot nodes overlap with clustering nodes simply by chance. In total, we conducted four tests, one per phylogenetic divergence metric (MPD or MNTD) and type of jackknifing node (clustering or overdispersion). All p -values were computed by Monte-Carlo simulation using 2000 replicates per test.

3 | RESULTS

3.1 | Phylogenetic divergence of medicinal categories across taxonomic resolutions

The phylogenetic divergence was strongly dependent on the metric and taxonomic resolution (Table 1). When analysing all medicinal plants together (i.e. category ‘whole medicinal flora’) at the genus level, MPD revealed strong clustering and MNTD was non-significant, whereas, at the species level, both MPD and MNTD showed significant overdispersion. Moreover, on the genus-level analysis, over half the use categories demonstrated significant clustering based on MPD,

and most exhibited random distribution according to MNTD, except for ‘symptoms’ and ‘muscular-skeletal system disorders’, which showed overdispersion. Conversely, most categories fit random patterns at the species level, except for ‘skin/subcutaneous cellular tissue disorders’ and ‘circulatory system disorders’, which showed significant clustering according to MPD and MNTD, respectively. Phylogenetic uncertainty had a negligible impact, as statistical significance remained consistent across all phylogenetic trees (Table S1).

3.2 | Influence of hot nodes on phylogenetic divergence metrics

The phylogeny used for the jackknifing and hot node analyses consisted of 2,030 internal nodes, of which 1,561 contained at least one medicinal genus, having therefore the potential to be significant in both analyses. Of these, the jackknifing analysis revealed 485 and 136 nodes significantly influencing MPD and MNTD, respectively (either for clustering or overdispersion), while the hot node analysis identified 124 nodes with a significantly high proportion of medicinal genera (Figures 1 and 2).

Among the 485 jackknifing nodes significantly influencing MPD, 249 and 236 contributed towards clustering and overdispersion, respectively. Therefore, the number of MPD-clustering nodes ($n = 249$) is double that of hot nodes ($n = 124$). Moreover, although the overlap between hot nodes and MPD-clustering nodes significantly exceeded random expectations ($n = 59$; $\chi^2 = 92.528$, $p < 0.001$), 52% of hot nodes did not overlap with MPD-clustering nodes, and the overlapping hot nodes accounted for only 24% of MPD-clustering nodes. The number of MPD-overdispersion nodes ($n = 236$) was also substantially higher than that of hot nodes, and their degree of overlap was undistinguishable from random expectations ($n = 24$; $\chi^2 = 1.7341$, $p = 0.214$).

Among the 136 jackknifing nodes significantly influencing MNTD, 49 pushed towards clustering and 87 towards overdispersion. Although the overlap between hot nodes and MNTD-clustering nodes narrowly exceeded random expectations ($n = 8$; $\chi^2 = 4.4753$, $p = 0.041$), 94% of hot nodes did not overlap with MNTD-clustering nodes and the overlapping hot nodes accounted for only 16% of MNTD-clustering nodes. The overlap between hot nodes and MNTD-overdispersion nodes was significantly higher than random expectation ($n = 17$; $\chi^2 = 15.598$, $p < 0.001$).

4 | DISCUSSION

4.1 | Phylogenetic divergence across taxonomic resolutions

The results presented here demonstrate that the choice of taxonomic resolution can significantly affect outcomes when analysing phylogenetic divergence patterns of medicinal plant assemblages. In this study, switching from genus to species level for MPD resulted in a

TABLE 1 Number of taxa and statistical significance of phylogenetic divergence metrics per medicinal category. The table shows the mean *p*-value for mean pairwise distance (MPD) and mean nearest taxon distance (MNTD) across 100 phylogenetic trees at each taxonomic resolution and medicinal category. Statistically significant results at 5% nominal alpha (mean *p* < 0.025 and mean *p* > 0.975 for clustering and overdispersion, respectively) are highlighted with *.

	Species level			Genus level		
	Taxa	MPD	MNTD	Taxa	MPD	MNTD
Whole medicinal flora	1,803	1.000*	0.999*	736	0.001*	0.895
Anaesthetic	15	0.045	0.118	13	0.026	0.192
Antidotes	293	0.602	0.076	140	0.493	0.606
Blood system disorders	16	0.429	0.783	16	0.282	0.553
Circulatory system disorders	75	0.063	0.006*	60	0.002*	0.036
Digestive system disorders	247	0.434	0.630	183	0.007*	0.660
Endocrine system disorders	78	0.118	0.872	64	0.016*	0.951
Genitourinary system disorders	145	0.324	0.114	113	0.015*	0.262
Immune system disorders	21	0.033	0.223	21	0.003*	0.060
Infections/infestations	520	0.805	0.490	303	0.001*	0.921
Inflammation	256	0.641	0.222	179	0.065	0.470
Injuries	341	0.193	0.536	215	0.004*	0.911
Mental disorders	26	0.048	0.158	20	0.111	0.604
Metabolic system disorders	21	0.094	0.305	17	0.142	0.672
Muscular-skeletal system disorders	149	0.501	0.950	114	0.027	0.991*
Nervous system disorders	110	0.148	0.742	88	0.005*	0.564
Nutritional disorders	35	0.216	0.864	34	0.046	0.697
Pregnancy/birth/puerperium disorders	143	0.174	0.521	109	0.018*	0.895
Respiratory system disorders	192	0.072	0.963	151	0.001*	0.816
Sensory system disorders	48	0.480	0.500	39	0.476	0.820
Skin/subcutaneous cellular tissue disorders	266	0.003*	0.331	187	0.001*	0.409
Symptoms	878	0.882	0.948	450	0.002*	0.991*
Tumours/cancers	80	0.741	0.206	62	0.733	0.520

radical shift from significant clustering to overdispersion when analysing the whole medicinal flora, and the number of individual use categories showing clustering substantially decreased from 12 to only one. This is crucial because, in the growing body of research within this field, species- and genus-level phylogenies are indiscriminately used, neglecting their potential impact.

Our study system involved hyperdiverse genera with an exceptionally small proportion of medicinal species, which may partially explain the observed discrepancies across taxonomic levels. For instance, the genera *Epidendrum* and *Stelis* comprise nine and only one medicinal representative out of 481 and 494 species in Ecuador. When analysed at the species level, this abundance of non-medicinal species contributes longer-than-expected distances among the medicinal ones (Figure 3). However, switching from species to genus level decreases the likelihood of randomly selecting two taxa that are more closely related than the medicinal ones, and, consequently, phylogenetic divergence metrics show higher clustering levels (Figure 3).

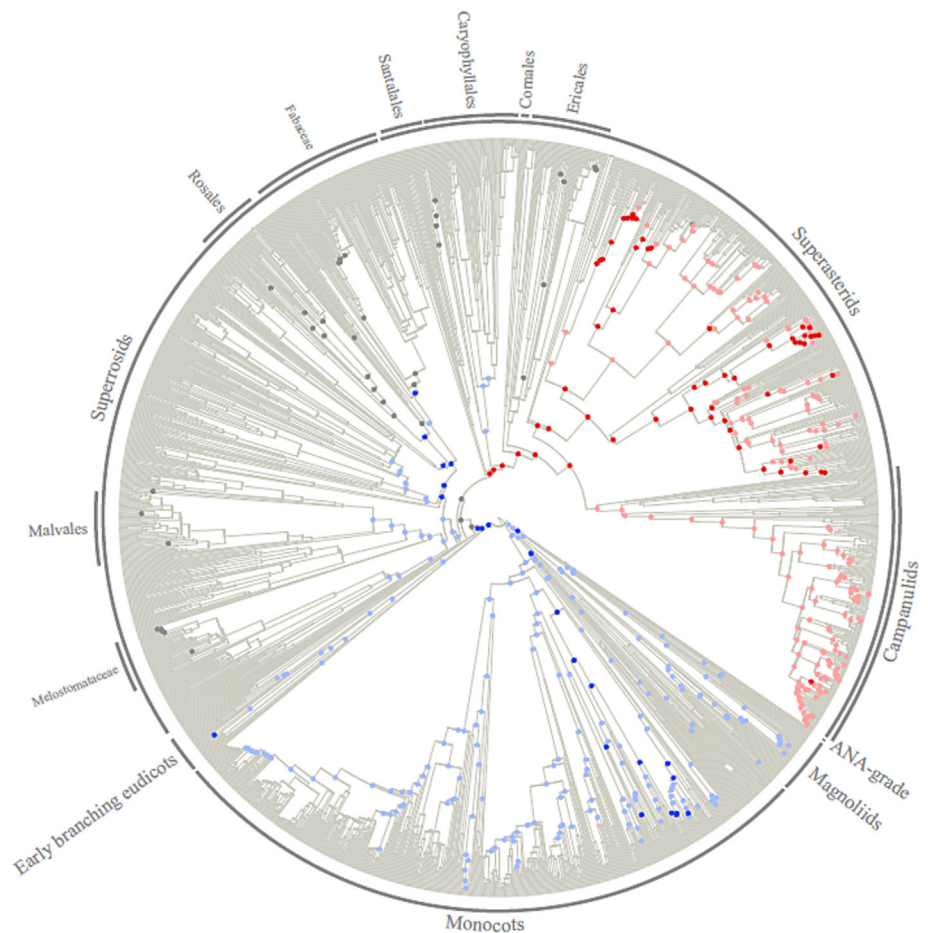
The absence of phylogenetic clustering found here at the species level may also be ascribed to the limited ethnobotanical documentation in the region. Research focused on the medicinal angiosperms of

China, a country with extraordinarily extensive historical documentation (Jaiswal et al., 2016), revealed strong clustering on species-level phylogenies (Zaman et al., 2021). Notably, approximately one-fourth of Chinese angiosperms are medicinal, whereas in Ecuador, only one in 10 species is registered as such (de la Torre et al., 2008). Therefore, our results may underscore the need to intensify documentation efforts in this region (Cámara-Leret et al., 2014).

On a different note, discrepancies between MPD and MNTD highlight the benefits of combining these metrics. MPD and MNTD are predominantly computed in tandem, for they provide complementary information on 'deep' and 'shallow' tree-branching structures, respectively (Mazel et al., 2016; Webb et al., 2002). For instance, in our genus-level analysis of the whole medicinal flora, MPD indicated significant clustering while MNTD was not significant. This suggests that Ecuadorian medicinal genera tend to cluster in large clades but do not show significant structure within them.

The combined application of both metrics and taxonomic resolutions may contribute to a deeper understanding of eco-phylogenetic patterns. However, for bioprospection objectives, the use of MNTD and the most resolved phylogeny possible may be preferable.

FIGURE 1 Differential contribution of phylogenetic clades to mean pairwise distance (MPD). Nodes that significantly contributed to phylogenetic clustering ($p > 0.975$) and overdispersion ($p < 0.025$) at the 5% nominal alpha are depicted in red and blue, respectively. [Correction added on 1 February, after first online publication: The labelling of p -values in preceding sentence has been corrected.] [Correction added on April 21 2025, after first online publication: The p -value symbols in preceding sentence has been inverted.] Brighter colours indicate an overlap with hot nodes, while unmatched hot nodes are shown in grey.



MNTD is often considered a better indicator than MPD for screening bioactive plants, as it narrows down potential target species (Sasli-Lagoudakis et al., 2011; Zaman et al., 2021), given its closer association with shallow clades (Mazel et al., 2016; Webb et al., 2002). Similarly, species-level phylogenies allow for a more precise identification of candidates, because traditional uses are commonly attributed to species rather than genera. Besides, considering that changes in taxonomic resolution produce different outcomes, genus-level analyses may not only yield less informative results but also lead to misleading conclusions. For example, a node encompassing several genera might be identified as 'hot' when analysed at the genus level if at least one species per genus is recorded as medicinal. However, conducting the same analysis at the species level may reveal that the node is not significantly 'hot' if only a small fraction of the species within each genus are recorded as medicinal, as observed with Ecuadorian *Stelis* and *Epidendrum*.

4.2 | Clades exerting a significant influence on phylogenetic divergence

While the hot nodes have been used as the primary method for identifying clades with a significant impact on phylogenetic clustering, this study demonstrates their limited revelatory power. Instead, we

highlight the potential of underexploited jackknifing techniques (Davies, 2021; Davies et al., 2016; Molina-Venegas et al., 2019). The hot nodes overlapped with only 24% and 16% of the nodes that significantly contributed to clustering in MPD and MNTD, respectively, thereby overlooking many relevant taxa, such as most medicinal plants within the campanulid lineage (Figures 1 and 2).

Additionally, our analysis revealed that many hot nodes did not significantly impact MPD and MNTD (e.g. certain groups in orders Ericales, Santalales, Rosales, Malvales and families Fabaceae and Melastomataceae) and, more critically, a significant number of them contributed to overdispersion instead of clustering (e.g. certain groups in Areaceae, Commelinales, Zingiberales and early branching eudicots in the case of MPD, as well as some deep nodes within the superrosid lineage for both MPD and MNTD; Figures 1 and 2). Therefore, if the clades that significantly influence phylogenetic clustering are indeed those with medicinal value, relying solely on hot nodes may misguide the search for bioactive potential.

Hot nodes depend exclusively on tree topology, while MPD and MNTD depend on both tree topology and phylogenetic distances. Thus, for instance, if a clade showing a significant overabundance of medicinal taxa (i.e. a hot node) has considerably long branches, it may have no significant impact on phylogenetic clustering or even contribute to overdispersion (Figure 4). However, the use of MPD and MNTD in the search for bioactive potential is based on the premise

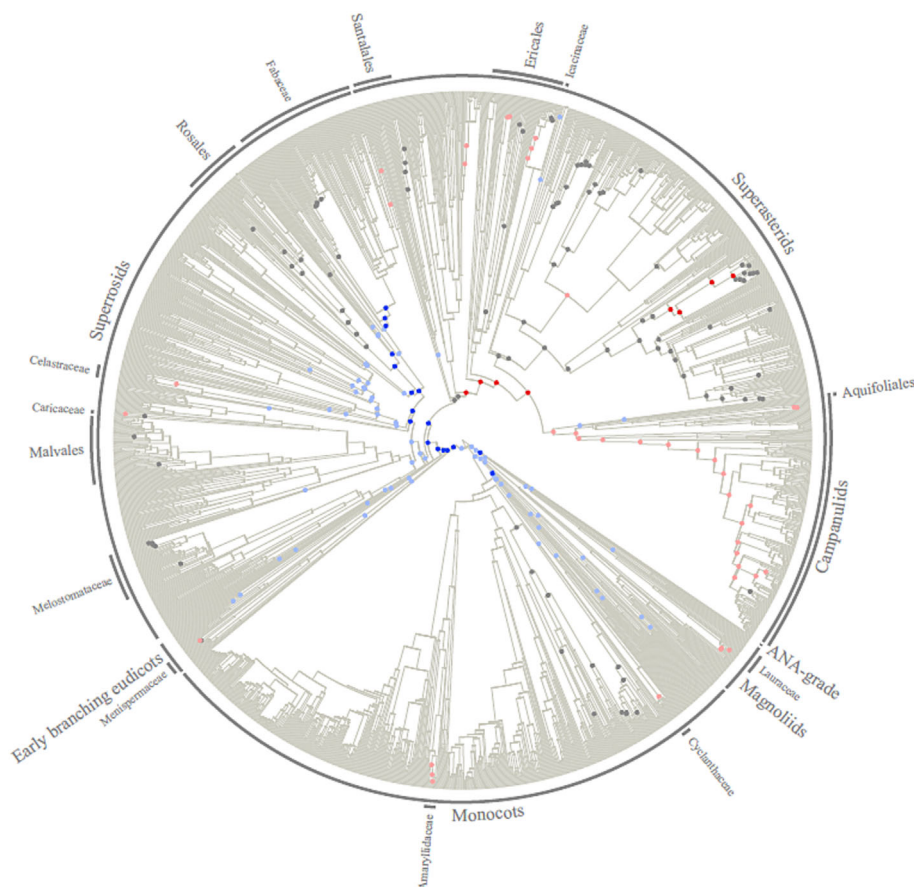


FIGURE 2 Differential contribution of phylogenetic clades to mean nearest taxon distance (MNTD). Nodes that significantly contributed to phylogenetic clustering ($p > 0.975$) and overdispersion ($p < 0.025$) at the 5% nominal alpha are depicted in red and blue, respectively. [Correction added on 1 February, after first online publication: The labelling of p -values in preceding sentence has been corrected.] [Correction added on April 21 2025, after first online publication: The p -value symbols in preceding sentence has been inverted.] Brighter colours indicate an overlap with hot nodes, while unmatched hot nodes are shown in grey.

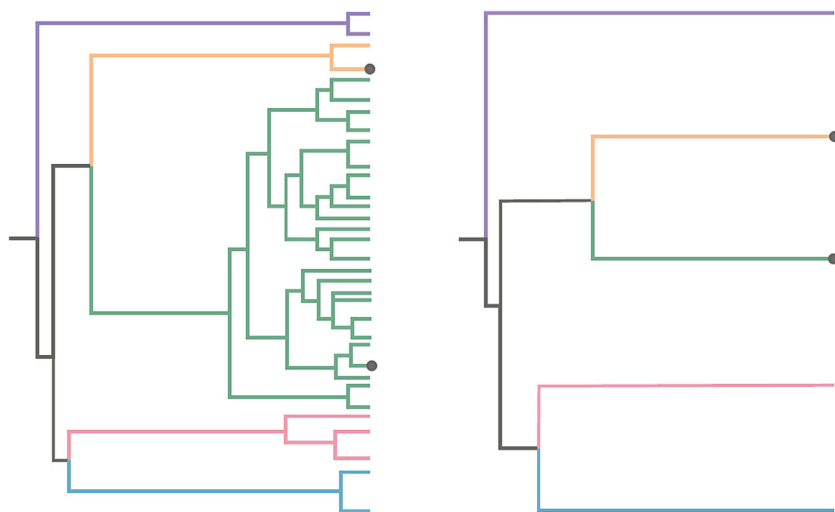


FIGURE 3 Hypothetical illustration of contrasting phylogenetic divergence patterns at different taxonomic resolutions. The left and right figures depict species- and genus-level phylogenies, respectively, with each genus in a distinct colour. Circle symbols indicate medicinal taxa in both scenarios. The green genus is highly diversified, so at the species level, there is a high probability of randomly picking two terminal nodes that are phylogenetically closer than the two medicinal ones (hence resulting in phylogenetic overdispersion of medicinal species), while at the genus level, such probability is equal to zero, as the medicinal taxa are the closest relatives in the phylogeny.

that phytochemical distances correlate with phylogenetic distances. Accordingly, bioactivity profiles would typically evolve gradually, so that species separated by shorter phylogenetic distances would more likely exhibit similar levels and/or molecular structures of these compounds. Nevertheless, this assumption is uncertain, prompting consideration of other alternative models like discontinuous evolution. Under that scenario, phylogenetic distances may not reliably indicate bioactive potential, and thus hot nodes may provide a more accurate identification of promising candidate clades for bioprospection. Given

this uncertainty, we recommend leveraging all available tools to improve the selection of target plants. For example, if a hot node is also significantly driving phylogenetic divergence towards clustering, it would be reasonable to think that the clade is a strong candidate for bioprospection.

In the context of identifying target species for bioprospection, emphasis typically rests on hot nodes and clustering-driving nodes. However, we argue that employing jackknifing techniques to pinpoint clades significantly contributing to overdispersion may also be

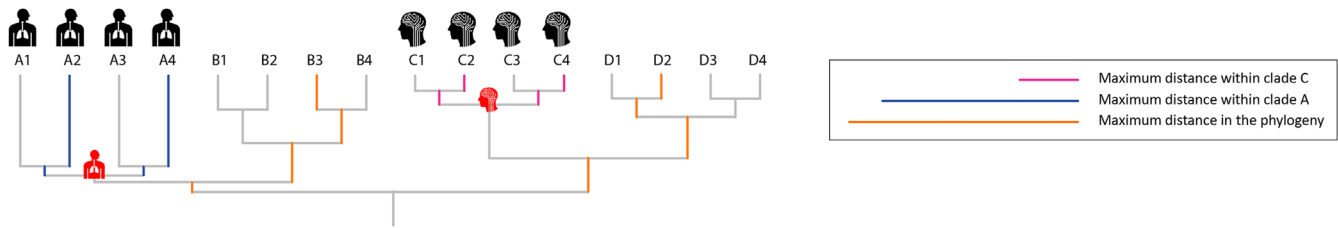


FIGURE 4 Hypothetical example to illustrate the matching between hot nodes and those contributing to phylogenetic divergence. The phylogeny includes 16 species, of which those constituting clades A and C are used to treat respiratory and nervous system diseases, respectively. The nodes highlighted in red are identified as hot nodes, indicating they include a higher-than-expected number of medicinal species. The species used for nervous system diseases are phylogenetically clustered, and the hot node analysis successfully identified the clade responsible for this pattern (clade C). Conversely, the species used for respiratory diseases are not phylogenetically clustered, yet their clade was still identified as a hot clade (clade A). Although the species in clade A are as topologically close as those in clade C, the pairwise phylogenetic distances within clade A are significantly longer than those within clade C and nearly as long as the maximum distance between any two species in the phylogeny (see inset). The respiratory system icon is protected under C BY 3.0 license, and it should be credited to Laymik UA (<https://thenounproject.com/icon/respiratory-system-2937635/>). The nervous system icon is protected under C BY 3.0 License, and it should be credited to Wes Breazell US (<https://thenounproject.com/icon/brain-189792/>).

valuable. These clades represent phylogenetically isolated groups distinct from the main clusters of beneficial species, hence potentially offering additional, complementary benefits (Molina-Venegas, 2021). The significance of phylogenetically isolated species in delivering multiple benefits has been validated globally in plants and birds (Gumbs et al., 2023; Molina-Venegas, 2021), and their potential medicinal value remains unknown and warrants future investigation.

In conclusion, the hot node analysis is a valuable tool that can be combined with phylogenetic divergence metrics for a more comprehensive approach. However, these methods rely on distinct components of the phylogeny and, consequently, the hot nodes only partially overlap with the nodes responsible for phylogenetic divergence. Therefore, we stress that if the goal is assessing the clades responsible for phylogenetic clustering, jackknifing techniques such as the one implemented here are preferable over the hot node approach. Beyond ethnomedicinal research, these techniques are widely used in other fields linked to community ecology, including forest dynamics (Arroyo-Rodríguez et al., 2012; Letcher, 2010; Ribeiro et al., 2016), biogeography (Souza-Neto et al., 2016; Bose et al., 2019) and plant invasion (Kalusová et al., 2021), among others. Hence, our findings bear important implications, not only for the exploration of the medicinal potential of biodiversity but also for advancing our understanding of how evolutionary histories influence the present-day ecologies of species assemblages.

AUTHOR CONTRIBUTIONS

RM-V conceived the idea with inputs from DA-B and MJM. DA-B curated the dataset, conducted the analyses with help from RM-V, drafted the figures and led the writing. MJM contributed the dataset and the ethnobotanical background of the research. All authors read and commented on the manuscript.

ACKNOWLEDGEMENTS

Our sincere gratitude goes to Lucía de la Torre, Hugo Navarrete, Priscilla Muriel M., Henrik Balslev and everyone involved in assembling the information forming the basis of this work. We also recognise the

invaluable contributions of the native communities that are the source of this data. We thank the Scientific Computation Centre of Andalusia (CICA) for the computing services they provided. This research was supported by the ‘Investigo’ programme, funded by the European Union (NextGenerationEU), granted to MJM by the Universidad Autónoma de Madrid (contract No. 39), and the project PID2022-138776NA-I00 granted to RM-V by the Spanish Ministry of Science, Innovation, and Universities.

DATA AVAILABILITY STATEMENT

The datasets and the phylogenies that support the findings of this study are openly available in figshare at [10.6084/m9.figshare.27134757.v1](https://figshare.com/10.6084/m9.figshare.27134757.v1) and the R code is available as part of the Supporting Information.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Dario Atienza-Barthelemy [ORCID](https://orcid.org/0009-0005-9554-168X) <https://orcid.org/0009-0005-9554-168X>

Manuel J. Macia [ORCID](https://orcid.org/0000-0002-4676-612X) <https://orcid.org/0000-0002-4676-612X>

Rafael Molina-Venegas [ORCID](https://orcid.org/0000-0001-5801-0736) <https://orcid.org/0000-0001-5801-0736>

REFERENCES

- Arroyo-Rodríguez, V., Cavender-Bares, J., Escobar, F., Melo, F. P. L., Tabarelli, M., & Santos, B. A. (2012). Maintenance of tree phylogenetic diversity in a highly fragmented rain forest: Tree phylogenetic diversity in fragmented forests. *Journal of Ecology*, 100(3), 702–711. <https://doi.org/10.1111/j.1365-2745.2011.01952.x>
- Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., & Supuran, C. T. (2021). Natural products in drug discovery: Advances and opportunities. *Nature Reviews Drug Discovery*, 20(3), 200–216. <https://doi.org/10.1038/s41573-020-00114-z>
- Bose, R., Ramesh, B. R., Pélissier, R., & Munoz, F. (2019). Phylogenetic diversity in the Western Ghats biodiversity hotspot reflects environmental filtering and past niche diversification of trees. *Journal of Biogeography*, 46(1), 145–157. <https://doi.org/10.1111/jbi.13464>

- Cámara-Leret, R., Paniagua-Zambrana, N., Balslev, H., & Macía, M. J. (2014). Ethnobotanical knowledge is vastly under-documented in northwestern South America. *PLoS ONE*, 9(1), e85794. <https://doi.org/10.1371/journal.pone.0085794>
- Choudhury, P. R., Talukdar, A. D., Nath, D., Saha, P., & Nath, R. (2020). Traditional folk medicine and drug discovery: Prospects and outcome. In J. K. Patra, A. C. Shukla, & G. Das (Eds.), *Advances in pharmaceutical biotechnology: Recent Progress and future applications* (pp. 3–13). Springer. https://doi.org/10.1007/978-981-15-2195-9_1
- Cook, F. E. M. (1995). *Economic botany data collection standard*. Royal Botanic Gardens (Kew).
- Davies, T. J. (2021). Ecophylogenetics redux. *Ecology Letters*, 24(5), 1073–1088. <https://doi.org/10.1111/ele.13682>
- Davies, T. J., Urban, M. C., Rayfield, B., Cadotte, M. W., & Peres-Neto, P. R. (2016). Deconstructing the relationships between phylogenetic diversity and ecology: A case study on ecosystem functioning. *Ecology*, 97(9), 2212–2222. <https://doi.org/10.1002/ecy.1507>
- de la Torre, L., Cerón, C. E., Balslev, H., & Borchsenius, F. (2012). A biodiversity informatics approach to ethnobotany: Meta-analysis of plant use patterns in Ecuador. *Ecology and Society*, 17(1), art15. <https://doi.org/10.5751/ES-04582-170115>
- de la Torre, L., Navarrete, H., Muriel, P., Macía, M. J., & Balslev, H. (2008). *Enciclopedia de las Plantas Útiles del Ecuador*. Herbario QCA & Herbario AAU.
- Ernst, M., Saslis-Lagoudakis, C. H., Grace, O. M., Nilsson, N., Simonsen, H. T., Horn, J. W., & Rønsted, N. (2016). Evolutionary prediction of medicinal properties in the genus *euphorbia* L. *Scientific Reports*, 6(1), 30531. <https://doi.org/10.1038/srep30531>
- GBIF. (2022). The Global Biodiversity Information Facility. Published on the Internet: <http://www.theplantlist.org>. (n.d.)
- Graham, C. H., Storch, D., Machac, A., & Isaac, N. (2018). Phylogenetic scale in ecology and evolution. *Global Ecology and Biogeography*, 27(2), 175–187. <https://doi.org/10.1111/geb.12686>
- Gumbs, R., Gray, C. L., Hoffmann, M., Molina-Venegas, R., Owen, N. R., & Pollock, L. J. (2023). Conserving avian evolutionary history can effectively safeguard future benefits for people. *Science Advances*, 9(38), eadh4686. <https://doi.org/10.1126/sciadv.adh4686>
- Gyllenhaal, C., Kadushin, M. R., Southavong, B., Sydara, K., Bouamanivong, S., Xaiveu, M., Xuan, L. T., Hiep, N. T., Hung, N. V., Loc, P. K., Dac, L. X., Bich, T. Q., Cuong, N. M., Ly, H. M., Zhang, H. J., Franzblau, S. G., Xie, H., Riley, M. C., Elkington, B. G., ... Soejarto, D. D. (2012). Ethnobotanical approach versus random approach in the search for new bioactive compounds: Support of a hypothesis. *Pharmaceutical Biology*, 50(1), 30–41. <https://doi.org/10.3109/13880209.2011.634424>
- Halse-Gramkow, M., Ernst, M., Rønsted, N., Dunn, R. R., & Saslis-Lagoudakis, C. H. (2016). Using evolutionary tools to search for novel psychoactive plants. *Plant Genetic Resources*, 14(4), 246–255. <https://doi.org/10.1017/S1479262116000344>
- Jaiswal, Y., Liang, Z., & Zhao, Z. (2016). Botanical drugs in Ayurveda and traditional Chinese medicine. *Journal of Ethnopharmacology*, 194, 245–259. <https://doi.org/10.1016/j.jep.2016.06.052>
- Jin, Y., & Qian, H. (2019). V.PhyloMaker: An R package that can generate very large phylogenies for vascular plants. *Ecography*, 42(8), 1353–1359. <https://doi.org/10.1111/ecog.04434>
- Jørgensen, P. M., & León-Yáñez, S. (1999). *Catalogue of the vascular plants of Ecuador*. Missouri Botanical Garden Press St.
- Kalusová, V., Padullés Cubino, J., Fristoe, T. S., Chytrý, M., Van Kleunen, M., Dawson, W., Essl, F., Kreft, H., Mucina, L., Pergl, J., Pyšek, P., Weigelt, P., Winter, M., & Lososová, Z. (2021). Phylogenetic structure of alien plant species pools from European donor habitats. *Global Ecology and Biogeography*, 30(12), 2354–2367. <https://doi.org/10.1111/geb.13387>
- Kemmel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26(11), 1463–1464. <https://doi.org/10.1093/bioinformatics/btq166>
- Khaiwa, N., Maarouf, N. R., Darwish, M. H., Alhamad, D. W. M., Sebastian, A., Hamad, M., Omar, H. A., Orive, G., & Al-Tel, T. H. (2021). Camptothecin's journey from discovery to WHO essential medicine: Fifty years of promise. *European Journal of Medicinal Chemistry*, 223, 113639. <https://doi.org/10.1016/j.ejmech.2021.113639>
- Kindt, R. (2020). WorldFlora: An R package for exact and fuzzy matching of plant names against the world Flora online taxonomic backbone data. *Applications in Plant Sciences*, 8(9), e11388. <https://doi.org/10.1002/aps3.11388>
- Kingston, D. G. I. (2011). Modern natural products drug discovery and its relevance to biodiversity conservation. *Journal of Natural Products*, 74(3), 496–511. <https://doi.org/10.1021/np100550t>
- Lei, D., Al Jabri, T., Teixidor-Toneu, I., Saslis-Lagoudakis, C. H., Ghazanfar, S. A., & Hawkins, J. A. (2020). Comparative analysis of four medicinal floras: Phylogenetic methods to identify cross-cultural patterns. *Plants, People, Planet*, 2(6), 614–626. <https://doi.org/10.1002/ppp3.10131>
- Letcher, S. G. (2010). Phylogenetic structure of angiosperm communities during tropical forest succession. *Proceedings of the Royal Society B: Biological Sciences*, 277(1678), 97–104. <https://doi.org/10.1098/rspb.2009.0865>
- Mazel, F., Davies, T. J., Gallien, L., Renaud, J., Groussin, M., Münkemüller, T., & Thuiller, W. (2016). Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. *Ecography*, 39(10), 913–920. <https://doi.org/10.1111/ecog.01694>
- Miao, S. M., Zhang, Q., Bi, X. B., Cui, J. L., & Wang, M. L. (2020). A review of the phytochemistry and pharmacological activities of ephedra herb. *Chinese Journal of Natural Medicines*, 18(5), 321–344. [https://doi.org/10.1016/S1875-5364\(20\)30040-6](https://doi.org/10.1016/S1875-5364(20)30040-6)
- Molina-Venegas, R. (2021). Conserving evolutionarily distinct species is critical to safeguard human well-being. *Scientific Reports*, 11(1), 24187. <https://doi.org/10.1038/s41598-021-03616-x>
- Molina-Venegas, R., Fischer, M., & Hemp, A. (2019). Disentangling the fundamental branching patterns of phylogenetic divergence to refine eco-phylogenetic analyses. *Journal of Biogeography*, 46(12), 2722–2734. <https://doi.org/10.1111/jbi.13692>
- Molina-Venegas, R., Fischer, M., Mollé, N. P., & Hemp, A. (2020). Connecting plant evolutionary history and human well-being at Mt. Kilimanjaro, Tanzania. *Botanical Journal of the Linnean Society*, 194(4), 397–409. <https://doi.org/10.1093/botlinnean/boaa049>
- Molina-Venegas, R., Ottaviani, G., Campetella, G., Canullo, R., & Chelli, S. (2022). Biogeographic deconstruction of phylogenetic and functional diversity provides insights into the formation of regional assemblages. *Ecography*, 2022(5), e06140. <https://doi.org/10.1111/ecog.06140>
- Montinari, M. R., Minelli, S., & De Caterina, R. (2019). The first 3500 years of aspirin history from its roots – A concise summary. *Vascular Pharmacology*, 113, 1–8. <https://doi.org/10.1016/j.vph.2018.10.008>
- Pellicer, J., Saslis-Lagoudakis, C. H., Carrió, E., Ernst, M., Garnatje, T., Grace, O. M., & Rønsted, N. (2018). A phylogenetic road map to anti-malarial Artemisia species. *Journal of Ethnopharmacology*, 225, 1–9. <https://doi.org/10.1016/j.jep.2018.06.030>
- POWO. (2022). Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; <http://www.plantsoftheworldonline.org>. (n.d.)
- Rangel, T. F., Colwell, R. K., Graves, G. R., Fučíková, K., Rahbek, C., & Diniz-Filho, J. A. F. (2015). Phylogenetic uncertainty revisited: Implications for ecological analyses. *Evolution*, 69(5), 1301–1312. <https://doi.org/10.1111/evo.12644>
- Ribeiro, E. M. S., Santos, B. A., Arroyo-Rodríguez, V., Tabarelli, M., Souza, G., & Leal, I. R. (2016). Phylogenetic impoverishment of plant communities following chronic human disturbances in the Brazilian Caatinga. *Ecology*, 97(6), 1583–1592. <https://doi.org/10.1890/15-1122.1>

- Rønsted, N., Symonds, M. R. E., Birkholm, T., Christensen, S., Meerow, A. W., Molander, M., Mølgaard, P., Petersen, G., Rasmussen, N., van Staden, J., Stafford, G. I., & Jäger, A. K. (2012). Can phylogeny predict chemical diversity and potential medicinal activity of plants? A case study of amaryllidaceae. *BMC Evolutionary Biology*, 12(1), 182. <https://doi.org/10.1186/1471-2148-12-182>
- Saslis-Lagoudakis, C. H., Klitgaard, B. B., Forest, F., Francis, L., Savolainen, V., Williamson, E. M., & Hawkins, J. A. (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(7), e22275. <https://doi.org/10.1371/journal.pone.0022275>
- Saslis-Lagoudakis, C. H., Savolainen, V., Williamson, E. M., Forest, F., Wagstaff, S. J., Baral, S. R., Watson, M. F., Pendry, C. A., & Hawkins, J. A. (2012). Phylogenies reveal predictive power of traditional medicine in bioprospecting. *Proceedings of the National Academy of Sciences*, 109(39), 15835–15840. <https://doi.org/10.1073/pnas.1202242109>
- Souza, E. N. F., Williamson, E. M., & Hawkins, J. A. (2018). Which plants used in ethnomedicine are characterized? Phylogenetic patterns in traditional use related to research effort. *Frontiers in Plant Science*, 9, 834. <https://doi.org/10.3389/fpls.2018.00834>
- Souza-Neto, A. C., Cianciaruso, M. V., & Collevatti, R. G. (2016). Habitat shifts shaping the diversity of a biodiversity hotspot through time: Insights from the phylogenetic structure of Caesalpinioideae in the Brazilian Cerrado. *Journal of Biogeography*, 43(2), 340–350. <https://doi.org/10.1111/jbi.12634>
- Thompson, J. B., & Hawkins, J. A. (2024). Phylogeny and bioprospecting: The diversity of medicinal plants used in cancer management. *Plants People Planet*, in press. <https://doi.org/10.1002/ppp3.10566>
- Tu, Y. (2011). The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nature Medicine*, 17(10), 1217–1220. <https://doi.org/10.1038/nm.2471>
- Villalobos, S., Sevenello-Montagner, J. M., & Vamosi, J. C. (2019). Specialization in plant–pollinator networks: Insights from local-scale interactions in Glenbow ranch Provincial Park in Alberta, Canada. *BMC Ecology*, 19(1), 34. <https://doi.org/10.1186/s12898-019-0250-z>
- Wangkheirakpam, S. (2018). Traditional and folk medicine as a target for drug discovery. In S. C. Mandal, V. Mandal, & T. Konishi (Eds.), *Natural products and drug discovery: An integrated approach* (pp. 29–56). Elsevier. <https://www.sciencedirect.com/science/article/pii/B978008102081400022>, <https://doi.org/10.1016/B978-0-08-102081-4.00002-2>
- Webb, C. O., Ackerly, D. D., & Kembel, S. W. (2008). Phylocom: Software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18), 2098–2100. <https://doi.org/10.1093/bioinformatics/btn358>
- Webb, C. O., Ackerly, D. D., McPeck, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, 33(1), 475–505. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448>
- WFO. (2022): World Flora Online. Published on the Internet; <http://www.worldfloraonline.org>. (n.d.).
- Yessoufou, K., Daru, B. H., & Muasya, A. M. (2015). Phylogenetic exploration of commonly used medicinal plants in South Africa. *Molecular Ecology Resources*, 15(2), 405–413. <https://doi.org/10.1111/1755-0998.12310>
- Zaman, W., Ye, J., Ahmad, M., Saqib, S., Shinwari, Z. K., & Chen, Z. (2022). Phylogenetic exploration of traditional Chinese medicinal plants: A case study on Lamiaceae. *Pakistan Journal of Botany*, 54(3), 1033–1040. [https://doi.org/10.30848/PJB2022-3\(19\)](https://doi.org/10.30848/PJB2022-3(19))
- Zaman, W., Ye, J., Saqib, S., Liu, Y., Shan, Z., Hao, D., Chen, Z., & Xiao, P. (2021). Predicting potential medicinal plants with phylogenetic topology: Inspiration from the research of traditional Chinese medicine. *Journal of Ethnopharmacology*, 281, 114515. <https://doi.org/10.1016/j.jep.2021.114515>

SUPPORTING INFORMATION

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

How to cite this article: Atienza-Barthelemy, D., Macia, M. J., & Molina-Venegas, R. (2025). Hot node limitations and impact of taxonomic resolution on phylogenetic divergence patterns: A case study on Ecuadorian ethnomedicinal flora. *Plants, People, Planet*, 7(3), 644–653. <https://doi.org/10.1002/ppp3.10594>