

Evolutionary history is a powerful means to evince significant levels of plant services for humankind

Rafael Molina-Venegas (✉ rmolina@us.es)

University of Alcalá <https://orcid.org/0000-0001-5801-0736>

Miguel Rodriguez

University of Alacala

Manuel Pardo-de-Santayana

Universidad Autónoma de Madrid

Cristina Ronquillo

Universidad de Alcalá

David J. Mabberley

National Herbarium of New South Wales

Brief Communication

Keywords: Biodiversity, Plant-use Records, Vascular Plant Phylogeny, Human Benefits

Posted Date: December 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-121288/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Ecology & Evolution on March 29th, 2021. See the published version at <https://doi.org/10.1038/s41559-021-01414-2>.

26 Evolution is the process that led to all living organisms and hence the foundation for the
27 human benefits that are directly provided by biodiversity¹, including not only basic resources
28 but also psychological and self-fulfilment needs². Since evolution is a divergent process, some
29 researchers have claimed that securing the services that are directly provided by biodiversity
30 may require counting on disparate lineages of the Tree of Life¹, because they might provide
31 complementary benefits³. Although this theoretical background is deeply rooted in the academic
32 literature⁴⁻⁶, empirical evidence connecting evolutionary history to human well-being is still
33 surprisingly tenuous⁷ and not without controversy^{8,9}. While some authors hold that maximizing
34 phylogenetic diversity should lead to recognition of high levels of useful feature diversity^{1,8},
35 others have suggested that the phylogenetic approach can be misleading⁹. This controversy
36 likely reflects that the connection between evolutionary history and human well-being remains
37 largely theoretical¹⁰ (but see Forest et al.¹¹ for an empirical local assessment), which is only an
38 initial move towards its consolidation as a scientific paradigm.

39 Here, we provide quantitative evidence that maximum levels of global plant
40 phylogenetic diversity (PD_{max}) capture more human benefits (i.e. plant-use records sorted into
41 28 standard categories of use¹²) and at higher diversity levels (i.e. records more evenly
42 distributed between the categories) than does random selection of taxa, supporting the long-
43 standing notion that maximizing phylogenetic diversity is a valuable means to retrieve high
44 levels of useful feature diversity⁴⁻⁶. Our genus-level analysis is based on the most
45 comprehensive time-calibrated vascular plant phylogeny available^{13,14}, including all accepted
46 vascular plant genera worldwide (a total of 13489) as well as 9478 genus-level plant-use records
47 (presence/absence) obtained from a systematic review of botanical literature and authoritative
48 websites¹⁵.

49 The PD_{max} strategy overcame random selection of taxa at any sample size (Fig. 1a), with
50 relative gains varying between 4% and 46% (Fig. 1b). This result suggests that in the absence of
51 any other source of information beyond evolutionary history, prospecting disparate lineages of
52 the phylogeny could help to make the most of the natural services that are the result of
53 evolution. With regard to individual plant-use categories, PD_{max} retrieved a higher number of

54 records relative to random selection in 92% of the comparisons (Fig. 2 and Supplementary Fig.
55 S1). Moreover, given that relative record gains with PD_{max} were overall higher for the less
56 common categories (Fig. S2), PD_{max} also retrieved significantly more equitable distributions of
57 records among categories at most sample sizes (Fig. 1c). This indicates that PD_{max} recovers
58 more plant-uses in general than random selection, and that it does so optimizing the capture of
59 some of the rarest uses, thus resulting in a more balanced palette of human benefits. Both PD_{max}
60 and random selection strategies retrieved the maximum possible richness of plant-use categories
61 ($n = 28$) across most sample sizes, yet random selection failed in retrieving maximum richness
62 of categories at 10% and 20% sample sizes in a few cases (data not shown). While our genus-
63 level approach is superior to the species level in that the latter would suffer from unacceptable
64 omission errors –ethnobotanical knowledge will most likely remain vastly under-documented
65 for long below the genus level¹⁶⁻¹⁸– and extreme lack of phylogenetic information¹³, it may
66 introduce some uncertainty because the operational unit of plant-use is often the species. As
67 such, retrieving a useful genus that comprises just a few species could be considered more
68 valuable than a highly diversified one with the same use, because the uncertainty regarding the
69 species that are actually useful within each genus would be less in the former case. Nonetheless,
70 a reanalysis of the data after downweighting our genus-level plant-use observations in direct
71 proportion to species richness per genus revealed an even stronger pattern (Supplementary Fig.
72 S3). Moreover, the relationship between PD and plant benefits held in separate continental
73 regions of the world (TDWG level-1 standards, Supplementary Figs. S4-S5), which suggests
74 that our results are consistent across floras that have evolved in distinct biogeographic regions
75 and over different timescales.

76 The striking success of the PD_{max} strategy lies in the phylogenetic structure of the
77 categories. As such, we found a strong positive relationship between the PD that is encapsulated
78 by each plant-use category and the relative gain in records per category under the PD_{max} strategy
79 (Fig. 3), meaning that greater gains are predicted for phylogenetically dispersed categories. In
80 fact, the only category that was significantly underrepresented with PD_{max} relative to random

81 selection concerns rubber plants (Fig. 2 and Supplementary Fig. S1), which are strongly
82 clumped in the phylogeny (Supplementary Table S2). Our results complement previous findings
83 reported in local studies that high levels of PD can increase multifunctionality via
84 complementarity of beneficial attributes among phylogenetically distant taxa³. For example,
85 regarding the production of natural poisons against harmful or nuisance invertebrates, we found
86 that maximum levels of global PD capture more plant taxa generating them than random
87 selection (Fig. 2), which in turn may imply an increased potential to control the detrimental
88 effects of disparate invertebrate lineages. While the latter hypothesis cannot be tested with our
89 data, observations that most of the antagonistic plant-invertebrate interactions that ultimately
90 shaped this benefit are phylogenetically conserved^{19,20} (i.e. invertebrate species often attack a
91 narrow range of closely-related host plants) and geographically restricted²¹ support this idea. It
92 follows that, in the shadow of global change, counting on a variety of invertebrate poisons and
93 deterrents from distinct plant lineages may help to counter phylogenetically diverse pests
94 coming from disparate parts of the world^{22,23}.

95 It is important to note that an unobserved link between a human need and a taxon does
96 not necessarily imply that the link will not be found in the future. The ecological apparency
97 hypothesis states that among equally valuable taxa with regard to a certain use, the most
98 apparent or salient ones are preferred simply because they are readily available²⁴. Furthermore,
99 cultural factors could also explain the preferential use of certain taxa at the expense of others
100 that might equally fulfill the need²⁵. By analogy to the ecological prediction that higher
101 competition between closely-related taxa of similar phenotypes can lead to greater phylogenetic
102 diversity²⁶, human preference patterns in the use of available plant resources might have
103 increased phylogenetic overdispersion in local ethnofloras. Therefore, ecological and cultural
104 factors, together with the fact that both plant lineages and the human cultures that prospect them
105 are geographically restricted to a greater or lesser extent, may have contributed to the striking
106 success of the PD_{max} strategy over random selection in capturing the human benefits that are
107 associated with plant biodiversity.

109 Services (IPBES) has recently approved undertaking the assessment of the use of ‘wild’ species,
110 including the identification of opportunities to establish measures that ensure and promote
111 sustainable practices²⁷. Concurring with the IPBES philosophy that the world is in need of a
112 broadly appreciation of nature’s contribution to people^{28,29}, our study establishes a solid
113 empirical foundation that links evolutionary history to human well-being, and it will serve as a
114 discussion baseline to promote better-grounded accounts of the services that are directly
115 provided by biodiversity^{30,31}.

117 **Methods**

118 **Plant-use dataset.** We compiled a genus-level dataset of plant-use records for all vascular plant
119 taxa described to date using the information gathered in the fourth edition of *Mabberley’s plant-*
120 *book*¹⁵. *Mabberley’s plant-book* is the most comprehensive and authoritative encyclopaedic
121 review of global plant classification (genera) and their uses published hitherto. From 1974 to
122 2017 all the information included in *Mabberley’s plant-book* was gathered, sorted, evaluated
123 and synthesized by David Mabberley, who systematically reviewed over 1000 botanical sources
124 including modern Floras, handbooks, periodicals, monographs and websites (all references can
125 be found in Mabberley¹⁵). We conducted a double-check manual screening of all plant-uses
126 described in *Mabberley’s plant-book* and sorted them into 28 standard categories of use
127 following the guidelines in the Economic Botany Data Collection Standard¹² (hereafter
128 “Collection Standard”). When two or more applications of the same category were described for
129 a given taxon, we considered them as a single plant-use record. For example, if the wood of a
130 taxon is used to build poles, furniture and toys (i.e. three different applications), we simply
131 recorded that the taxon provides timber. This procedure resulted in a binary classification of
132 9478 plant-use records across the 28 categories, including benefits related to human and animal
133 nutrition (human food, human-food additives, vertebrate food, invertebrate food), materials
134 (wood, stems, fibres, leaves, seeds/fruits, tannins/dyestuffs, gums/resins, lipids, waxes, scents,
135 latex/rubber), fuels (fuelwood, charcoal, biofuels), medicine (both human and veterinary),

136 poisons (vertebrate poison, invertebrate poison), social (antifertility agents, smoking
137 materials/drugs, symbolic/magic/inspiration) and environmental uses (ornamental,
138 bioindicators/bioremediators, soil improvers, hedging/shelter). A detailed description of the
139 categories is provided in Supplementary Table S1. Despite the use of leaves and seeds/fruits as
140 materials are considered as “miscellaneous” in the Collection Standard, we took them up front
141 as independent categories because we found many records in *Mabberley’s plant-book* that fit
142 into these categories (typically leaves for thatching and seeds/fruits for handicrafts). The
143 environmental categories “erosion control”, “revegetators”, “soil improvers” and “agroforestry”
144 described in the Collection Standard were considered as one single category (i.e. soil improvers)
145 because they were very difficult to tease apart in many cases (e.g. some plants are used in
146 agroforestry because they prevent soil erosion, and revegetators often improve soil quality). The
147 same rationale applies to the Collection Standard categories “shade/shelter” and
148 “boundaries/barriers/supports”, which were merged into a single category (i.e. hedges and
149 shelters). The Collection Standard also recognized different sub-categories of medicine, human
150 food and poisons¹², but we did not distinguish between them here because such information is
151 often unknown and does not make much sense in the context of our global assessment. For
152 example, while we are interested in recording the value of a taxon as human food,
153 distinguishing between the parts of the plant that are actually eaten (sub-categories for human
154 food in the Collection Standard) is rather irrelevant for the purposes of the study. A few records
155 could not be assigned to any of the categories described in the Collection Standard (e.g. spores
156 and inflorescences used as materials), which recommends gathering such cases into
157 “miscellaneous” categories¹². However, we simply disregarded them because such a mixture of
158 poorly represented categories would not make sense in the context of our study. Finally, the
159 category “cork and cork substitutes” described in the Collection Standard was disregarded
160 because we found very few records in *Mabberley’s plant-book* (likely because cork and cork
161 substitutes are provided only by a few species and primarily from *Quercus*). We considered
162 both fully realized (> 99% of the cases) and mooted uses (as long as they were properly
163 documented in the literature), and doubtful entries were disregarded in any case. The resultant

164 plant-use binary matrix (i.e. presence/absence of uses per genus) was used in all the analyses
165 described below. Additionally, we derived a downweighted plant-use matrix by dividing the
166 entries in the binary matrix (plant-use observations at the genus-level) by the total number of
167 accepted species per genus (following *Plants of the World Online*³²). This second matrix was
168 used in a second round of analyses to take into account the uncertainty in the relationship
169 between plant-use records in the genus-level dataset and the species that are actually useful, as
170 the latter information is often unknown.

171 Of all the taxa included in the dataset, 33% showed at least one category of use, with a
172 maximum number of plant-use records per taxa of 17 (Supplementary Fig. S6). The most
173 common category was “ornamental” (26%), followed by “medicine” (16%), “human food”
174 (13%) and “timber” (8%), while the rest of categories occurred at a frequency lower than 5%
175 (Supplementary Fig. S7). The phi correlation coefficient among the categories varied between -
176 0.008 and 0.332, suggesting overall weak relationships among them.

177

178 **Phylogenetic data.** We generated a genus-level time-calibrated molecular phylogeny using the
179 mega-tree GBOTB.extended¹⁴, which is a combination of the GBOTB tree for seed plants of
180 Smith & Brown¹³ and the pteridophytes clade in Zanne et al.³³ phylogeny with updates and
181 corrections (i.e. taxonomic standardization to *The Plant List*³⁴ nomenclatural and spelling
182 criteria). This combined phylogeny represents the most comprehensive and sophisticated
183 molecular phylogeny for vascular plants published hitherto. For each accepted genus in
184 *Mabberley’s plant-book*, we picked one representative species at random from the largest
185 monophyletic cluster of the genus in GBOTB.extended (if available). In the very few cases
186 where more than one largest monophyletic cluster was found, we first selected one of the
187 clusters at random and then picked one representative species. The GBOTB.extended phylogeny
188 was then pruned to retain only the representative species of the genera. After resolving a few
189 discrepancies and synonymy issues between *Mabberley’s plant-book*¹⁵ and *The Plant List*³⁴
190 (using the nomenclatural criteria in *Plants of the World Online*³² as a complementary reference
191 to solve disputes), we found that 71% of the genera accepted in *Mabberley’s plant-book*

192 included at least one representative species in the phylogeny. This purely molecular
193 phylogenetic topology (hereafter “molecular tree”) revealed that all the taxonomic families of
194 the genera included in the tree formed monophyletic clades except for Nymphaeaceae,
195 Olacaceae, and Tectariaceae, which were paraphyletic, and the polyphyletic Diplaziopsidaceae
196 (see Supplementary Table S3 for a list of genera with taxonomic families). To take into account
197 uncertainty in the phylogenetic relationships of the taxa that were missed in the molecular tree
198 (hereafter “phylogenetically uncertain taxa” or PUT³⁵), we derived a distribution of
199 phylogenetic hypotheses from the latter using a systematic randomization procedure that was
200 taxonomically and phylogenetically informed³⁵. The workflow proposed by Rangel et al.³⁵
201 implies defining for each PUT its “most derived consensus clade” (MDCC) (i.e. the clade in the
202 molecular tree that most certainly contains the PUT) based on expert knowledge (e.g.
203 taxonomy, morphology, geographic distribution, etc). Once the MDCCs of the PUTs are
204 defined, a distribution of phylogenetic hypotheses can be generated by replicating the random
205 insertion of the PUTs within their respective MDCCs a high number of times (e.g. 100 times per
206 posterior tree³⁵). The resultant phylogenetic hypotheses can be then used to replicate the
207 analyses and average the results over the entire distribution of trees^{9,14,35}. Smith & Brown¹³
208 provided just one maximum likelihood tree rather than a posterior distribution, and therefore we
209 derived 100 alternative phylogenetic hypotheses from the maximum likelihood tree as follows.

210 First, we retrieved for each genus in the dataset the taxonomic rank immediately above
211 in the taxonomic hierarchy (typically subtribe, tribe or subfamily in ascending order, hereafter
212 “taxonomic ranks”) from the NCBI Taxonomy database, the standard nomenclature and
213 classification repository for the International Nucleotide Sequence Database Collaboration³⁶.
214 For some families, this information was not available in the NCBI repository, in which case we
215 retrieved the taxonomic ranks from *Mabberley’s plant-book*¹⁵. In the cases where taxonomic
216 ranks were neither available in the latter source, we simply assigned the family rank to the
217 genera. The mapping of taxonomic ranks in the molecular tree reveals whether or not they
218 represent natural lineages (i.e. monophyletic or paraphyletic³⁷), and we took advantage of such
219 information to define the MDCCs for our PUTs. If the taxonomic rank of a PUT mapped as

220 purely monophyletic or purely paraphyletic in the molecular tree, the subset of phylogenetic
221 branches connecting all the genera in the tree that shared the same taxonomic rank as the PUT
222 (hereafter “sharing taxa”) defined the MDCC (see Supplementary Figs. S8a and S9a). In few
223 cases, the taxonomic ranks did not map as purely monophyletic or paraphyletic due to (1) the
224 presence of “outliers” that mapped away from the main cluster of sharing taxa or (2) the
225 presence of “intruders” from a different taxonomic rank within the main cluster. Such outliers
226 and intruders might represent incorrect taxonomic assignments or even artefacts derived from
227 the phylogenetic inference rather than evidence of unnatural (i.e. polyphyletic) groups. Thus, we
228 calculated two different indices for each potential monophyletic or paraphyletic cluster of
229 sharing taxa (because of the presence of outliers, intruders, or both) in the phylogeny. The
230 outlier ratio (OR) for a given set of sharing taxa is the ratio between the number of outliers
231 observed for the set (relative to the largest cluster) and the number of sharing taxa in the set, and
232 the intruder ratio (IR) is the ratio between the number of intruders observed within the largest
233 cluster of sharing taxa and the size of the cluster (see Supplementary Figs. S8 and S9). If (and
234 only if) both ratios were ≤ 0.05 , the subset of phylogenetic branches connecting all the sharing
235 taxa in the largest cluster (i.e. including intruders if any but not outliers) defined the MDCC of
236 the PUT. Otherwise, the MDCC was defined as the smallest phylogenetic clade that included all
237 the sharing taxa in the tree (i.e. including outliers and/or intruders, see Supplementary Figs. S8
238 and S9). In those cases where one single genus represented the only sharing taxon of a PUT in
239 the molecular tree, the terminal node (i.e. the phylogenetic tip) defined the MDCC of the PUT
240 only if the node represented a singleton taxonomic family or subfamily. Otherwise (e.g.
241 singleton tribes or subtribes), the parent node of the singleton sharing taxon defined the MDCC
242 instead (see Supplementary Fig. S10). Once all the PUTs were assigned to a MDCC (see
243 Supplementary Table S4), they were added to a randomly selected branch of their
244 corresponding MDCC, the probability of being added along any branch of the clade being
245 directly proportional to the length of the branch. We used a uniform distribution to determine
246 the exact position to insert the PUTs along the selected branches³⁸. This procedure was
247 replicated 100 times to obtain a distribution of phylogenetic hypotheses.

248

249 **Finding the subsets of genera that maximize phylogenetic diversity.** We used the
250 phylogenetic diversity (PD) index as a metric of the evolutionary history encompassed by a set
251 of taxa⁴ because PD is the most commonly used metric in exercises that aim at maximizing
252 phylogenetic diversity^{4,8,9,39}. The greedy algorithm⁴⁰ was used to find heuristically the subset of
253 genera in the phylogeny that maximized the PD metric (PD_{max}) for a sample size $S = 10, 20, 30,$
254 $40, 50, 60, 70, 80,$ and 90% of the total pool ($n = 13489$), respectively. Because there are
255 multiple subsets of size S that maximize PD in a phylogeny, we produced 10 PD_{max} subsets of
256 genera per alternative phylogenetic hypothesis ($n = 100$) and sample size S . Thus, we obtained
257 1000 different PD_{max} subsets for each sample size S ⁹.

258

259 **Assessing the performance of the PD_{max} strategy.** With regard to human benefits provided by
260 plant biodiversity, the PD_{max} strategy could be considered more powerful than random selection
261 of taxa if the former captures (i) a greater richness of plant-use categories, (ii) a greater number
262 of plant-use records (in total and per category), and (iii) a greater equitability in the distribution
263 of the records among the categories (Pielou's evenness index⁴¹). Thus, for each sample size S ,
264 we computed these variables using 1000 PD_{max} subsets and averaged the results to obtain one
265 observed value per sample size and variable⁹. We used standardized effect sizes (SES) to
266 compare observed values against null distributions generated by randomly picking subsets of S
267 taxa 1000 times:

$$268 \quad \text{SES} = \frac{M_{\text{obs}} - M_{\text{null}}}{SD_{\text{null}}} \quad (1)$$

269 where SES is the standardized effect size score for a given variable and sample size,
270 M_{obs} is the observed averaged value of the variable when taxa selection is phylogenetically
271 informed (i.e. using PD_{max} subsets), M_{null} is the mean of the null distribution (averaged value of
272 the variable when taxa are picked at random), and SD_{null} is the standard deviation of the null
273 distribution.

274

275 **Phylogenetic diversity of plant-use categories.** We computed the amount of evolutionary
276 history (PD) that is encapsulated in each plant-use category in our dataset⁴. PD is not
277 statistically independent of taxa richness, which differed very greatly between the categories
278 (Supplementary Table S2). Therefore, in order to make PD values comparable between them,
279 we computed SES scores using equation 1. Null distributions of PD were generated for each
280 category by shuffling taxa labels across the phylogenetic tips 1000 times⁴², and SES scores were
281 averaged across the 100 phylogenetic hypotheses. All analyses were conducted in R⁴³ using the
282 packages *picante*⁴⁴, *phytools*³⁸ and the greedyPD function developed by Mazel et al.⁹.

283

284 **Continental-scale analyses.** In order to assess whether the relationship between PD and plant
285 benefits holds across floras that have evolved in distinct biogeographic regions, we also
286 conducted all the analyses described above at the continental scale. To do so, we compiled a
287 checklist of the native genera of each TDWG level-1 region (Biodiversity Information
288 Standards⁴⁵), namely, Africa (n = 4487), Australasia (n = 2067), Europe + Asia-Temperate (n =
289 4117), North America (n = 3307), Asia-Tropical (n = 4071) and South America (n = 4783),
290 using distributional information available in *Plants of the World Online*³² and also *Mabberley's*
291 *plant-book*¹⁵ in the few cases where this information could not be retrieved from the former
292 source. The TDWG regions “Pacific” (minor Pacific islands) and “Antarctic” were disregarded
293 because they showed comparatively lower diversities, and “Europe” and “Asia-Temperate”
294 were merged into one single unit because the taxonomic turnover between the two regions (i.e.
295 β_{sim} distance⁴⁶) was very low (Table S5), meaning that most of the genus-level flora of “Europe”
296 (the less diverse of the two) is shared with that of “Asia-Temperate”. Thus, we finally analyzed
297 six continental datasets separately. We note that widespread genera might not always include
298 useful species across their entire distribution range, which would lead to overestimating the
299 ethnofloras of the regions. Thus, in order to account for this potential bias, we also conducted
300 the continental-scale analyses using only the genera that were endemic to each region (Africa =
301 2294; Australasia = 776; Europe + Asia-Temperate = 1887; North America = 824; Asia-
302 Tropical = 809; South America = 2387).

303

304 **Data availability**

305 The plant-use dataset will be immediately published as an Open Access data paper upon
306 acceptance of this article. All the information included in the dataset was obtained from the
307 fourth edition of *Mabberley's plant-book*.

308

309 **Code availability**

310 All the code used in this research is available as functions that were either implemented in
311 published R packages or provided as supplementary material in a previous Open Access study.

312

313 **References**

- 314 1. D. P. Faith, S. Magallón, A. P. Hendry, E. Conti, T. Yahara, M. J. Donoghue,
315 Evosystem services: an evolutionary perspective on the links between biodiversity and
316 human well-being. *Curr. Opin. Env. Sust.* **2**, 66–74 (2010).
- 317 2. R. Cámara-Leret, S. Faurby, M. J. Macía, H. Balslev, B. Gödel, J.-C. Svenning, W.
318 D. Kissling, N. Rønsted, C. H. Saslis-Lagoudakis, Fundamental species traits explain
319 provisioning services of tropical American palms. *Nat Plants.* **3**, 16220 (2017).
- 320 3. C. Oka, M. Aiba, T. Nakashizuka, Phylogenetic clustering in beneficial attributes of
321 tree species directly linked to provisioning, regulating and cultural ecosystem services.
322 *Ecol. Indic.* **96**, 477–495 (2019).
- 323 4. D. P. Faith, Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–
324 10 (1992).
- 325 5. R. I. Vane-Wright, C. J. Humphries, P. H. Williams, What to protect?—Systematics
326 and the agony of choice. *Biol. Conserv.* **55**, 235–254 (1991).
- 327 6. R. H. Crozier, Genetic diversity and the agony of choice. *Biol. Conserv.* **61**, 11–15
328 (1992).
- 329 7. C. M. Tucker, T. Aze, M. W. Cadotte, J. L. Cantalapiedra, C. Chisholm, S. Díaz, R.
330 Grenyer, D. Huang, F. Mazel, W. D. Pearse, Assessing the utility of conserving
331 evolutionary history. *Biol. Rev.* (2019).
- 332 8. N. R. Owen, R. Gumbs, C. L. Gray, D. P. Faith, Global conservation of phylogenetic
333 diversity captures more than just functional diversity. *Nat. Commun.* **10**, 859 (2019).
- 334 9. F. Mazel, M. W. Pennell, M. W. Cadotte, S. Diaz, G. V. Dalla Riva, R. Grenyer, F.
335 Leprieur, A. O. Mooers, D. Mouillot, C. M. Tucker, W. D. Pearse, Prioritizing

- 336 phylogenetic diversity captures functional diversity unreliably. *Nat. Commun.* **9**
337 (2018), doi:10.1038/s41467-018-05126-3.
- 338 10. F. Mazel, M. W. Pennell, M. W. Cadotte, S. Diaz, G. V. D. Riva, R. Grenyer, F.
339 Leprieur, A. O. Mooers, D. Mouillot, C. M. Tucker, W. D. Pearse, Reply to: “Global
340 conservation of phylogenetic diversity captures more than just functional diversity.”
341 *Nat. Commun.* **10**, 858 (2019).
- 342 11. F. Forest, R. Grenyer, M. Rouget, T. J. Davies, R. M. Cowling, D. P. Faith, A.
343 Balmford, J. C. Manning, Ş. Procheş, M. van der Bank, others, Preserving the
344 evolutionary potential of floras in biodiversity hotspots. *Nature*. **445**, 757–760 (2007).
- 345 12. F. E. M. Cook, Economic Botany Data Collection Standard. International Working
346 Group on Taxonomic Databases for Plant Sciences (TDWG). Royal Botanic Gardens,
347 Kew, UK (1995).
- 348 13. S. A. Smith, J. W. Brown, Constructing a broadly inclusive seed plant phylogeny. *Am.*
349 *J. Bot.* **105**, 302–314.
- 350 14. Y. Jin, H. Qian, V. PhyloMaker: an R package that can generate very large
351 phylogenies for vascular plants. *Ecography*. **42**, 1353–1359 (2019).
- 352 15. D. J. Mabberley, Mabberley’s plant-book: a portable dictionary of plants, their
353 classification and uses. Fourth edition. Cambridge University Press (2017).
- 354 16. P. A. Cox, Will tribal knowledge survive the millennium? *Science*. **287**, 44–45 (2000).
- 355 17. R. Cámara-Leret, N. Paniagua-Zambrana, H. Balslev, M. J. Macía, Ethnobotanical
356 knowledge is vastly under-documented in northwestern South America. *PLOS ONE*.
357 **9**, e85794 (2014).
- 358 18. R. Cámara-Leret, Z. Dennehy, Information gaps in indigenous and local knowledge
359 for science-policy assessments. *Nat. Sustain.* **2**, 736–741 (2019).
- 360 19. V. Novotny, Y. Basset, S. E. Miller, G. D. Weiblen, B. Bremer, L. Cizek, P. Drozd,
361 Low host specificity of herbivorous insects in a tropical forest. *Nature*. **416**, 841–844
362 (2002).
- 363 20. G. S. Gilbert, R. Magarey, K. Suiter, C. O. Webb, Evolutionary tools for
364 phytosanitary risk analysis: phylogenetic signal as a predictor of host range of plant
365 pests and pathogens. *Evol. Appl.* **5**, 869–878 (2012).
- 366 21. J. Calatayud, J. L. Hórreo, J. Madrigal-González, A. Migeon, M. Á. Rodríguez, S.
367 Magalhães, J. Hortal, Geography and major host evolutionary transitions shape the
368 resource use of plant parasites. *PNAS*. **113**, 9840–9845 (2016).
- 369 22. G. T. Pecl, M. B. Araújo, J. D. Bell, J. Blanchard, T. C. Bonebrake, I.-C. Chen, T. D.
370 Clark, R. K. Colwell, F. Danielsen, B. Evengård, L. Falconi, S. Ferrier, S. Frusher, R.
371 A. Garcia, R. B. Griffis, A. J. Hobday, C. Janion-Scheepers, M. A. Jarzyna, S.
372 Jennings, J. Lenoir, H. I. Linnetved, V. Y. Martin, P. C. McCormack, J. McDonald, N.

- 373 J. Mitchell, T. Mustonen, J. M. Pandolfi, N. Pettorelli, E. Popova, S. A. Robinson, B.
374 R. Scheffers, J. D. Shaw, C. J. B. Sorte, J. M. Strugnell, J. M. Sunday, M.-N. Tuanmu,
375 A. Vergés, C. Villanueva, T. Wernberg, E. Wapstra, S. E. Williams, Biodiversity
376 redistribution under climate change: Impacts on ecosystems and human well-being.
377 *Science*. **355** (2017).
- 378 23. P. Lehmann, T. Ammunét, M. Barton, A. Battisti, S. D. Eigenbrode, J. U. Jepsen, G.
379 Kalinkat, S. Neuvonen, P. Niemelä, J. S. Terblanche, B. Økland, C. Björkman,
380 Complex responses of global insect pests to climate warming. *Front. Ecol. Environ.*
381 **18**, 141–150 (2020).
- 382 24. R. F. P. de Lucena, P. M. de Medeiros, E. de L. Araújo, A. G. C. Alves, U. P. de
383 Albuquerque, The ecological apparency hypothesis and the importance of useful
384 plants in rural communities from Northeastern Brazil: An assessment based on use
385 value. *J. Environ. Manag.* **96**, 106–115 (2012).
- 386 25. G. Menendez-Baceta, L. Aceituno-Mata, V. Reyes-García, J. Tardío, M. Salpeteur, M.
387 Pardo-de-Santayana, The importance of cultural factors in the distribution of
388 medicinal plant knowledge: A case study in four Basque regions. *J. Ethnopharmacol.*
389 **161**, 116–127 (2015).
- 390 26. C. O. Webb, D. D. Ackerly, M. A. McPeck, M. J. Donoghue, Phylogenies and
391 Community Ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505 (2002).
- 392 27. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
393 Global. Information on scoping for the thematic assessment of sustainable use of wild
394 species. IPBES (2018). Available at: [https://ipbes.net/sustainable-use-wild-species-](https://ipbes.net/sustainable-use-wild-species-assessment)
395 [assessment](https://ipbes.net/sustainable-use-wild-species-assessment)
- 396 28. M. Pardo-de-Santayana, M. Macía, The benefits of traditional knowledge. *Nature*.
397 **518**, 487–488 (2015).
- 398 29. S. Díaz, U. Pascual, M. Stenseke, B. Martín-López, R. T. Watson, Z. Molnár, R. Hill,
399 K. M. A. Chan, I. A. Baste, K. A. Brauman, S. Polasky, A. Church, M. Lonsdale, A.
400 Larigauderie, P. W. Leadley, A. P. E. van Oudenhoven, F. van der Plaat, M. Schröter,
401 S. Lavorel, Y. Aumeeruddy-Thomas, E. Bukvareva, K. Davies, S. Demissew, G.
402 Erpul, P. Failler, C. A. Guerra, C. L. Hewitt, H. Keune, S. Lindley, Y. Shirayama,
403 Assessing nature’s contributions to people. *Science*. **359**, 270–272 (2018).
- 404 30. A. Antonelli, et al. (205 authors), State of the world’s plants and fungi 2020. Royal
405 Botanic Gardens, Kew (2020).
- 406 31. T. Ulian, M. Diazgranados, S. Pironon, S. Padulosi, U. Liu, L. Davies, M.-J. R.
407 Howes, J. S. Borrell, I. Ondo, O. A. Pérez-Escobar, S. Sharrock, P. Ryan, D. Hunter,
408 M. A. Lee, C. Barstow, Ł. Łuczaj, A. Pieroni, R. Cámara-Leret, A. Noorani, C. Mba,
409 R. N. Womdim, H. Muminjanov, A. Antonelli, H. W. Pritchard, E. Mattana,

- 410 Unlocking plant resources to support food security and promote sustainable
411 agriculture. *Plants, People, Planet*. **2**, 421–445 (2020).
- 412 32. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published
413 on the Internet; <http://www.plantsoftheworldonline.org/>
- 414 33. A. E. Zanne, D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn,
415 D. J. McGlinn, B. C. O'Meara, A. T. Moles, P. B. Reich, D. L. Royer, D. E. Soltis, P.
416 F. Stevens, M. Westoby, I. J. Wright, L. Aarssen, R. I. Bertin, A. Calaminus, R.
417 Govaerts, F. Hemmings, M. R. Leishman, J. Oleksyn, P. S. Soltis, N. G. Swenson, L.
418 Warman, J. M. Beaulieu, Three keys to the radiation of angiosperms into freezing
419 environments. *Nature*. **506**, 89–92 (2014).
- 420 34. The Plant List. Version 1.1. Published on the Internet; <http://www.theplantlist.org/>
- 421 35. T. F. Rangel, R. K. Colwell, G. R. Graves, K. Fučíková, C. Rahbek, J. A. F. Diniz-
422 Filho, Phylogenetic uncertainty revisited: Implications for ecological analyses.
423 *Evolution*. **69**, 1301–1312 (2015).
- 424 36. S. Federhen, The NCBI Taxonomy database. *Nucleic Acids Res.* **40**, D136–D143
425 (2012).
- 426 37. E. Hörandl, T. F. Stuessy, Paraphyletic groups as natural units of biological
427 classification. *Taxon*. **59**, 1641–1653 (2010).
- 428 38. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other
429 things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
- 430 39. A. S. L. Rodrigues, K. J. Gaston, Maximising phylogenetic diversity in the selection
431 of networks of conservation areas. *Biol. Conserv.* **105**, 103–111 (2002).
- 432 40. M. Bordewich, A. G. Rodrigo, C. Semple, Selecting taxa to save or sequence:
433 desirable criteria and a greedy solution. *Syst. Biol.* **57**, 825–834 (2008).
- 434 41. E. C. Pielou, The measurement of diversity in different types of biological collections.
435 *J. Theor. Biol.* **13**, 131–144 (1966).
- 436 42. S. W. Kembel, Disentangling niche and neutral influences on community assembly:
437 assessing the performance of community phylogenetic structure tests. *Ecol. Lett.* **12**,
438 949–960 (2009).
- 439 43. R Core Team. R: A language and environment for statistical computing. (R foundation
440 for statistical computing Vienna, Austria, 2020).
- 441 44. S. W. Kembel, P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D.
442 Ackerly, S. P. Blomberg, C. O. Webb, Picante: R tools for integrating phylogenies and
443 ecology. *Bioinformatics*. **26**, 1463–1464 (2010).
- 444 45. Brummitt, R. K. (2001). World geographical scheme for recording plant distributions
445 (PDF) (2nd ed.). International working group on taxonomic databases for plant
446 sciences (TDWG).

447 46. A. Baselga, Partitioning the turnover and nestedness components of beta diversity.
448 *Glob. Ecol. Biogeogr.* **19**, 134–143 (2010).

449

450

451 **Acknowledgments**

452 We thank the Scientific Computation Center of Andalusia (CICA) for the computing services
453 they provided and Herlander Lima for his assistance in downloading plant distributional
454 information from the web. This work was supported by the Regional Government of the
455 Community of Madrid and the University of Alcalá through the project “Plant evolutionary
456 history and human well-being in a changing world; assessing theoretical foundations using
457 empirical evidence and new phylogenetic tools”, which was granted to RMV (CM/JIN/2019-
458 005). RMV was supported by the TALENTO program of the Regional Government of the
459 Community of Madrid (2018-T2/AMB-10332). MAR was supported by the Ministry of Science
460 and Innovation of Spain (grant CGL2017-86926-P).

461

462 **Author contributions**

463 RMV conceived the ideas, led the assemblage of the plant-use dataset with the help of MPS and
464 DJM, conducted the analyses and led the writing. CR led the assemblage of the continental
465 datasets. MAR suggested using *Mabberley’s plant-book* as the source of global plant-use
466 information. All the authors read, edited and commented on the manuscript.

467

468 **Competing interests**

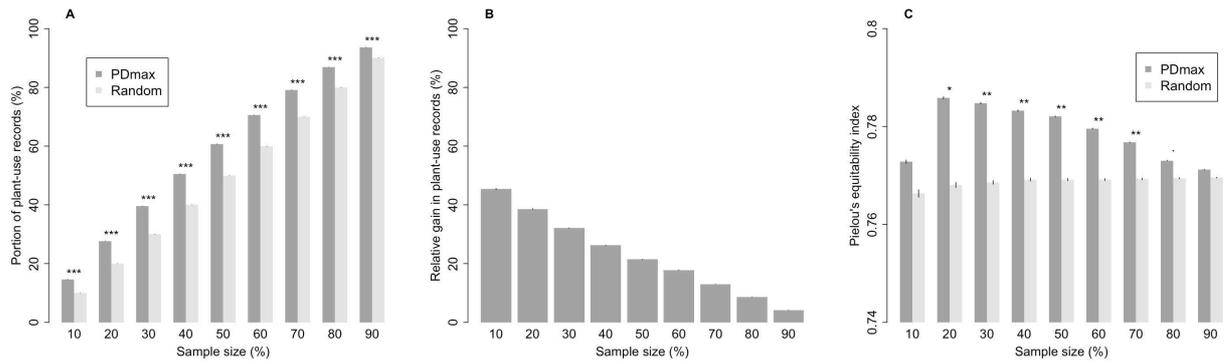
469 The authors declare no competing interests.

470

471 **Additional information**

472 Supplementary information is available for this paper at XXX

473



474

475

476

477

478

479

480

481

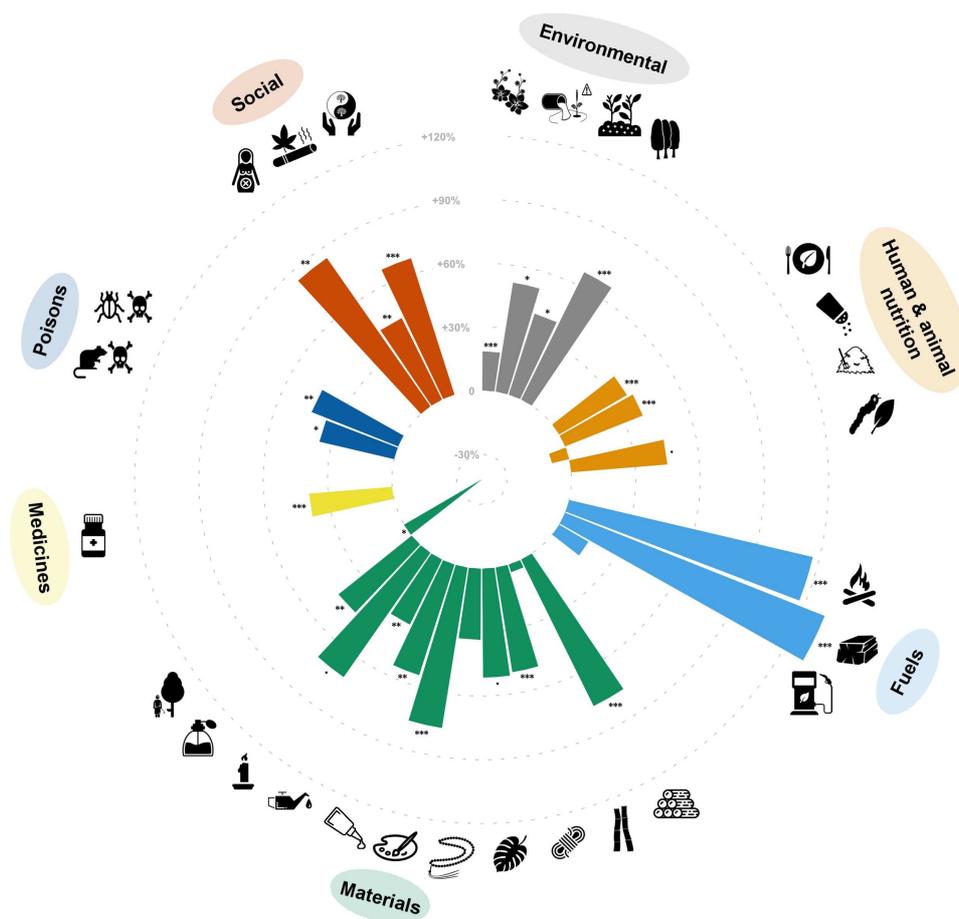
482

483

484

485

Fig. 1. Relative gain in plant-use records and equitability in their distribution among categories. a, Portion of the total pool of plant-use records ($n = 9478$ counted for all use categories combined) retrieved with the PD_{max} and random selection strategies across sample sizes. b, Gain in plant-use records obtained with PD_{max} relative to random selection across sample sizes. c, Equitability (Pielou's evenness index) in the distribution of plant-use records among the 28 categories with PD_{max} and random sampling strategies across sample sizes. The symbols in a and c indicate statistical significance (based on SES scores) for a nominal alpha of 10% “.”, 5% “*”, 1% “**” and 0.1% “***”, respectively (two-tailed tests), and the vertical thin bars at the center of the percentage bars represent confidence intervals at 95%.

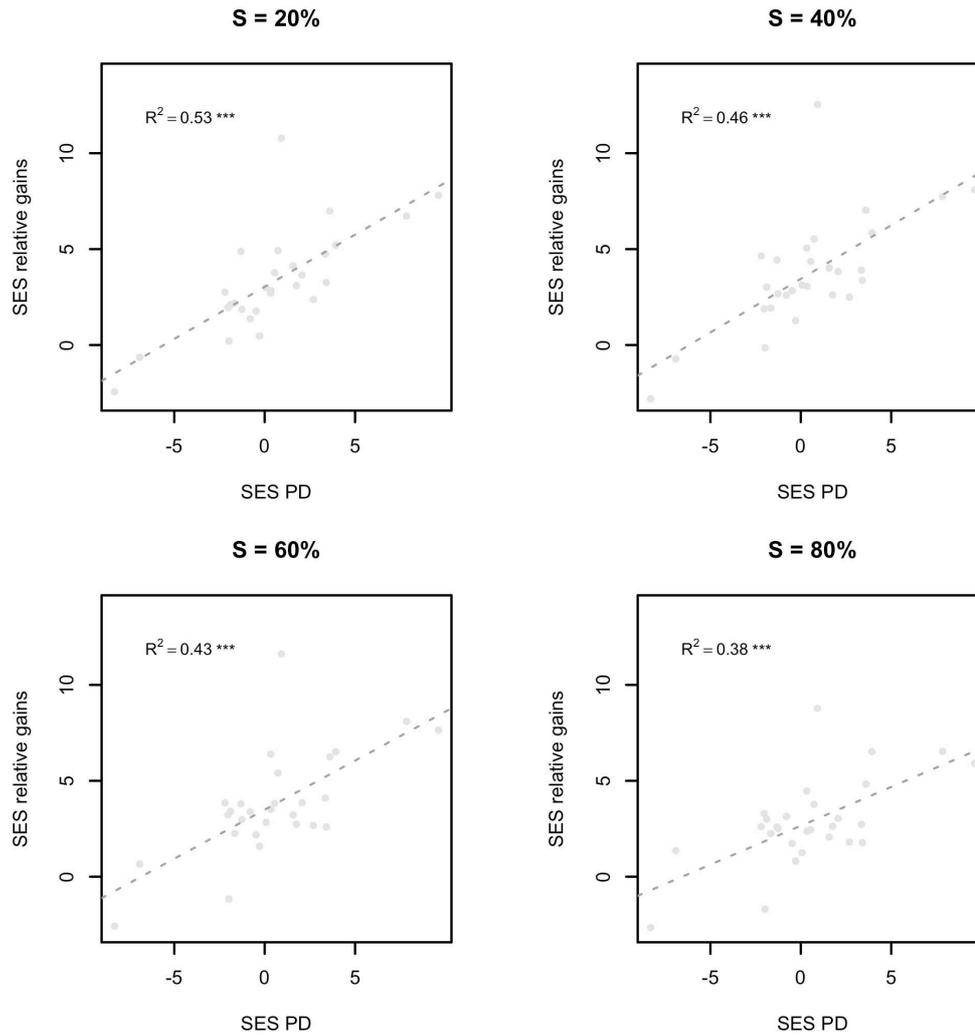


486

487

488 **Fig. 2. Relative gains in plant-use records per category.** The bars represent the relative
 489 gains obtained with PD_{max} relative to random selection at $S = 20\%$ of the total pool of taxa, the
 490 sample size at which the maximum equitability in the distribution of records among use categories
 491 was observed (see Fig. 1c). The symbols on the bars indicate statistical significance (based on
 492 SES scores) for a nominal alpha of 10% “.”, 5% “*”, 1% “***” and 0.1% “****”, respectively (two-
 493 tailed tests). The colours represent different groups of categories following the Economic Botany
 494 Data Collection Standard (see Supplementary Table S1). Note that in order to optimize the
 495 visibility of the figure, values below the -30% threshold do not scale linearly (only one category
 496 with relative gain at -61%). From twelve o’clock and clockwise: (1) ornamental, (2) bioindicators
 497 and bioremediators, (3) soil improvers, (4) hedges and shelters, (5) human food, (6) human food
 498 additives, (7) vertebrate food, (8) invertebrate food, (9) fuelwood, (10) charcoal, (11) biofuels,
 499 (12) timber, (13) stems, (14) fibres, (15) leaves, (16) seeds and fruits, (17) tannins and dyestuffs,
 500 (18) resins and gums, (19) lipids, (20) waxes, (21) scents, (22) rubber, (23) medicines, (24)
 501 vertebrate poisons, (25) invertebrate poisons, (26) antifertility agents, (27) smoking materials and
 502 drugs, (28) symbolism, magic and inspiration.

503



504

505

506 **Fig. 3. Relationship between the phylogenetic structure of plant-use categories and relative**
 507 **gains per category under the PD_{max} strategy.** The dotted lines represent the regression models
 508 between the phylogenetic structure of plant-use categories (SES scores of PD averaged across
 509 100 phylogenetic hypotheses) and SES scores of the relative gains per category across different
 510 sample sizes (S = 20, 40, 60 and 80% of the total pool). All regressions were significant for a
 511 nominal alpha of 0.1%.

512

513

514

Figures

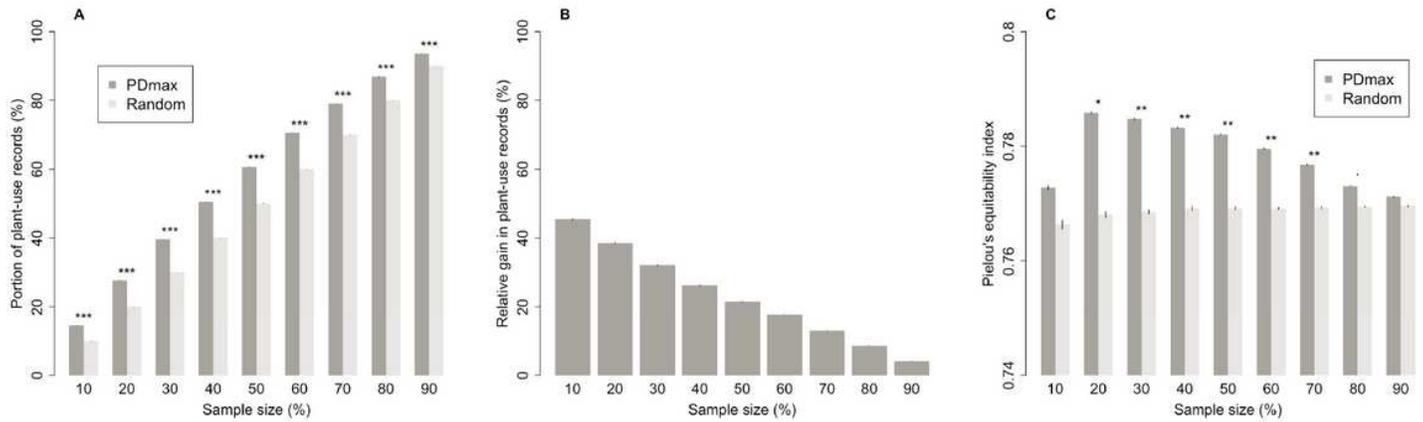


Figure 1

Relative gain in plant-use records and equitability in their distribution among categories. a, Portion of the total pool of plant-use records ($n = 9478$ counted for all use categories combined) retrieved with the PDmax and random selection strategies across sample sizes. b, Gain in plant-use records obtained with PDmax relative to random selection across sample sizes. c, Equitability (Pielou's evenness index) in the distribution of plant-use records among the 28 categories with PDmax and random sampling strategies across sample sizes. The symbols in a and c indicate statistical significance (based on SES scores) for a nominal alpha of 10% ".", 5% "*", 1% "**" and 0.1% "***", respectively (two-tailed tests), and the vertical thin bars at the center of the percentage bars represent confidence intervals at 95%.

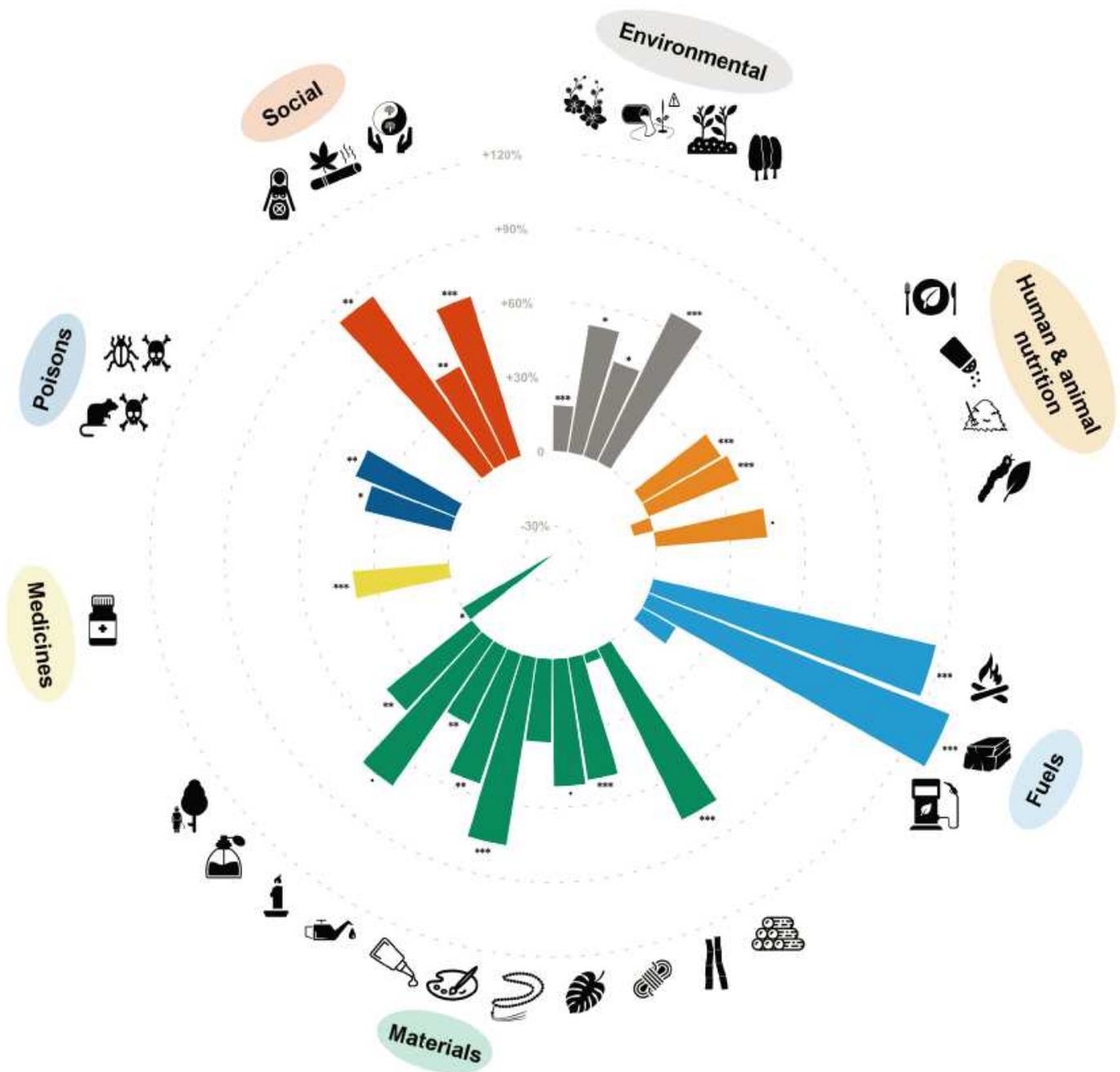


Figure 2

Relative gains in plant-use records per category. The bars represent the relative gains obtained with PDmax relative to random selection at $S = 20\%$ of the total pool of taxa, the sample size at which the maximum equitability in the distribution of records among use categories was observed (see Fig. 1c). The symbols on the bars indicate statistical significance (based on SES scores) for a nominal alpha of 10% “.”, 5% “*”, 1% “**” and 0.1% “***”, respectively (two tailed tests). The colours represent different groups of categories following the Economic Botany Data Collection Standard (see Supplementary Table S1). Note that in order to optimize the visibility of the figure, values below the -30% threshold do not scale linearly (only one category with relative gain at -61%). From twelve o’clock and clockwise: (1) ornamental, (2)

bioindicators and bioremediators, (3) soil improvers, (4) hedges and shelters, (5) human food, (6) human food additives, (7) vertebrate food, (8) invertebrate food, (9) fuelwood, (10) charcoal, (11) biofuels, (12) timber, (13) stems, (14) fibres, (15) leaves, (16) seeds and fruits, (17) tannins and dyestuffs, (18) resins and gums, (19) lipids, (20) waxes, (21) scents, (22) rubber, (23) medicines, (24) vertebrate poisons, (25) invertebrate poisons, (26) antifertility agents, (27) smoking materials and drugs, (28) symbolism, magic and inspiration.

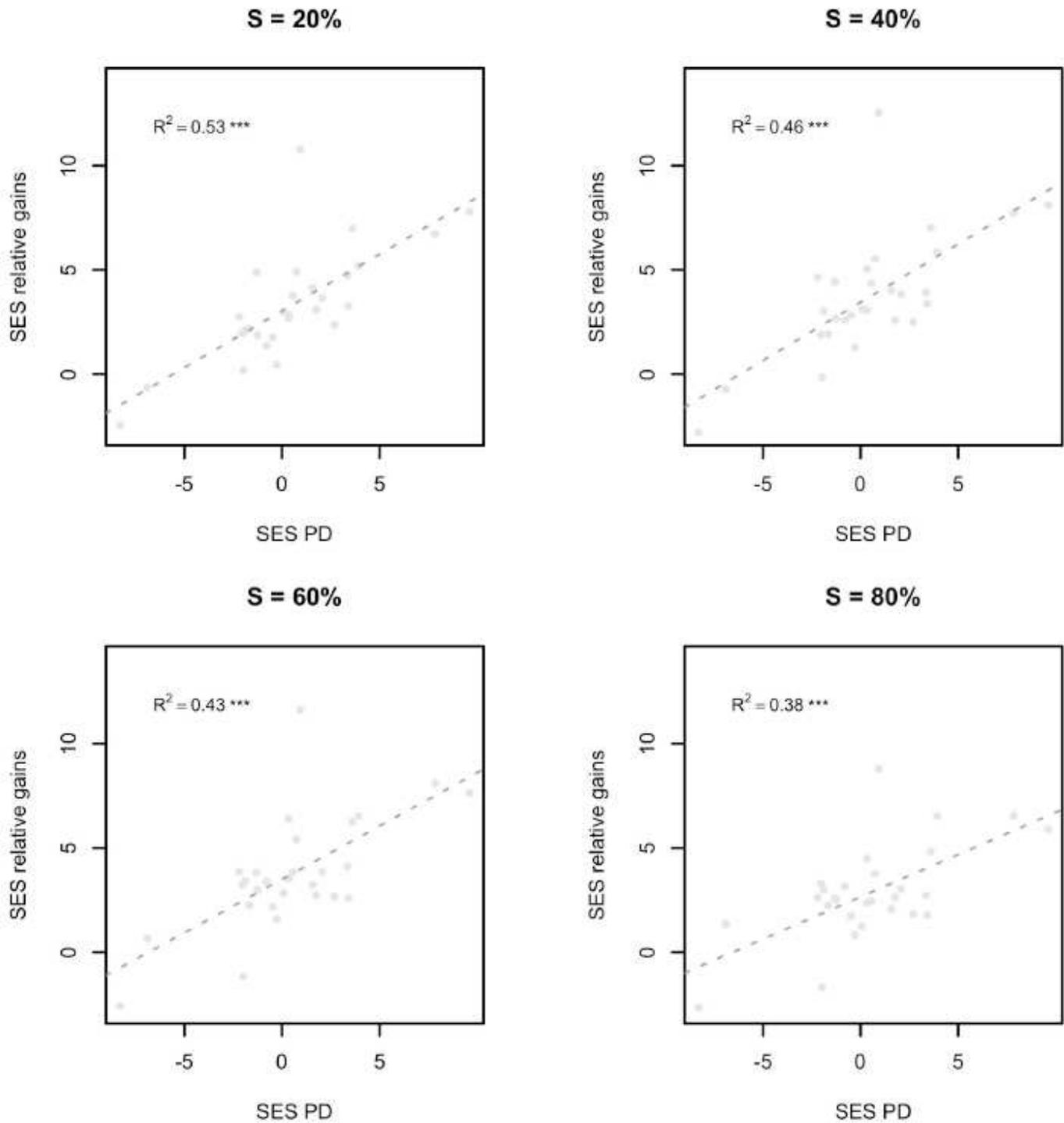


Figure 3

Relationship between the phylogenetic structure of plant-use categories and relative gains per category under the PDmax strategy. The dotted lines represent the regression models between the phylogenetic structure of plant-use categories (SES scores of PD averaged across 100 phylogenetic hypotheses) and SES scores of the relative gains per category across different sample sizes (S = 20, 40, 60 and 80% of the total pool). All regressions were significant for a nominal alpha of 0.1%.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryNEE.pdf](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)