



Plant scientists' research attention is skewed towards colourful, conspicuous and broadly distributed flowers

Martino Adamo¹✉, Matteo Chialva¹, Jacopo Calevo^{2,3}, Filippo Bertoni⁴, Kingsley Dixon³ and Stefano Mammola^{5,6}

Scientists' research interests are often skewed toward charismatic organisms, but quantifying research biases is challenging. By combining bibliometric data with trait-based approaches and using a well-studied alpine flora as a case study, we demonstrate that morphological and colour traits, as well as range size, have significantly more impact on species choice for wild flowering plants than traits related to ecology and rarity. These biases should be taken into account to inform more objective plant conservation efforts.

Throughout human history, plants have played the role of silent partners in the growth of virtually every civilization¹. Humans have exploited wild plants and crops as sources of food², used trees as combustible material and to craft manufactured goods^{3,4} and taken inspiration from the beauty of flowers for poetic and artistic endeavours^{4,5}. Since the birth of modern science, plants have also become the subjects of intense investigation. As scientists systematically studied the natural history of plants⁶, they soon realized that many of these species could function as model organisms to address fundamental scientific questions⁷. Edward O. Wilson famously stated that '[...] for every scientific question, there is the ideal study system to test it' and thus, the choice of a researcher to study one species or another is often driven by functional criteria (for example, ploidy level for genetics studies and ease of growth under controlled conditions). Still, outside of the laboratory or the greenhouse, field scientists may be challenged in their choice of focus organisms by concerns that exceed strictly scientific research interests. As a result, when plant scientists select to study a specific wild plant among the pool of species available in a given study region, it may be that factors unrelated to the biological question end up influencing species choice and introducing biases in the research outcome. Whereas this is not a problem per se, a disparity in scientific attention towards certain species may become a concern in conservation biology, where it is paramount to ensure a 'level playing field' in selecting conservation priorities^{8,9}.

Given their global diversity¹⁰ and ecological importance^{11,12}, plants should be prominent in conservation biology's effort to curb species loss under mounting anthropogenic pressures^{13–15}. Yet, it is well documented that plants receive less attention and consequently less funding in conservation than do animals^{16,17}. This particular case of taxonomic bias has been connected to 'plant blindness'¹⁸ or 'plant awareness disparity'¹⁹, two terms proposed to indicate the lack of awareness for plants. Associated with both the evolutionary

history of human cognition and the effect of cultural, scientific and educational tendencies, this disparity translates into serious real-life impacts, as it affects the knowledge base of conservation and its policies. As addressing this bias is urgent but also often outside the scope of plant sciences, we want to identify more specific biases that can be addressed from within the scientific community dedicated to plants, thereby informing better research practices.

With this goal in mind, we chose a well-defined case study in which to consider specific traits and factors that could influence the choice of species studied. By combining the strengths of bibliometrics and trait-based approaches, we asked what kind of biases might operate in plant sciences, resulting in some species being more studied than others. To resolve this question, we chose a model system of 113 species typical of the Southwestern Alps, one of the largest biodiversity hotspots within the Mediterranean region²⁰ (Fig. 1a). By focusing on a well-known flora in a delimited area, this study design allowed us to control for several confounding factors related to sampling biases, trait heterogeneity and research interest.

We tested whether there is a relationship between research focus on a plant species (measured using bibliometric indicators) and species-specific traits related to ecology, morphology and rarity (Fig. 1a). In the Web of Science, we sourced 280 papers focusing on the selected plant species (average (\pm s.d.) of 2.15 ± 2.96 scientific papers per plant), published between 1975 and 2020. Given that the number of publications, their average annual number of citations and average *h*-index were all reciprocally correlated (all Pearson's $r > 0.7$; Supplementary Fig. 1), we expressed research attention simply as the total number of publications. By means of variance partitioning analysis²¹, we ruled out the relative contribution of ecology, morphology and rarity in determining the observed pattern of research attention. This analysis indicated how the choice of investigated species across the literature in the last 45 years has been strongly influenced by plant traits related to aesthetics. Using marginal R^2 , we observed how morphological and colour traits explain the greater proportion of variance (15.0%), whereas the contribution of ecology and rarity was negligible (Fig. 1b). However, 75.6% of model variance remained unexplained. When reassessing variance partitioning using conditional R^2 , which describe the proportion of variance explained by both fixed effect and random factors (species taxonomic relatedness), we found that 54% of unexplained variance was due to the random effect. This reveals that certain clusters of closely related plants are more studied than others and share

¹Department of Life Sciences and Systems Biology, University of Torino, Torino, Italy. ²Department of Biology, University of Napoli Federico II, Napoli, Italy. ³School of Molecular and Life Sciences, Curtin University, Bentley, Western Australia, Australia. ⁴Humanities of Nature, Museum für Naturkunde, Berlin, Germany. ⁵Molecular Ecology Group, Water Research Institute (IRSA), National Research Council (CNR), Verbania Pallanza, Italy. ⁶Laboratory for Integrative Biodiversity Research, Finnish Museum of Natural History (LUOMUS), University of Helsinki, Helsinki, Finland. ✉e-mail: martino.adamo@unito.it

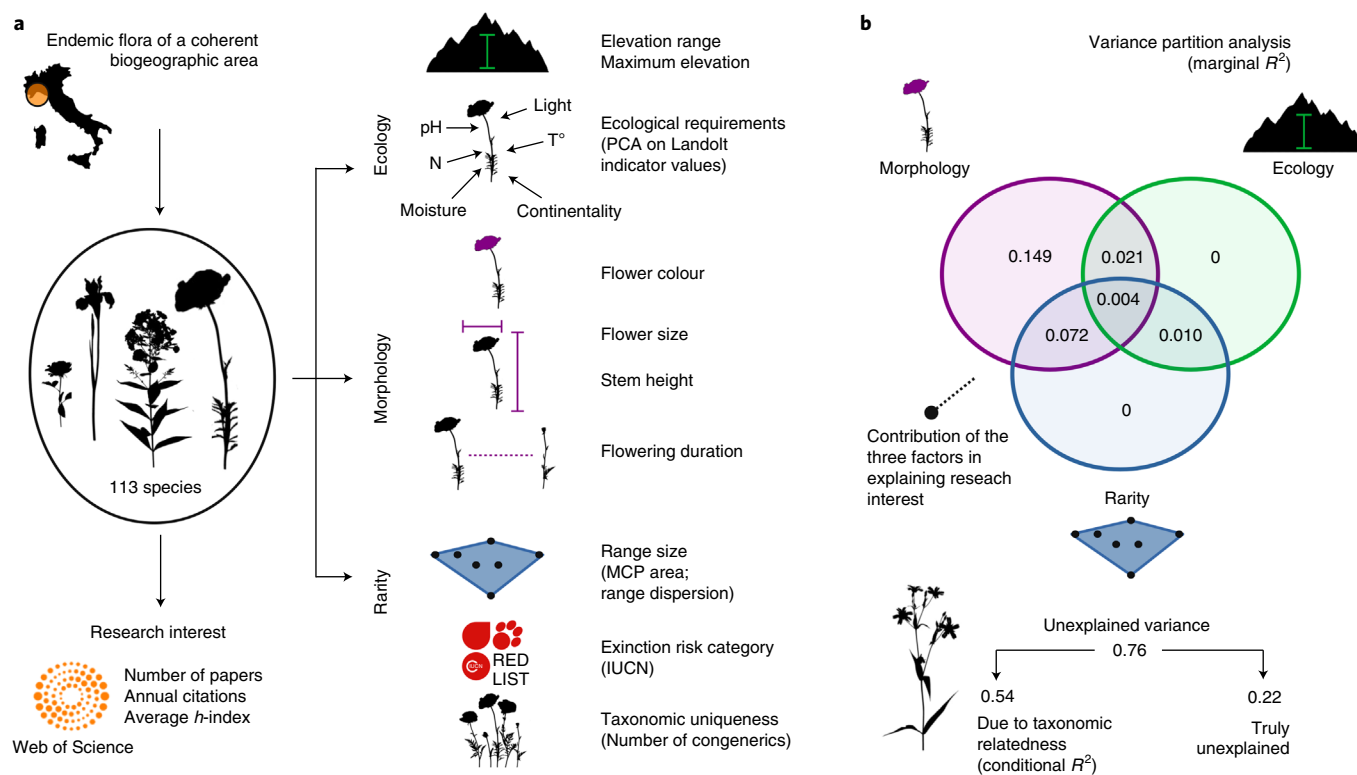


Fig. 1 | Study workflow and most important factors in explaining research interest. **a**, Schematic representation of the data collection and the subdivision of the plant traits in three categories of ecology, morphology and rarity. **b**, Outcomes of the variance partitioning analysis, whereby the relative contribution of traits related to ecology, morphology and rarity is ruled out, as well as the random effect of species' taxonomic relatedness at family level. T°, temperature.

more similar traits than expected from a random pool (examples in Fig. 2a).

This first result was surprising, as species rarity and scientific interest for narrow-range endemics or International Union for Conservation of Nature (IUCN) listed taxa did not emerge as significant drivers. Moreover, a preference for species with particular ecological features seemed likely, as some of the endemic species of the Southwestern Alps are adapted to stressful habitats characterized by a narrow range of environmental conditions, such as rocky lands and xerophilous grasslands^{22,23}. While these adaptations might be desirable for studies on evolution, ecological niche theory, ecophysiology and conservation, the lack of correlation between variables related to 'rarity' and 'ecology' highlights the absence of cross-study guidelines to help plant scientists prioritize such research areas in their choice of species studied.

To obtain a more nuanced understanding of which specific traits are driving research attention, we explored the relationships between traits and number of published papers with a Poisson generalized linear mixed model (GLMM) that accounted for taxonomic non-independence among species²⁴ (Supplementary Table 1). Using backward model selection, we identified a best-performing model that included colour, range size, flower size and stem size as fixed terms (Fig. 2b,c). All other variables introduced in the model had no significant effects and were therefore removed during model selection (Supplementary Table 3). We observed a significant relationship between the number of published papers and flower colour, with blue-coloured flowers being the most studied and white and red/pink significantly more studied than the baseline (brown/green flowers that stand out the least from the environmental background). Moreover, there was a significant positive effect of plant stem height and a (rather weak) negative effect of flower size on research interest. A greater stem height implies that species are more conspicuous

but also taller; thus their inflorescences are more easily accessible without investigators having to stoop to the ground. Furthermore, several plants with small flowers in the Maritime Alps may have intrinsic human appeal, for example flowers constituting conspicuous inflorescences (such as *Gymnadenia conopsea* and *Saxifraga florulenta*) that are more striking than single large flowers, introducing an 'inflorescence effect'. Finally, there was a positive effect of range size on research interest. Tentatively, this is because a broader distribution could make a species accessible to more researchers and thus more likely to be studied. It is interesting to note that, incidentally, broad geographical ranges generally make species less prone to extinction, in line with our finding that species with greater IUCN extinction risk are not subject to more research interest.

The statistical relevance of similar trends across our dataset, where morphological traits such as bright colours, accessible inflorescences and conspicuousness are shown to drive research attention, highlight what we call an aesthetic bias in plant research. While aesthetics is today used to refer to art and beauty (often in direct opposition to scientific values like objectivity), the Greek root of the word refers to sensory perception (as evident in its cognates 'anaesthetic' and 'synaesthetic'). As such, the term highlights sensorial perception, both in its physiological, evolved cognitive structures and in its learned sociocultural articulations. Here it is interesting to note that humans have evolved trichromacy, that is the separate perception of wavelength ranges corresponding to blue, red and green regions through specialized structures²⁵. It has been speculated that the evolutionary acquisition of colour vision in humans and other primates led to an increased ability to locate ripe fruits against a green background^{26,27}. The human eye is thus optimized to perceive green, red and blue which, according to colour psychology theory²⁸, also greatly impacts people's affection, cognition and behaviour. The evolved and physiological aspect of human perception is also

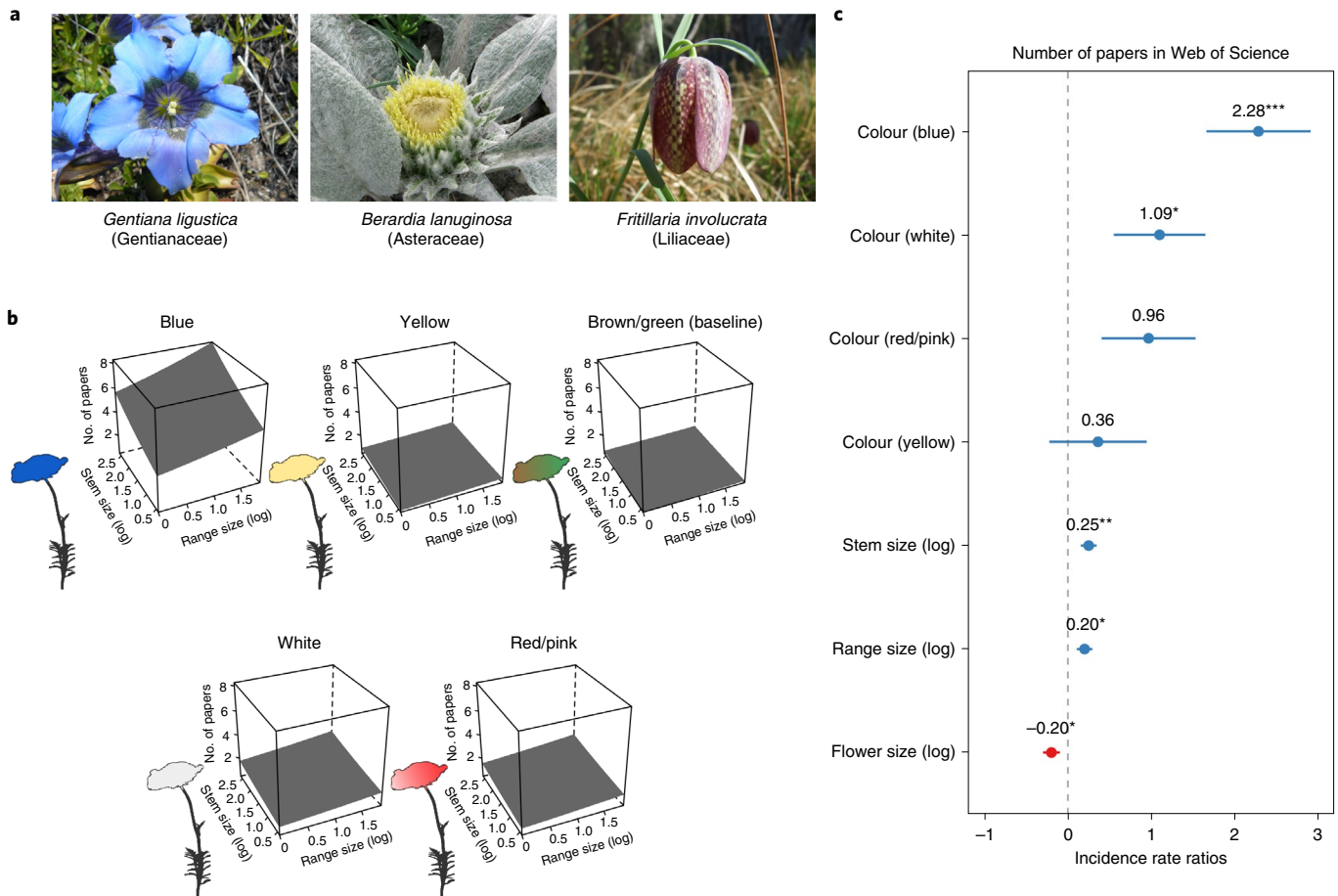


Fig. 2 | Regression analysis of plant traits. a, Example of representative plant species with different traits and research attention: *Gentiana ligustica* R. Vilm. & Chopinet (Gentianaceae, blue inflorescence and many published papers), *Berardia lanuginosa* (Lam.) Fiori (Asteraceae, small and single yellow flower head, short stem and no published papers) and *Fritillaria involucreta* All. (Liliaceae, large brownish flower, short stem and no published papers). **b**, Predicted number of papers in relation to stem and flower size by flower colour according to the best-performing GLMM. **c**, Incidence rate ratios and significance levels ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$) for all the explanatory variables included in the final model (exact P values: colour (blue), 0.00025; colour (white), 0.04562; colour (red/pink), 0.08603; colour (yellow), 0.53695; range size, 0.03308; stem size, 0.00858; flower size, 0.04987). Error bars indicate standard errors. P values for parametric terms were based on two-sided z -test.

demonstrably affected by sociocultural factors, since education, class, gender, age, cultural background all shape how we perceive the world²⁹. Yet, while these above speculations about the origin of the aesthetic bias are interesting, they are beyond the scope of this communication. What matters is that this bias affects the representativity of data used to ground research priorities and conservation policies and, as such, risks compromising efforts to effectively focus plant conservation activities and preserve plant biodiversity.

In conclusion, our analysis identified the traits a plant must possess to be attractive to a scientist, emphasizing the trade-off between aesthetic characteristics, research attention and conservation need. While many factors can determine the choice of studied plant species, we showed how research interests and conservation concerns are less important than aesthetic characteristics in driving research attention. This apparently superficial preference has implicit and undesired effects, as it translates into an aesthetic bias in the data that form the basis for scientific research and practices. Whether this bias is grounded in an evolutionary adaptation of human cognition or in cultural and learned preferences or is simply the effect of practical constraints in the field, it would be desirable to develop measures to counteract it, given the potentially negative impact on our understanding of the ecology and evolution of plants and the conservation of vital plant biodiversity such as species of high

phylogenetic value or with unique ecological traits and ecosystem functions. Statistical modelling has been widely used in conservation ecology to predict ecological niches in space and time and to develop a practical conservation agenda³⁰. Whereas many potential issues, including geographical-relatedness and sampling biases³¹ or metrics selection³², have been routinely considered in modelling exercises, the well-known problem of observer-related biases³³ is largely overlooked³⁴. Against this background, our study demonstrates the need to consider aesthetic biases more explicitly in experimental design and choice of species studied. As Kéry and Greg³⁵ stated: ‘although plants stand still and wait to be counted, they sometimes hide’. Often in plain view, we would add.

Methods

Species selection. We focused the analysis on the flora of the Italian and French Maritime alps, a plant biodiversity hotspot in the Southwestern Alps³⁰. By restricting the analysis to a flora from an intensively studied and confined area, we were able to control for three confounding factors:

- (1) Since the Maritime Alps flora has been extensively studied for over two centuries³⁶, the number of described plant species in this area has already reached the asymptote³⁷ compared with under-studied floras outside Europe³⁸.
- (2) Narrow-range plants on similar substrates and localities are characterized by a restricted range of physicochemical features and would be expected to show similar adaptations. This excludes confounding factors that would occur

if a study was undertaken on species from different biomes and ecological regions.

- (3) Narrow-range species are primarily studied by local researchers (mostly from France, Italy and Switzerland), which are expected to share a similar cultural background and thus share cultural biases. This would not occur in the case of cosmopolitan plants studied by different researchers from mixed cultural backgrounds from around the world.

We selected a representative list of 113 plant species from checklists^{39,40}. For the purpose of this analysis, we excluded subspecies and species of uncertain taxonomic status.

Bibliometric data. We obtained bibliometric data from the Web of Science⁴¹. We searched all published works focusing on each of the 113 species, using the accepted Latin names and synonyms reported in The Plant List⁴². For each species, we derived three values: number of published papers, their average number of citations per year and their average *h*-index. We acknowledge that our search for papers was not exhaustive: we have only included articles in English⁴³, used a single bibliographic database and focused the bibliometric search to the abstract, title and keywords. This implies, for example, that species with no studies in the Web of Science ($n = 43$; 38%) may have actually been the focus of grey literature or of studies that did not mention the Latin name in the abstract of keywords. This is a common practice, for example, in multispecies studies. However, we assumed that this bias was homogeneously distributed across species and thus unlikely to affect the observed patterns.

Species traits. We derived flower colour, stem size, flowering duration and altitude data from Tela Botanica⁴⁴, Actaplantarum (www.actaplantarum.org) and InfoFlora⁴⁵. We obtained species' ecological preferences using Landolt indicator values available in Flora Indiativa⁴⁶. We extracted flower size from FlorAlpes (www.floreAlpes.com/index.php) and conservation status from the IUCN red list⁴⁷. We expressed taxonomic uniqueness of each species as the number of congeneric species, on the basis of ref. ⁴². Finally, we approximated species range size using species occurrences available in the Global Biodiversity Information Facility⁴⁸. We calculated two measures: (1) the area of the minimum convex polygon (MCP) encompassing all localities (range area) and (2) the dispersion of points around the distribution centroid (range dispersion). The latter measure is a more robust measure of range if distribution data are biased, which is often the case with GBIF datasets where sampling effort is uneven (for example, refs. ^{49,50}). We grouped species traits into three categories (Supplementary Table 2 and Fig. 1a): ecology (minimum altitude, altitude range, maximum altitude and Landolt Indexes), morphology (flower colour, flower diameter, stem size and flowering duration) and rarity (range area, range dispersion, IUCN category and number of congeneric species).

Data analysis. We performed all analyses in R⁵¹. We conducted data exploration following ref. ⁵². We checked homogeneity of continuous variables and log₁₀-transformed non-homogeneous variables, when appropriate (Supplementary Table 2). We verified multicollinearity among predictors with pairwise Pearson's *r* correlations (Supplementary Figs. 1 and 3). We visualized potential associations between continuous and categorical variables with boxplots. We summarized the main eight Landolt indicator values variations as the first two principal component (PC) axes of a Principal Component Analysis (PCA), describing the environment in which the different species live. PC1 explained 30.5% of the variance and PC2 explained 17.1% of the variance. We excluded salinity tolerance in the calculation of PCA because it is not applicable in the analysed geographical and ecological context.

IUCN categories were compared with the other 'rarity' variables, revealing a strong association with range area and dispersion (extinction risk is often inferred on the basis of range size⁵³) and a consistent association with the number of congeneric species. Also, collinearity analysis revealed a high correlation ($|r| > 0.7$) between minimum and maximum elevation, and range area and dispersion. We thus excluded the IUCN category, minimum elevation and range area from the analysis. Moreover, to balance the levels of the variable flower colour, we grouped together red with pink and brown with green coloured flowers. The category 'green/brown' was used as a baseline in all analyses, being the least prominent colours from the background^{25,26}.

Variance partitioning analysis. We used variance partitioning analysis²¹ to resolve the relative contribution of ecology, morphology and rarity in determining the observed pattern of research attention. We fitted seven GLMMs (modelling details in the next section), one for each individual set of variables (ecology, morphology and rarity) and their combined effects (ecology + morphology; ecology + rarity; morphology + rarity; ecology + morphology + rarity). In turn, we used the model pseudo *R*² (both conditional and marginal)⁵⁴ to evaluate the contribution of each variable and combination of variables the research attention each species receives, by partitioning their explanatory power using the modEvA⁵⁵ and results visualized as a Venn diagram.

Regression model. We used regressions to explore relationships between the research attention each species receives and plant traits²⁴. Given that number

of published sources, average number of citations and average *h*-index were all reciprocally correlated (Pearson's $r > 0.7$), we only selected the number of publications as a response variable (Supplementary Fig. 1). GLMM with lme4 (ref. ⁵⁶) were fitted to these data using a Poisson distribution and a log link function. The Poisson distribution is often used for count data (in our case, number of papers in the Web of Science) and the log link function ensures positive fitted values²⁴. We scaled all variables and optimized GLMM with bound optimization by quadratic approximation to facilitate model convergence. We used the family taxonomic rank of each plant species as a random factor, to take into account data dependence under the assumption that species within the same family are more likely to share similar traits. Even though 38% of values in the response variable were zeros (that is, species never studied in scientific papers in the Web of Science), zero-inflation was considered as acceptable because these are 'true zeros'⁵⁷.

We built an initial GLMM using all the non-collinear variables and the non-associated factors (Supplementary Table 1) selected after data exploration (the equation is in R notation):

$$\begin{aligned} \text{Number of Papers} \sim & \text{Flower colour} + \text{Flowering duration} + \text{Flower diameter} \\ & + \text{Stem size} + \text{Landolt values PC1} + \text{Landolt values PC2} + \text{Maximum altitude} \\ & + \text{Altitude range} + \text{Range size} + \text{Congeneric species} + \text{random (Family)} \end{aligned}$$

Once the initial model had been fitted, we performed model selection by backward elimination. We based model reduction on Aikake information criterion values (Supplementary Table 3), to simplify the model and avoid overfitting⁵⁸. We validated models with performance⁵⁹ by checking overdispersion and standard residuals plots²⁴ (Supplementary Fig. 2).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data and R script to reproduce the analysis are available in figshare (<https://doi.org/10.6084/m9.figshare.13655456>).

Received: 24 June 2020; Accepted: 1 April 2021;
Published online: 10 May 2021

References

- Schaal, B. Plants and people: our shared history and future. *Plants People Planet* **1**, 14–19 (2019).
- Bates, D. M. People, plants and genes: the story of crops and humanity. *Q. Rev. Biol.* **84**, 206–207 (2009).
- Nedelcheva, A., Dogan, Y., Obratov-Petkovic, D. & Padure, I. M. The traditional use of plants for handicrafts in southeastern Europe. *Hum. Ecol. Interdiscip. J.* **39**, 813–828 (2011).
- Willes, M. A *Shakespearean Botanical* (Bodleian Library, 2015).
- Shoemaker, C. A. Plants and human culture. *J. Home Consum. Hortic.* **1**, 3–7 (1994).
- Alfred, J. & Baldwin, I. T. The natural history of model organisms: new opportunities at the wild frontier. *eLife* **4**, e06956 (2015).
- Hedges, S. B. The origin and evolution of model organisms. *Nat. Rev. Genet.* **3**, 838–849 (2002).
- Clark, J. A. Taxonomic bias in conservation research. *Science* **297**, 191–192 (2002).
- Mammola, S. et al. Towards a taxonomically unbiased European Union biodiversity strategy for 2030. *Proc. R. Soc. B* **287**, 20202166 (2020).
- Christenhusz, M. J. M. & Byng, J. W. The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201–217 (2016).
- Quijas, S., Schmid, B. & Balvanera, P. Plant diversity enhances provision of ecosystem services: a new synthesis. *Basic Appl. Ecol.* **11**, 582–593 (2010).
- Zak, D. R., Holmes, W. E., White, D. C., Peacock, A. D. & Tilman, D. Plant diversity, soil microbial communities, and ecosystem function: are there any links? *Ecology* **84**, 2042–2050 (2003).
- Barnosky, A. D. et al. Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57 (2011).
- Cardinale, B. J. et al. Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
- Ripple, W. J. et al. World scientists' warning to humanity: a second notice. *Bioscience* **67**, 1026–1028 (2017).
- Balding, M. & Williams, K. J. H. Plant blindness and the implications for plant conservation. *Conserv. Biol.* **30**, 1192–1199 (2016).
- Fukushima, C. S., Mammola, S. & Cardoso, P. Global wildlife trade permeates the Tree of Life. *Biol. Conserv.* **247**, 108503 (2020).
- Wandersee, J. H. & Schussler, E. E. Preventing plant blindness. *Am. Biol. Teach.* **61**, 82–86 (1999).
- Parsley, K. M. Plant awareness disparity: a case for renaming plant blindness. *Plants People Planet* **2**, 598–601 (2020).

20. Villemant, C. et al. The Mercantour/Alpi Marittime All Taxa Biodiversity Inventory (ATBI): achievements and prospects. *Zoosystema* **37**, 667–679 (2015).
21. Borcard, D., Legendre, P. & Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–1055 (1992).
22. Médail, F. & Verlaque, R. Ecological characteristics and rarity of endemic plants from Southeast France and Corsica: implications for biodiversity conservation. *Biol. Conserv.* **80**, 269–281 (1997).
23. Noble, V. & Diadema, K. In *La flore des Alpes-Maritimes et de la Principauté de Monaco* (eds Noble, V. & Diadema, K.) 57–72 (Naturalia, 2011).
24. Zuur, A. F. & Ieno, E. N. A protocol for conducting and presenting results of regression-type analyses. *Methods Ecol. Evol.* **7**, 636–645 (2016).
25. Horiguchi, H., Winawer, J., Dougherty, R. F. & Wandell, B. A. Human trichromacy revisited. *Proc. Natl Acad. Sci. USA* **110**, E260–E269 (2013).
26. Gerl, E. J. & Morris, M. R. The causes and consequences of color vision. *Evol. Educ. Outreach* **1**, 476–486 (2008).
27. Bompas, A., Kendall, G. & Sumner, P. Spotting fruit versus picking fruit as the selective advantage of human colour vision. *i-Perception* **4**, 84–94 (2013).
28. Elliot, A. J. & Maier, M. A. Color psychology: effects of perceiving color on psychological functioning in humans. *Annu. Rev. Psychol.* **65**, 95–120 (2014).
29. Chiao, J. Y. et al. Dynamic cultural influences on neural representations of the self. *J. Cogn. Neurosci.* **22**, 1–11 (2010).
30. Cotto, O. et al. A dynamic eco-evolutionary model predicts slow response of alpine plants to climate warming. *Nat. Commun.* **8**, 15399 (2017).
31. Costa, G. C., Nogueira, C., Machado, R. B. & Colli, G. R. Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodivers. Conserv.* **19**, 883–899 (2010).
32. Tucker, C. M. et al. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol. Rev. Camb. Philos. Soc.* **92**, 698–715 (2017).
33. De Boeck, H. J., Liberloo, M., Gielen, B., Nijs, I. & Ceulemans, R. The observer effect in plant science. *New Phytol.* **177**, 579–583 (2008).
34. Morrison, L. W. Observer error in vegetation surveys: a review. *J. Plant Ecol.* **9**, 367–379 (2016).
35. Kéry, M. & Gregg, K. B. Effects of life-state on detectability in a demographic study of the terrestrial orchid *Cleistes bifaria*. *J. Ecol.* **91**, 265–273 (2003).
36. Allioni, C. *Flora Pedemontana: Sive Enumeratio Methodica Stirpium Indigenarum Pedemontii* Vol. 1 (Joannes Michael Briolus, 1785).
37. Aeschimann, D., Rasolofo, N. & Theurillat, J.-P. Analyse de la flore des alpes. 1: Historique et biodiversité. *Candollea* **66**, 27–55 (2011).
38. Boakes, E. H. et al. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol.* **8**, e1000385 (2010).
39. Julve, P. *Baseflor. Index Botanique, Ecologique et Chorologique de la Flore de France* <http://philippe.julve.pagesperso-orange.fr/baseflor.xlsx> (1998).
40. Bartolucci, F. et al. An updated checklist of the vascular flora native to Italy. *Plant Biosyst.* **152**, 179–303 (2018).
41. *Web of Science* (Clarivate Analytics, accessed 16 March 2020); <https://www.webofknowledge.com>
42. Kalwij, J. M. Review of ‘The Plant List, a working list of all plant species’. *J. Veg. Sci.* **23**, 998–1002 (2012).
43. Konno, K. et al. Ignoring non-English-language studies may bias ecological meta-analyses. *Ecol. Evol.* **10**, 6373–6384 (2020).
44. Heaton, L., Millerand, F. & Proulx, S. Tela Botanica: une fertilisation croisée des amateurs et des experts. *Hermès* **57**, 61–68 (2010).
45. Lauber, K., Wagner, G. & Gyğax, A. *Flora Helvetica: Illustrierte Flora der Schweiz* (Haupt Verlag, 2018).
46. Landolt, E. et al. *Flora Indicativa: Okologische Zeigerwerte und Biologische Kennzeichen zur Flora der Schweiz und der Alpen* (Haupt, 2010).
47. *The IUCN Red List of Threatened Species* (IUCN, 2020).
48. *Global Biodiversity Information Facility* (2020); <https://www.gbif.org>
49. Beck, J., Böller, M., Erhardt, A. & Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species’ geographic distributions. *Ecol. Inform.* **19**, 10–15 (2014).
50. Shirey, V., Belitz, M. W., Barve, V. & Guralnick, R. A complete inventory of North American butterfly occurrence data: narrowing data gaps, but increasing bias. *Ecography* **44**, 537–547 (2021).
51. R. Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
52. Zuur, A. F., Ieno, E. N. & Elphick, C. S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **1**, 3–14 (2010).
53. Brooks, T. M. et al. Measuring terrestrial area of habitat (AOH) and its utility for the IUCN Red List. *Trends Ecol. Evol.* **34**, 977–986 (2019).
54. Barton, K. *MuMIn: multi-model inference*. R package version 1.43.17 <https://cran.r-project.org/package=MuMIn> (2020).
55. Barbosa, A. M., Real, R., Munoz, A. R. & Brown, J. A. New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Divers. Distrib.* **19**, 1333–1338 (2013).
56. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1 (2015).
57. Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M. & Castells, E. What does a zero mean? Understanding false, random and structural zeros in ecology. *Methods Ecol. Evol.* **10**, 949–959 (2019).
58. Johnson, J. B. & Omland, K. S. Model selection in ecology and evolution. *Trends Ecol. Evol.* **19**, 101–108 (2004).
59. Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P. & Makowski, D. Assessment, testing and comparison of statistical models using R. Preprint at *PsyArXiv* <https://doi.org/10.31234/osf.io/vtq8f> (2021).

Acknowledgements

We developed this study as a time-burning activity during the COVID-19 Italian lockdown, when laboratory and field work was precluded. We thank D. Fontaneto for useful discussions, P. Bonfante for the critical reading of the manuscript and D. Adamo for their help in data mining. K.D. is funded under the Australian Government through the Australian Research Council Industrial Transformation Training Centre for Mine Site Restoration (project no. CI150100041). S.M. acknowledges support from the European Commission (program H2020-MSCA-IF-2019; grant no. 882221).

Author contributions

M.A. mined and curated data. M.A. and S.M. analysed data and led the paper writing. F.B. assisted with sociological interpretations. M.C. and S.M. prepared figures. M.A., M.C., J.C. and K.D. provided botanical arguments. All authors contributed to the writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41477-021-00912-2>.

Correspondence and requests for materials should be addressed to M.A.

Peer review information *Nature Plants* thanks Paulo Borges, Elena Conti and Sonja Wipf for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Trait data were manually collected by the first author (MA); species distribution data were extracted and calculated using "adehabitatHR" package v(0.1-29) and "rgbif" v(2.2.0) package in R.

Data analysis

Data were analyzed using R v. 4.0.3 and R studio v. 1.4.1103

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings and the R script to generate the analysis are available on figshare database (doi: 10.6084/m9.figshare.13655456)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The study is based on plants typical to the Maritime Alps (n=111). For each plant we derived 16 standardized traits and we extracted all relevant papers in Web of Science.
Research sample	We listed 113 plant species available in the study area using most recent available checklists.
Sampling strategy	We sampled the whole available population, namely the entire set of plants of the study area, which was obtained from the most recent available checklists. The entire population of relevant papers was also extracted from Web of Science.
Data collection	All data were collected online by the first author (MA), using standardized bibliometric procedures. Traits data were compiled by MA, using different standard sources that are detailed in the paper's methods.
Timing and spatial scale	In Web of Science, we sourced all published papers within the available temporal span (1975–2020).
Data exclusions	As a result of data exploration, one outlier was removed from the dataset.
Reproducibility	The study is fully reproducible being based on a standardized search of Web of Science and estimation of traits from different standard sources and a specified accession data (IUCN, gbif, Flora Helvetica, etc.). The data that support the findings and the R script to generate the analysis are available on figshare database (doi: 10.6084/m9.figshare.13655456).
Randomization	Not applicable. The only random structure that could bias the result is the taxonomic rank of species. The Family taxonomic rank of each plant species was used as a random factor in the regression models, to take into account data dependence under the assumption that species within the same family are more likely to share similar traits.
Blinding	Not applicable. The study is based on bibliometric data.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging