

Drivers of species knowledge across the Tree of Life

Stefano Mammola , Martino Adamo, Dragan Antić, Jacopo Calevo, Tommaso Cancellario, Pedro Cardoso, Dan Chamberlain, Matteo Chialva, Furkan Durucan, Diego Fontaneto, Duarte V. Gonçalves, Alejandro Martínez, Luca Santini, Iñigo Rubio-López, Ronaldo Sousa, David Villegas-Ríos, Aida Verdes, Ricardo A. Correia

Reviewed Preprint

Revised by authors after peer review.

About eLife's process

Reviewed preprint version 2

September 21, 2023 (this version)

Reviewed preprint version 1

August 15, 2023

Sent for peer review

April 16, 2023

Posted to bioRxiv

March 27, 2023

Molecular Ecology Group (MEG), Water Research Institute (CNR-IRSA), National Research Council, Verbania Pallanza, Italy • Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History (LUOMUS), University of Helsinki, Helsinki, Finland • National Biodiversity Future Center, Palermo, Italy • Department of Life Sciences and Systems Biology, University of Turin, Torino, Italy • University of Belgrade - Faculty of Biology, Belgrade, Serbia • Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK • School of Molecular and Life Sciences, Curtin University, Bentley, WA 6102, Australia • Department of Aquaculture, Isparta University of Applied Sciences, Isparta, Türkiye • CIIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Matosinhos, Portugal • Department of Biology and Biotechnologies "Charles Darwin", Sapienza University of Rome, Rome, Italy • CBMA - Centre of Molecular and Environmental Biology, Department of Biology, University of Minho, Campus Gualtar, Braga, Portugal • Instituto de Investigaciones Marinas, CSIC, Eduardo Cabello, 6. 36208. Vigo, Pontevedra, Spain • Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales, Madrid, Spain • Helsinki Lab of Interdisciplinary Conservation Science (HELICS), Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland • Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland • CESAM - Centre for Environmental and Marine Studies, University of Aveiro, Aveiro, Portugal • Biodiversity Unit, University of Turku, 20014 Turku, Finland

 https://en.wikipedia.org/wiki/Open_access

 <https://creativecommons.org/licenses/by/4.0/>

Abstract

Knowledge of biodiversity is unevenly distributed across the Tree of Life. In the long run, such disparity in awareness unbalances our understanding of life on Earth, influencing policy decisions and the allocation of research and conservation funding. We investigated how humans accumulate knowledge of biodiversity by searching for consistent relationships between scientific (number of publications) and societal (number of views in Wikipedia) interest, and species-level morphological, ecological and socio-cultural factors. Across a random selection of 3,019 species spanning 29 Phyla/Divisions, we show that socio-cultural factors are the most important correlates of scientific and societal interest in biodiversity, including the fact that a species is useful or harmful to humans, has a common name and is listed in the IUCN Red List. Furthermore, large-bodied, broadly distributed and taxonomically unique species receive more scientific and societal attention, whereas colorfulness and phylogenetic proximity to humans correlates exclusively with societal attention. These results highlight a favoritism towards limited branches of the Tree of Life, and that scientific and societal priorities in biodiversity research broadly align. This suggests that we may be missing out on key species in our research and conservation agenda simply because they are not on our cultural radar.

eLife assessment

With a carefully collected dataset and **compelling** analyses, this **fundamental** manuscript demonstrates detailed links between societal and academic interest and natural species across the globe. In doing so, the authors reveal biases that may be diminishing our abilities to care for the species on our planet that may need our care the most. While some parts of this manuscript reflect previously published work, the authors are commended for putting all the puzzle pieces together for the first time. Their work highlights our uneven knowledge of biodiversity and its potential causes.

Introduction

Human relationships with biodiversity trace back to our dawn as a species (Wilson, 1993). Wildlife permeates art, myths, and traditions, it constitutes an irreplaceable source of food and goods, and, even in the digital age, it remains one of the most powerful triggers of human emotions (Correia and Mammola, 2023 [↗](#); Hicks and Stewart, 2018 [↗](#); Jacobs, 2012 [↗](#); Soga and Gaston, 2016 [↗](#)). Furthermore, the birth of modern science has turned biodiversity into a subject of intense investigation. However, scientific and societal attention towards biodiversity is unevenly distributed across the branches of the Tree of Life (Wilson et al., 2007 [↗](#)). Whether for utilitarian reasons or due to conflictual emotional stimuli (Nyhus, 2016 [↗](#)), we have better knowledge of some species than others (Jarić et al., 2022 [↗](#)).

Widespread evidence indicates that biodiversity research has concentrated on certain lineages, habitats and geographic regions over others (Clark and May, 2002 [↗](#); García-Roselló et al., 2023 [↗](#); Hortal et al., 2015; Mammola et al., 2023 [↗](#); Šmíd, 2022 [↗](#); Troudet et al., 2017 [↗](#)). At the species level, for example, research interests and conservation efforts are often skewed toward vertebrates rather than other animals (Cardoso et al., 2011b [↗](#), 2011a [↗](#); Leather, 2013 [↗](#)), plants (Adamo et al., 2022 [↗](#); Balding and Williams, 2016 [↗](#)) or fungi (Gonçalves et al., 2021 [↗](#); Oyanedel et al., 2022 [↗](#)). Furthermore, scientific and societal attention towards species may correlate, to some degree, with aesthetic features (Adamo et al., 2021 [↗](#); Borgi et al., 2014 [↗](#); Ward et al., 1998 [↗](#)), online popularity (Correia et al., 2016 [↗](#); Mammola et al., 2020 [↗](#)) and phylogenetic proximity to humans (Miralles et al., 2019 [↗](#)), although the relative importance of these factors is likely to vary across cultural settings and societal groups. Indeed, even the selection of model organisms is not always based on functional criteria (e.g., ease of growth under controlled conditions, cell size, genome size, ploidy level; Hedges, 2002 [↗](#)) and instead may be driven by economic, affective, cultural, or other subjective attributes (Dietrich et al., 2020 [↗](#)).

Importantly, most attempts to quantify which features make species attractive to humans have focused on vertebrates—typically mammals and birds (Haukka et al., 2023; Miralles et al., 2019 [↗](#)). This means we now possess a growing understanding of research biases for selected taxa (Guedes et al., 2023; Šmíd, 2022 [↗](#); Sumner et al., 2018 [↗](#); Zvaríková et al., 2021 [↗](#)), but we still lack a comprehensive picture of cross-taxa features that could drive human interest in biodiversity. Here, we explored research and societal interest in organisms across the Tree of Life, asking two general questions: What are the species-level and cultural drivers of scientific interest throughout the Tree of Life? And, how do those drivers differ from those explaining societal interest? To this end, we randomly sampled 3,019 species spanning 29 Phyla and Divisions (Figure S1). We sourced the number of scientific papers focusing on each species as a measure of scientific interest (Figure S2A), and the number of views of the Wikipedia page of each species as a measure of societal interest (Figure S2B). Furthermore, we collected species-level traits referring to morphology and ecology (size, coloration, range size, biome and taxonomic uniqueness) and cultural factors

reflecting how humans perceive and interact with biodiversity (usefulness and harmfulness for humans, presence of a common name in English, phylogenetic distance to humans, IUCN conservation status).

Results

The number of scientific papers focusing on these randomly selected species varied by four orders of magnitude and showed a highly skewed distribution (**Figure 1A**). While 52% of species lacked scientific papers associated with their scientific name in the Web of Science (median $\pm$ S.E. = 0 ± 3.96), there was a long tail of comparatively few species attracting substantial scientific attention (the most studied species in our selection, *Ginkgo biloba* L., appeared in as many as 7,280 scientific papers) (Figure S2A). In contrast, the distribution of the number of views in Wikipedia was less skewed (**Figure 1A**), but there was enormous disparity in societal attention across species ($266 \pm 25,217$; range = 0–50,727,745) (Figure S2B). With the notable exception of Chordata (the Phylum encompassing all vertebrates), most species from other taxonomic groups attracted more scientific interest than expected from societal attention (**Figure 1B**). The few species that attracted disproportionately more societal than scientific attention were colorful, of larger size, and possessed a common name (Figure S3).

Next, we modeled scientific and societal interest in relation to species-level traits and cultural features using generalized linear mixed effects models, controlling for phylogenetic and geographic effects. This analysis revealed a set of drivers that were associated with a high scientific and societal interest (**Figure 2A**; see methods for driver-specific hypotheses), with scientific and societal priorities largely mirroring each other. First, larger species were more attractive to both scientists and the general public. Second, species with broader geographic distributions and taxonomically unique species (i.e., with fewer congeners) all received greater scientific and societal attention. Third, several cultural features strongly correlated with both scientific and societal interest, including the presence of a common name, whether a species is useful and/or harmful for humans, and whether a species had been assessed in the International Union for Conservation of Nature (IUCN) Red List of Threatened Species. Finally, there were three traits uniquely associated with societal interest in organisms: colorful species, freshwater-dwelling species, and species phylogenetically closer to humans all received greater societal attention.

Overall, both models explained ~60% of variance, with an additional ~20% captured by random effects related to taxonomic relatedness and geographic provenance. Using variance partitioning analysis, we compared the relative contribution of morphological, ecological and cultural factors in determining the observed pattern of research and societal attention. Cultural features were the most important in explaining the choice of investigated species across the scientific literature (31% of explained variance) and, to an even greater extent, the number of views on Wikipedia (38%). Species-level traits explained 12% of the variance in the scientific model and 15% of the variance in the societal interest model, whereas both sets of drivers jointly contributed an additional 19 and 16%, respectively, to the two models (**Figure 2B**).

Discussion

We found that the strongest drivers of research and societal interest are utilitarian cultural features, namely whether a species is useful and/or harmful for humans in some way (**Figure 2A**), matching previous evidence based on restricted taxonomic samples. For example, Vardi et al. (2021) showed that in Israel, the most popular plants in terms of online representation often have some use for humans. Similarly, Ladle et al. (2019) found that bird representation online is strongly associated with long histories of human interactions, for example in the form of hunting

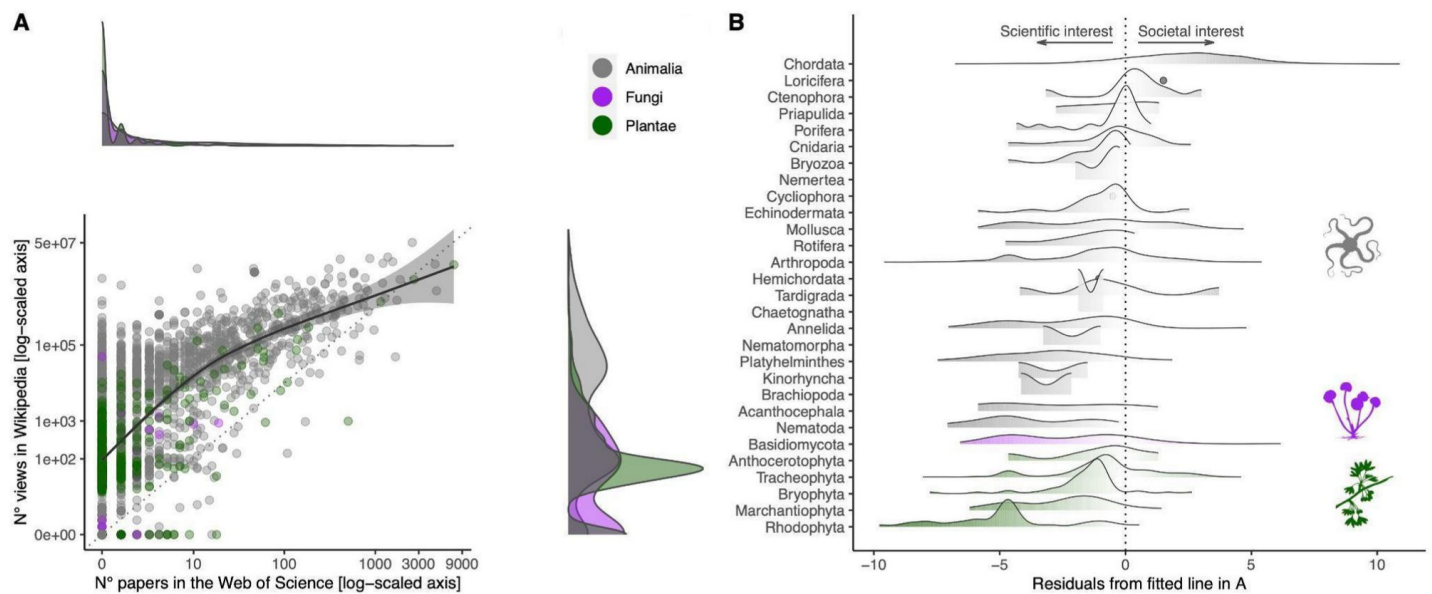


Figure 1.

Relationship between societal and scientific interest across the eukaryotic Tree of Life.

A) Relationship between number of views in Wikipedia (popular interest) and number of papers in the Web of Science (scientific interest) for each species. Both axes are log-scaled to ease visualization. Density functions are provided for both scientific (above scatter plot) and societal interests (right of scatter plot) to illustrate the distribution of values. Color coding refers to the three realms of Animalia, Fungi and Plantae. The regression line is obtained by fitting a Gaussian generalized additive model through the data ($F_{1,3017} = 2497.5$; $p < 0.001$). The farther away a dot is from the fitted line, the more the attention is unbalanced towards either scientific (negative residuals) or societal interest (positive residuals). **B)** Distribution of negative and positive residuals (from the regression line in A) across the species sampled for each Phylum/Division. Phyla/Divisions with only one sampled species are represented with dots.

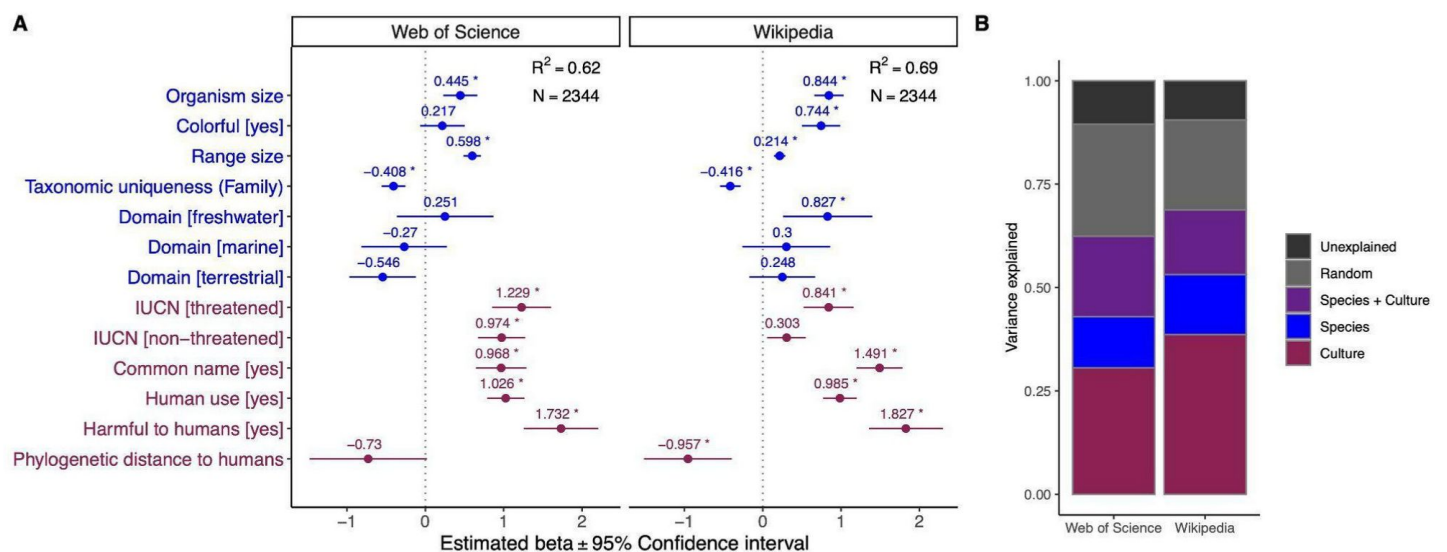


Figure 2.

Influence of species-level traits (blue) and cultural factors (red) on the scientific and societal interest across the eukaryotic Tree of Life.

A) Forest plots summarize the estimated parameters based on Negative binomial generalized linear mixed models (Eq. 1). Baseline levels for multilevel factor variables are: Domain [Multiple] and IUCN [Unknown]. Error bars mark 95% confidence intervals. Variance explained is reported as marginal R^2 , i.e. those explained by fixed factors. Asterisks (*) mark significant effects ($\alpha = 0.01$). Exact estimated regression parameters and p-values are in Tables S1. **B)** Outcomes of the variance partitioning analysis, whereby we partitioned out the relative contribution of species-level traits (blue) and culture factors (red). Joint explained variance (Species + Culture) is highlighted in purple. Unexplained variance is the amount of unexplained variance after considering the contribution of random factors related to species' taxonomy and biogeographic origin (as obtained via conditional R^2).

or pet-keeping. From a cognitive standpoint, an interpretation of this relationship may be rooted in our ancestral past, when we more often relied on wildlife products and we were more frequently subject to predation and other hazards related to wildlife. Experimental evidence suggests that, even in today's society, images of dangerous animals are better able to arouse and maintain human attention (Yorzinski et al., 2014). Interestingly, harmfulness to humans was not a significant driver of scientific and societal interest in Tracheophyta (**Figure 3**). This result may partly be an artifact because plants dangerous to humans are those that are poisonous, but many poisonous plants are contemporary medicinal plants, making it difficult to draw a clear border between usefulness and dangerousness. This is also the case for many poisonous animals, but since vascular plants do not move, the value of their poison as a medicine might overrun our perception of them as a threat. More broadly, many species with deeply-rooted histories of interactions with humans retain their importance in specific cultural contexts, particularly among indigenous peoples, and are thus more likely to remain salient nowadays. Disrupting these connections can have important biocultural consequences, and negatively affect both the species and the communities that value them (Ladle et al. 2023; Reyes-Garcia et al. 2023).

Species with a common name also attracted more scientific and popular interest, matching previous studies (e.g., Vardi et al., 2021). This result should be interpreted with caution, however, because we considered only whether a species possesses an English common name. While we recognize the limitations of this approach, English was selected due to the lack of a comprehensive list of common species names in multiple languages, and because most species that are relevant in other cultural and language settings are also likely to have been attributed English common names as part of legislative, scientific and other societal processes. It must also be noted that this variable entails some circularity, given that humans tend to assign common names to popular species and/or those that are relevant to humans in some way. For example, a recent study showed that across nine local villages in Mozambique, species perceived as dangerous were more likely to have a local name (Farooq et al., 2021). Interestingly, this speaks about the possible existence of specific interactions among different cultural traits and cultural settings which our results do not capture and could be further explored with targeted studies.

The positive effect of body size on scientific and societal interest suggests our attention is likely best captured by organisms with sizes similar (or larger than) our own, rather than organisms that are barely visible. Furthermore, larger species are easier to study and more detectable in the field (Johnston et al., 2014; Kéry and Gregg, 2003). Previous studies documented positive relationships between human interest and body size, e.g., in different vertebrate groups (Berti et al., 2020; Guedes et al., 2023; Ward et al., 1998; Żmihorski et al., 2013) and flowering plants (Adamo et al., 2021), while others observed negative relationships, e.g., in passerine birds (Garrett et al., 2018) and butterflies (Żmihorski et al., 2013). This hints that there may be some within-group variability that is not captured in our broad-scale analysis. However, it is worth noting that most previous studies have focused on organisms that are within the same approximate size range as humans. Indeed, when we repeated regression analyses within subsets of data corresponding to the phylum Chordata, Arthropoda and Tracheophyta, we found that the effect of body size was not significant in Tracheophyta (**Figure 3**). While our random sample of Tracheophyta encompassed an enormous range of sizes—from a duckweed to a sequoia—it may be that attractiveness in plants is primarily controlled by other aesthetic drivers (Adamo et al., 2021).

Different variables reflecting both commonness and rarity contributed markedly in explaining scientific and, to a lesser extent, societal interest. The positive relationship between scientific and societal interest and geographic range size suggests a broader area of distribution could make a species accessible and visible to more people, including researchers, and thus more likely to be studied and searched for in Wikipedia. This result aligns with previous studies observing a positive correlation between proxies of species familiarity and online popularity (Correia et al., 2016; Żmihorski et al., 2013) or scientific interest (Adamo et al., 2021). Furthermore,

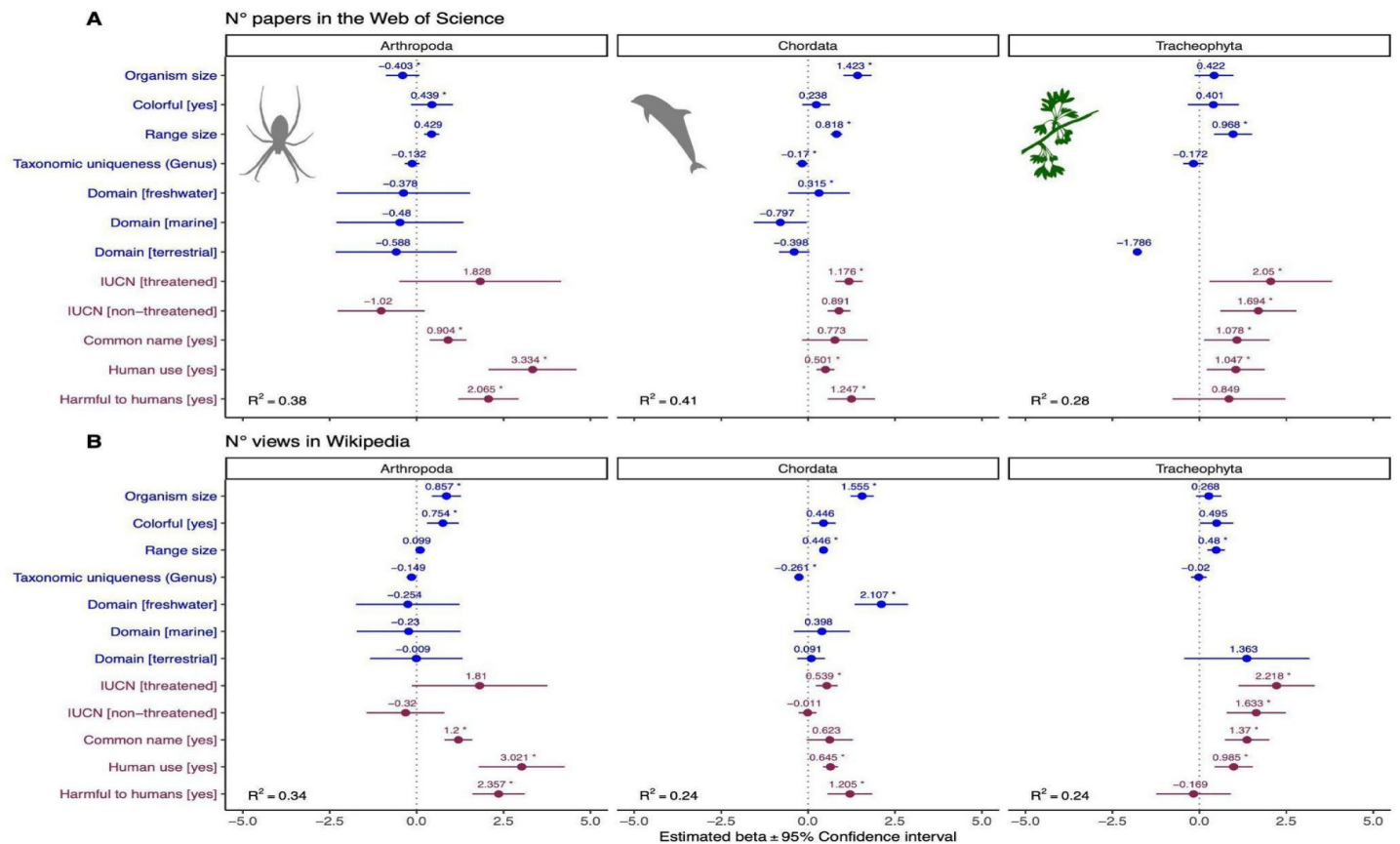


Figure 3.

Influence of species-level traits (blue) and cultural factors (red) on scientific (A) and societal (B) interest for Arthropoda, Chordata, and Tracheophyta.

Forest plots summarize the estimated parameters based on Negative binomial generalized linear mixed models (Eq. 2). Baseline levels for multilevel factor variables are: Domain [Multiple] and IUCN [Unknown]. Error bars mark 95% confidence intervals. Variance explained is reported as marginal R^2 , i.e. those explained by fixed factors. Asterisks (*) mark significant effects ($\alpha = 0.01$). Estimated regression parameters and p-values are in Tables S2 (A) and Table S3 (B).

taxonomically unique species often attracted more scientific and societal interest. These species may represent unique adaptations and phylogenetic distinctiveness and thus be of interest from research or conservation standpoints. Taxonomic uniqueness may also appeal and fascinate the general public, as in famous cases of the discovery of living individuals belonging to taxa previously restricted to the fossil record such as the maidenhair tree (*Ginkgo biloba* L.) or the coelacanth fish (*Latimeria chalumnae* Smith). Conservation rarity, measured as presence and status on the IUCN Red List, was also an important driver of scientific and societal interest. Concerning scientific interest, this was true regardless of the threatened status, namely both endangered and least concern species were more studied and popular across our dataset compared to unlisted species. This variable also entails a certain degree of circularity: IUCN assessments require a lot of data, making it possible to confidently assess species only when there is background information on their distribution and threats.

Finally, colorfulness and phylogenetic proximity to humans correlated exclusively with societal attention. Colorfulness is an important proxy for the aesthetic value of biodiversity (Langlois et al., 2022) and has been shown to often match cultural and economic interests—for example, it was recently shown that colorful birds and fish are more frequently targeted in wildlife trade (Borges et al., 2022 [↗](#); Senior et al., 2022 [↗](#)). Phylogenetic proximity to humans seemingly correlates with a range of traits including the degree of empathy and anthropomorphism toward species. This result resonates with a recent study by Miralles et al. (2019) [↗](#), who used an online survey to assess the empathy of 3500 raters towards 52 taxa (animals, plants and fungi) and observed a strong negative correlation between empathy scores and the divergence time separating the different taxa from *Homo sapiens* (Miralles et al., 2019 [↗](#)). It is more difficult to explain the fact that freshwater-dwelling species were significantly more searched for in Wikipedia than species inhabiting multiple habitats. Speculatively, this may reflect human preference for species inhabiting habitats that are more foreign to human experience, but may also be a sampling artifact (only 103 species in the model, less than 4% of the total, were freshwater-dwelling).

The fact that subjectivity might drive scientific and societal attention towards biodiversity is not a problem *per se*, but, in the long run, it may bias our general understanding of life on Earth to the point of influencing policy decisions and allocation of research and conservation funding. For example, more popular species tend to receive more funding and resources for conservation efforts (Adamo et al., 2022 [↗](#); Davies et al., 2018 [↗](#); Mammola et al., 2020 [↗](#)) and the allocation of protected areas has not adequately considered non-vertebrate species, as up to two-thirds of threatened insect species are not currently covered by existing protected areas (Chowdhury et al., 2023 [↗](#)). This disparity in awareness may also influence species' long-term conservation prospects—a species is less likely to go extinct if humans choose to protect it. Bluntly put, it may be that we are concentrating our attention on species that humans generally consider to be useful, beautiful, or familiar, rather than species that deserve more research effort due to a higher extinction risk and/or due to the key role they play in ecosystem stability and functioning.

Excluding subjectivity when developing any research agenda is certainly challenging. However, once we are aware that utilitarian needs and emotional and familiarity factors play a key role in the development of biodiversity research globally, we can start moving toward more balanced research agendas by carefully selecting which criteria we want to focus on. Ideally, we should aim, over time, for all parameter estimates in **Figure 2A** [↗](#) to move towards the middle (with the possible exceptions of IUCN categories). This strategy would minimize the effect of aesthetic and cultural factors in the selection of research and conservation priorities, and can be achieved over time through a more even repartition of research and conservation funds (see, e.g., Mammola et al., 2020 [↗](#) for a concrete agenda).

Global biodiversity is disappearing at an accelerating pace, not only from the physical world (Barnosky et al., 2011 [↗](#); Cowie et al., 2022 [↗](#)) but also from our minds (Donadio Linares, 2022; Jarić et al., 2022 [↗](#); Soga and Gaston, 2016 [↗](#)). Given that the long-term survival of humanity is

intertwined with the natural world, preserving biodiversity in all its forms and functions (including cultural awareness of it) is a central imperative of the 21st century (Díaz and Malhi, 2022 [↗](#); Jarić et al., 2022 [↗](#); Loreau et al., 2021 [↗](#)). However, biodiversity goals can only be reached by ensuring a ‘level playing field’ in the selection of conservation priorities, rather than looking exclusively at the most appealing branches of the Tree of Life.

Material methods

Species sampling

We carried out random stratified sampling of the eukaryotic multicellular Tree of Life [Animalia, Fungi (restricted to Agaricomycetes), and Plantae (excluding unicellular Algae)] using the Global Biodiversity Information Facility (GBIF) backbone taxonomy (www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c (<http://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c>) [↗](#); accessed on 01 May 2020). To our knowledge, GBIF is the only available backbone taxonomy covering all our target groups using a congruent classification. Note that we restricted our analyses to pluricellular organisms to by-pass issues with the unstable taxonomic classification of protists (Adl et al., 2019 [↗](#), 2012 [↗](#)) and the challenge of extracting comparable traits between unicellular and multicellular eukaryotes.

Initially, we cleaned the GBIF backbone taxonomy by sub-selecting only accepted names (taxonomicStatus = “accepted”), removing subspecies and varieties (taxonRank = “subspecies” and “variety”), and fossil species [by removing both entirely extinct groups (e.g., Dinosauria) and single species labeled as “Fossil_Specimen”]. We chose the following criteria for the stratified random sampling:

- i) The sample was at the species level within each order of Animalia, Fungi and Plantae (this way, we sampled all extant phyla and classes in the database).
- ii) For each order, we sampled a fraction of 0.002 species. To avoid having an excessively uneven number of species among orders, we set the following thresholds:
 - - If the number of species in an order was comprised of between 10,001 and 50,000, we arbitrarily sampled 20 species;
 - - If the number of species in an order was comprised of between 50,001 and 100,000, we arbitrarily sampled 40 species;
 - - If the number of species in an order was >100,000, we arbitrarily sampled 60 species.
- iii) We incorporated a broader sample of tetrapods so as to reflect the typical knowledge bias (“Institutional vertebratism”; Leather, 2013). For each tetrapod order, we arbitrarily sampled 20 species. However, for small tetrapod orders with less than 10 species, we only sampled 1 species. *Institutional vertebratism*; Leather, 2013 [↗](#)). For each tetrapod order, we arbitrarily sampled 20 species. However, for small tetrapod orders with less than 10 species, we only sampled 1 species.

This random sampling procedure yielded a database consisting of 3,019 species (Figure S2C). Despite the initial cleaning procedure of the dataset, due to the fact that some taxonomic names were not properly labeled in GBIF, 129 of the sampled names were synonyms, doubtful (*nomina dubia*) or fossils. We therefore manually inspected all records and dealt with taxonomic issues.

Each expert involved in the study made decisions for their focal organisms on the invalid taxonomic names, e.g., reclassifying subspecies to the species rank, replacing eventual synonyms with the currently valid name, and substituting fossils with extant species.

Measures of scientific and societal interest

We collected data on two indicators of human attention towards species, pertaining to scientific and societal interest.

We measured scientific interest as the number of articles indexed in the *Web of Science* that refer to a given species. This is a standard quantitative estimate of research effort towards individual species (Adamo et al., 2021 [↗](#); dos Santos et al., 2020; Tam et al., 2022 [↗](#); Wilson et al., 2007 [↗](#)). We collected data using the R package ‘wosr’ version 0.3.0 (Baker, 2018 [↗](#)). Specifically, we queried the *Web of Science*’s Core Collection database using topic searches (“TS”) and the species scientific name as the search term, and recorded the total number of references published between 1945 and the date of sampling returned by each query. The use of scientific names returns comparable results to searches using vernacular names (Correia et al., 2017 [↗](#); Jarić et al., 2016 [↗](#)) but avoids common problems associated with vernacular language queries [e.g., words with multiple meanings (homonyms) or used as brand names (theronyms)].

We measured societal interest for each species as the total number of pageviews across the languages where the species is represented on Wikipedia. Wikipedia is one of the top 10 most visited websites in the world (<https://www.similarweb.com/top-websites/> [↗](#), accessed on 3 February 2023) nowadays and is often visited as a source of information for wildlife enthusiasts, many species containing a page in this digital encyclopedia. Wikipedia data has been widely used to explore patterns of popular interest in biodiversity, and total pageviews may be a particularly useful metric in instances where some pages have very few visits overall (Vardi et al., 2021 [↗](#)). To extract the number of pageviews for each species, we first obtained the identification number of each species from the Wikidata knowledge base using the R package ‘WikidataQueryServiceR’ version 1.0.0 (Popov, 2020 [↗](#)). We then used each species’ identifier to compile a list of available Wikipedia pages for the species in any language using the same query service. Once we identified the full list of Wikipedia pages for the species, we used the R package ‘pageviews’ version 0.5.0 (Keyes and Lewis, 2020 [↗](#)) to extract monthly user pageviews (i.e., excluding views by bots) for the period between January 1st 2016 and December 31st 2021.

Species-level traits and associated hypotheses

To investigate the relationship between species-level traits, cultural factors and scientific and popular interest, we selected a set of candidate variables hypothesized to relate to species morphology, ecology and scientific and societal preferences of humans. Extracting comparable traits across distantly related taxa is challenging (Chichorro et al., 2022 [↗](#); Palacio et al., 2022 [↗](#); Weiss and Ray, 2019 [↗](#)), thus we restricted the analysis to a small number of scalable traits and kept trait resolution low (i.e., we scored most traits as categorical variables rather than on continuous scales). Importantly, to ensure cross-taxon comparability of traits, we made specific decisions on how to score traits for the different organisms (details of decisions made and sources of traits are provided in Supplementary Text S1).

Species-level traits

First, we extracted the average body size for each species (in mm). Size is among the most conspicuous and ubiquitous traits in ecology, relating to diverse body functions and ecological strategies (Calder, 1996 [↗](#); Peters and Peters, 1986). Furthermore, we expected an innate preference for large-sized species among scientists, the media and the public alike (Berti et al., 2020 [↗](#); Hall et al., 2011 [↗](#); Mammola et al., 2017 [↗](#); McClain et al., 2015 [↗](#)). We also extracted the

average size of males and females to calculate sexual size dimorphism as a possible driver of interest. However, as sex-specific size values were available for <20% of species in the database, we ended up excluding this variable from analyses.

We also scored, as binary variables (Yes/No), whether individuals within a species are colorful overall (brightly-colored and/or multi-colored species), blue-colored (i.e., when the species has bright blue/light blue markings or overall coloration), or red-colored (when the species has bright red/purple markings or overall coloration). In the case of sexually dichromatic species, we scored these traits as “Yes” even if only one sex displayed colorations. While there are more sophisticated ways to compute color variables (e.g., by extracting RGB pixels from standardized photographs; Delhey et al., 2021 [\[link\]](#)), this was not possible in our case since photographs were available for only 57% of the species included in our database. Given the role of aesthetics in driving human preference across diverse domains (Hoyer and Stokburger-Sauer, 2012 [\[link\]](#)) we hypothesized colorfulness to be a strong driver of attention toward biodiversity (Langlois et al., 2022). Furthermore, we scored red and blue patterns because these colors are known to impact people’s affection, cognition and behavior (Elliot and Maier, 2014 [\[link\]](#)). Recent studies on European plants, for example, have highlighted that species with blue/purple flowers are more frequently studied in the scientific literature (Adamo et al., 2021 [\[link\]](#)) and receive more conservation funds (Adamo et al., 2022 [\[link\]](#)).

For each species, we calculated taxonomic uniqueness as the number of species in the same family (Family uniqueness) or the number of congeneric species (Genus uniqueness). Taxonomic uniqueness may be interesting to scientists and the general public for different reasons. On the one hand, monospecific genera or families may capture divergent phylogenetic lineages defined by the presence of rare or exclusive characters (i.e., unique synapomorphies), and thus be of interest from research or conservation standpoints. On the other hand, families or genera rich in species may be useful as case studies (e.g., to explore evolutionary radiations; Gillespie et al., 2020) or be of interest to the general public simply because of greater accessibility and familiarity.

We marked the main domain inhabited by each species, namely “freshwater”, “marine”, “terrestrial”, or “multiple”. Finally, we used the R package ‘rgbif’ version 3.7.1 (Chamberlain et al., 2022) to extract distribution points for each species. As in Adamo et al. (2021) [\[link\]](#), we expressed the geographical range size of each species as the average distance between occurrence points. This measure (dispersion) is less influenced by sampling effort than commonly used proxies of range size (e.g., minimum convex polygon or the area of occupancy). Hence, it should be better suited when dealing with opportunistically collected occurrence data such as in GBIF (Hughes et al., 2021 [\[link\]](#)). Geographical range size is not only a measure of ecological commonness (Gaston, 2011 [\[link\]](#)), but also reflects species’ accessibility and familiarity to scientists and the general public. Indeed, there is a tendency for humans to be more interested in wildlife species with which they have direct experience (Ladle et al., 2016 [\[link\]](#)), e.g., common species that are available to us through direct experience (Adamo et al., 2021 [\[link\]](#); Schuetz and Johnston, 2019 [\[link\]](#)). Using the GBIF coordinates, we also extracted the coordinate of the centroid of each species’ range, providing a rough indication of their geographic provenance (Figure S1). Using the FADA Faunistic Regions database (Balian et al., 2008) (available at www.marineregions.org (<http://www.marineregions.org/>) [\[link\]](#); accessed on 1 November 2022), we extracted the biogeographic region in which each species occurs (Afrotropical, Antarctic, Australasian, Nearctic, Neotropical, Oriental, Pacific, and Palearctic) based on the centroid coordinates.

Cultural features

To express cultural knowledge and relationships between humans and wildlife, we scored, as binary variables (Yes/No), whether: i) a species has a popular name in English (*Common name*); ii) is an established scientific model organism beyond ecology and evolution (*Model organism*); iii) is harmful to humans in some way—e.g., crop pests, invasive species, species potentially dangerous

to humans (large carnivores, venomous snakes, etc.) (*Harmful to human*); iv) has any commercial and/or cultural use (e.g. used as pets, as food or for pharmaceuticals) (*Human use*); and v) whether it has been assessed by the IUCN. Although we acknowledge that for the variables *Harmful to human* and *Human use* further subcategories could be used (e.g., crop pests, invasive, and harmful to humans may elicit different reactions and interests from a scientific and societal perspective), we decided not to split them due to sample size limitations.

We obtained divergence time (in millions of years) between each organism and *Homo sapiens* from the Time Tree database (Kumar et al., 2022 [link](#)). For this, we used a modified version of the *timetree()* function in the R package ‘timetree’ version 1.0 (<https://github.com/FranzKrah/timetree> [link](#); accessed on 8 November 2021). First, we obtained pairwise divergence time between each taxon and *H. sapiens* by running the function at the genus rank. If the assignment failed, we ran the function iteratively up to the family rank. If still missing, we manually assigned values to the first occurring rank in Time Tree (78 taxa, 2.3% of total). We hypothesize divergence time from *H. sapiens* to be a key factor that may explain human interest in biodiversity (Wilson, 1993), relating to empathy and compassion towards species (Miralles et al., 2019 [link](#)) and the degree of anthropomorphism in human-organism interactions (Servais, 2018 [link](#)).

Finally, we expressed the conservation status of each species as their IUCN extinction risk, which we extracted from the IUCN Red List of Threatened species using the R package ‘rredlist’ version 0.7.0 (Chamberlain, 2022 [link](#)). We assigned each species to one of the following categories: Extinct (EX), Extinct in the Wild (EW), Critically Endangered (CR), Endangered (EN), Vulnerable (VU), Near Threatened (NT), Least Concern (LC), Data Deficient (DD), and Not evaluated (NE). To balance the factor levels, we later re-grouped the different categories into three levels: “Threatened” (EX, EW, CR, EN and VU), “Non-Threatened” (NT and LC), and “Unknown” (DD and NE).

Data analysis

We used regression analyses (Zuur and Ieno, 2016 [link](#)) to test whether there were consistent relationships between scientific (number of scientific papers) and societal (number of views in Wikipedia) interest in an organism and species-level traits and cultural features. We carried out all analyses in R version 4.1.0 (R Core Team, 2021). We used the package ‘glmmTMB’ version 1.1.1 for modeling (Brooks et al., 2017) and ‘ggplot2’ version 3.3.4 (Wickham, 2016 [link](#)) for visualizations. In all analyses, we followed the general approach by Zuur & Ieno (2016) [link](#) for data exploration, model fitting and validation. For data exploration, we visually inspected variable distribution, the presence of outliers, collinearity among continuous predictors (using pairwise Pearson’s correlations) and the balance of factor levels (Zuur et al., 2009 [link](#)). For model validation, we used the suite of functions of the package ‘performance’ version 0.0.0.6 (Lüdtke et al., 2020 [link](#)) to visually inspect model residuals and evaluate overdispersion, zero-inflation and multicollinearity. Given the large sample size of our dataset, we used a conservative approach in the identification of significance, setting an alpha level for significance at 0.01 instead of the usually accepted 0.05 (Benjamin et al., 2018 [link](#)). Furthermore, in interpreting and discussing results, we gave more relevance to explained variance and effect sizes rather than significance (Muff et al., 2021 [link](#)).

In a first set of models, we explored the role of species-level and cultural traits in explaining scientific and popular interest (dependent variables). As a result of data exploration, we log-transformed the variables *Organism size*, *Range size*, *Family uniqueness* and *Phylogenetic distance to humans* to homogenize their distribution and minimize the effect of a few outlying observations. We dropped the categorical variable *Model organism* because it was highly unbalanced—our random sample of species across the Tree of Life only captured 15 species classified as model organisms. Likewise, the variables blue colored and red colored were unbalanced and, to a certain extent, associated with the variable *Colorful*. We used only the latter in the analyses. Finally, we scaled continuous variables to a mean of zero and a standard deviation

of one to facilitate model convergence and interpretation of the effect sizes. We fitted the initial models assuming a Poisson error structure (suitable for count data) and a log-link function (ensuring positive fitted data). The models had the formula (in R notation):

(Eq. 1) $y \sim \text{Organism Size} + \text{Colorful} + \text{Range size} + \text{Domain} + \text{Taxonomic uniqueness} + \text{Common name} + \text{IUCN} + \text{Human use} + \text{Harmful to human} + \text{Phylogenetic distance to humans} + (1 \mid \text{Phylum} / \text{Class} / \text{Order}) + (1 \mid \text{Biogeographic region})$

Where y was either the N° of articles in the Web of Science (*Scientific interest*) or the N° of views in Wikipedia (*Popular interest*). We introduced random factors to take into account the non-independence of observations. We accounted for taxonomic relatedness among species with a nested random intercept structure ($1 \mid \text{Phylum} / \text{Class} / \text{Order}$), under the assumption that closely related species should share more similar traits than would be expected from a random sample of species. Likewise, we used the random intercept structure ($1 \mid \text{Biogeographic region}$) under the assumption that people from the same region, including researchers, might be geographically biased in their interests, i.e., share common appreciation for similar species. Both models were overdispersed (Scientific interest: dispersion ratio = 47.2; Pearson's $\chi^2 = 109874.8$; $p < 0.001$; Popular interest: dispersion ratio = 632366.5; Pearson's $\chi^2 = 1471516950.1$; $p < 0.001$). Therefore, we fitted new models assuming a negative binomial distribution—i.e., a generalization of Poisson distribution that loosens the assumption that the variance should be equal to the mean.

Model validation for the scientific interest model revealed the existence of a highly influential observation corresponding to the Asian elephant (*Elephas maximus* L.). We therefore refitted the model removing this observation, which yielded almost identical model estimates but a better distribution of residuals versus fitted values. Also in the case of the popular interest model, there was a highly influential observation corresponding to the Mugger crocodile [*Crocodylus palustris* (Lesson)], which we removed. Model validation further revealed that the popular interest model was underfitting zeros (Observed zeros: 176; Predicted zeros: 95; Ratio: 0.54), suggesting probable zero-inflation. Therefore, we refitted the model as a standard zero-inflated negative binomial model, using the default “NB2” parameterization implemented in ‘glmmTMB’ (Hardin and Hilbe, 2007). This substantially improved model fit (Akaike Information Criterion of 42727.9 versus 42805.1). No multicollinearity affected either final model, with all Variance Inflation Factors for covariates below 3 (Zuur et al., 2009).

Once the models were fitted and validated, we used variance partitioning analysis (Borcard et al., 1992) to estimate the relative contribution of species-level traits and cultural factors in determining the observed pattern of scientific and societal interest. We used variance explained (marginal R^2) to evaluate the contribution of each variable and combination of variables to the research and societal attention each species receives, by partitioning their explanatory power with the R package ‘modEvA’ version 2.0 (Barbosa et al., 2015).

Next, we tested whether the importance of traits would change across the main groups of organisms by running three models within subsets of data corresponding to Arthropoda, Chordata, and Tracheophyta (i.e., the Phyla/Divisions with most observations). The structure of the models was:

(Eq. 2) $y \sim \text{Organism Size} + \text{Colorful} + \text{Range size} + \text{Domain} + \text{Genus uniqueness} + \text{Common name} + \text{IUCN} + \text{Human use} + \text{Harmful to human} + (1 \mid \text{Class} / \text{Order}) + (1 \mid \text{Biogeographic region})$

The formula is essentially the same as Eq. 1, but for the exclusion of *Phylum* from the random part (as we modeled at the Phylum/Division level) and *Phylogenetic distance to humans* from the fixed part (as we lacked enough resolution in the phylogenetic distance information within Phyla). We

also used Genus uniqueness instead of Family uniqueness given that we modeled at the Phylum level. Also in this case, since Poisson models were overdispersed, we switched to a negative binomial distribution.

Finally, we ran an analysis to understand which species-level traits drive the relative interest of scientists and the general public in different taxa. First, we used a generalized additive model to model the relationship between *Popular interest* and *Scientific interest* (**Figure 1A** [↗](#)). For each species, we extracted the residuals from this regression curve, whereby positive residuals indicate species with a greater popular than scientific interest, residuals close to zero indicate species with a balanced popular and scientific interest, and negative residuals indicate species with a greater scientific than popular interest (**Figure 1B** [↗](#)). Next, we used a Gaussian linear mixed model to model the relationship between the residuals and species-level traits. This model had the same general formula as Eq. 1.

Author contribution

Conceptualization: SM, MA, MC, TC, JC, PC, DF, AM, RC Study design: SM

Data collection (traits): all authors except RC and PC

Data mining (online databases): MC, RC, SM, and TC

Analysis: SM, RC

Interpretation: all authors

Writing, first draft: SM; RC

Writing, contributions: all authors

Acknowledgements

Thanks to Caio Graco-Roza for helping with ggplot2. Filipe Chichorro kindly compiled traits for ants. The authors acknowledge the support of NBFC, funded by the Italian Ministry of University and Research, P.N.R.R., Missione 4 Componente 2, “Dalla ricerca all’impresa”, Investimento 1.4, Project CN00000033.

Conflict of interest

None declared

Data and code availability

The database used in the analyses is available in Figshare (doi: 10.6084/m9.figshare.22731440). The R code to generate analyses and figures is available on GitHub (https://github.com/StefanoMammola/Mammola_et_al_ToL_research_interest [↗](#)).

Supporting information

Table S1–S3

Figure S1–S3

Supplementary text S1

References

1. Adamo M, Chialva M, Calevo J, Bertoni F, Dixon K, Mammola S (2021) **Plant scientists' research attention is skewed towards colourful, conspicuous and broadly distributed flowers** *Nat Plants* **7**:574–578 <https://doi.org/10.1038/s41477-021-00912-2>
2. Adamo M, Sousa R, Wipf S, Correia RA, Lumia A, Mucciarelli M, Mammola S (2022) **Dimension and impact of biases in funding for species and habitat conservation** *Biol Conserv* **272** <https://doi.org/10.1016/j.biocon.2022.109636>
3. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, Cárdenas P, Čepička I, Chistyakova L, del Campo J, Dunthorn M, Edvardsen B, Eglit Y, Guillou L, Hampl V, Heiss AA, Hoppenrath M, James TY, Karnkowska A, Karpov S, Kim E, Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, Le Gall L, Lynn DH, Mann DG, Massana R, Mitchell EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L, Shimano S, Spiegel FW, Torruella G, Youssef N, Zlatogursky V, Zhang Q (2019) **Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes** *J Eukaryot Microbiol* **66**:4–119 <https://doi.org/10.1111/jeu.12691>
4. Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, le Gall L, Lynn DH, McManus H, Mitchell EAD, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick L, Schoch CL, Smirnov A, Spiegel FW (2012) **The Revised Classification of Eukaryotes** *J Eukaryot Microbiol* **59**:429–514 <https://doi.org/10.1111/j.1550-7408.2012.00644.x>
5. Baker C. (2018) **wosr: Clients to the “Web of Science” and “InCites” APIs**
6. Balding M, Williams KJH (2016) **Plant blindness and the implications for plant conservation** *Conserv Biol* **30**:1192–1199 <https://doi.org/10.1111/cobi.12738>
7. Balian E V, Segers H, Martens K, Lévêque C (2008) **An introduction to the freshwater animal diversity assessment (FADA) project** *Hydrobiologia* **600** <https://doi.org/10.1007/s10750-007-9235-6>
8. Barbosa AM, Brown JA, Jimenez-Valverde A, Real A (2015) **modEvA: Model Evaluation and Analysis**
9. Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B, Ferrer EA (2011) **Has the Earth's sixth mass extinction already arrived?** *Nature* **471** <https://doi.org/10.1038/nature09678>
10. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, Bollen KA, Brembs B, Brown L, Camerer C, Cesarini D, Chambers CD, Clyde M, Cook TD, De Boeck P, Dienes Z, Dreber A, Easwaran K, Efferson C, Fehr E, Fidler F, Field AP, Forster M, George EI, Gonzalez R, Goodman S, Green E, Green DP, Greenwald AG, Hadfield JD, Hedges L V, Held L, Hua Ho T, Hoijtink H, Hruschka DJ, Imai K, Imbens G, Ioannidis JPA, Jeon M, Jones JH, Kirchler M, Laibson D, List J, Little R, Lupia A, Machery E, Maxwell SE, McCarthy M, Moore DA, Morgan SL, Munafó M, Nakagawa S, Nyhan B, Parker TH, Pericchi L, Perugini M, Roudier J, Rousseau J, Savalei V, Schönbrodt FD, Sellke T, Sinclair B, Tingley D, Van Zandt T, Vazire S, Watts DJ, Winship C, Wolpert RL, Xie Y, Young C, Zinman J, Johnson VE (2018) **Redefine statistical significance** *Nat Hum Behav* **2**:6–10 <https://doi.org/10.1038/s41562-017-0189-z>

11. Berti E, Monsarrat S, Munk M, Jarvie S, Svenning J-C (2020) **Body size is a good proxy for vertebrate charisma** *Biol Conserv* **251** <https://doi.org/10.1016/j.biocon.2020.108790>
12. Borcard D, Legendre P, Drapeau P (1992) **Partialling out the Spatial Component of Ecological Variation** *Ecology* **73**:1045–1055 <https://doi.org/10.2307/1940179>
13. Borges AKM, Oliveira TPR, Alves RRN (2022) **Marine or freshwater: the role of ornamental fish keeper's preferences in the conservation of aquatic organisms in Brazil** *PeerJ* **10** <https://doi.org/10.7717/peerj.14387>
14. Borgi M, Cogliati-Dezza I, Brelsford V, Meints K, Cirulli F (2014) **Baby schema in human and animal faces induces cuteness perception and gaze allocation in children** *Front Psychol* **5** <https://doi.org/10.3389/fpsyg.2014.00411>
15. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Mächler M, Bolker BM (2017) **glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling** *R J* **9**:378–400 <https://doi.org/10.32614/RJ-2017-066>
16. Calder WA (1996) **Size, function, and life history** *Courier Corporation*
17. Cardoso P, Borges PA V, Triantis KA, Ferrandez MA, Martin JL (2011) **Adapting the IUCN Red List criteria for invertebrates** *Biol Conserv* **144**:2432–2440 <https://doi.org/10.1016/j.biocon.2011.06.020>
18. Cardoso P, Erwin TL, Borges PAV, New TR (2011) **The seven impediments in invertebrate conservation and how to overcome them** *Biol Conserv* **144**:2647–2655 <https://doi.org/10.1016/j.biocon.2011.07.024>
19. Chamberlain S. (2022) **rredlist: “IUCN” Red List Client**
20. Chamberlain S (2022) **Oldoni D** *Waller J*
21. Chichorro F, Urbano F, Teixeira D, Väre H, Pinto T, Brummitt N, He X, Hochkirch A, Hyvönen J, Kaila L, Juslén A, Cardoso P (2022) **Trait-based prediction of extinction risk across terrestrial taxa** *Biol Conserv* **274** <https://doi.org/10.1016/j.biocon.2022.109738>
22. Chowdhury S, Zalucki MP, Hanson JO, Tiatragul S, Green D, Watson JEM, Fuller RA (2023) **Three-quarters of insect species are insufficiently represented by protected areas** *One Earth* <https://doi.org/10.1016/j.oneear.2022.12.003>
23. Clark JA, May RM (2002) **Taxonomic Bias in Conservation Research** *Science* <https://doi.org/10.1126/science.297.5579.191b>
24. Correia RA, Jepson P, Malhado ACM, Ladle RJ (2017) **Internet scientific name frequency as an indicator of cultural salience of biodiversity** *Ecol Indic* **78**:549–555 <https://doi.org/10.1016/j.ecolind.2017.03.052>
25. Correia RA, Jepson PR, Malhado ACM, Ladle RJ (2016) **Familiarity breeds content: assessing bird species popularity with culturomics** *PeerJ* **4** <https://doi.org/10.7717/peerj.1728>
26. Correia RA, Mammola S (2023) **The searchscape of fear: A global analysis of internet search trends for biophobias** *People Nat* <https://doi.org/10.1002/pan3.10497>

27. Cowie RH, Bouchet P, Fontaine B (2022) **The Sixth Mass Extinction: fact, fiction or speculation?** *Biol Rev* **97**:640–663 <https://doi.org/10.1111/brv.12816>
28. Davies T, Cowley A, Bennie J, Leyshon C, Inger R, Carter H, Robinson B, Duffy J, Casalegno S, Lambert G, Gaston K (2018) **Popular interest in vertebrates does not reflect extinction risk and is associated with bias in conservation investment** *PLoS One* **14** <https://doi.org/10.1371/journal.pone.0203694>
29. Delhey K, Dale J, Valcu M, Kempenaers B (2021) **Migratory birds are lighter coloured** *Curr Biol* **31**:R1511–R1512 <https://doi.org/10.1016/j.cub.2021.10.048>
30. Díaz S, Malhi Y (2022) **Biodiversity: Concepts, Patterns, Trends, and Perspectives** *Annu Rev Environ Resour* **47**:31–63 <https://doi.org/10.1146/annurev-environ-120120-054300>
31. Dietrich MR, Ankeny RA, Crowe N, Green S, Leonelli S (2020) **How to choose your research organism** *Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci* **80** <https://doi.org/10.1016/j.shpsc.2019.101227>
32. Linares LM Donadio (2022) **The awkward question: What baseline should be used to measure biodiversity loss? The role of history, biology and politics in setting up an objective and fair baseline for the international biodiversity regime** *Environ Sci Policy* **135**:137–146 <https://doi.org/10.1016/j.envsci.2022.04.019>
33. dos Santos JW, Correia RA, Malhado ACM, Campos-Silva J V, Teles D, Jepson P, Ladle RJ (2020) **Drivers of taxonomic bias in conservation research: a global analysis of terrestrial mammals** *Anim Conserv* **23**:679–688 <https://doi.org/10.1111/acv.12586>
34. Elliot AJ, Maier MA (2014) **Color Psychology: Effects of Perceiving Color on Psychological Functioning in Humans** *Annu Rev Psychol* **65**:95–120 <https://doi.org/10.1146/annurev-psych-010213-115035>
35. Farooq H, Bero C, Guilengue Y, Elias C, Massingue Y, Mucopote I, Nanvonamuquitxo C, Marais J, Antonelli A, Faurby S (2021) **Species perceived to be dangerous are more likely to have distinctive local names** *J Ethnobiol Ethnomed* **17** <https://doi.org/10.1186/s13002-021-00493-6>
36. García-Roselló E, González-Dacosta J, Lobo JM (2023) **The biased distribution of existing information on biodiversity hinders its use in conservation, and we need an integrative approach to act urgently** *Biol Conserv* **283** <https://doi.org/10.1016/j.biocon.2023.110118>
37. Garrett KL, Shuford WD, Gill RE, Handel CM (2018) **Introducing change: A current look at naturalized bird species in western North America** *Trends Tradit Avifaunal Chang West North Am* :116–130
38. Gaston KJ (2011) **Common Ecology** *Bioscience* **61**:354–362 <https://doi.org/10.1525/bio.2011.61.5.4>
39. Gillespie RG, Bennett GM, De Meester L, Feder JL, Fleischer RC, Harmon LJ, Hendry AP, Knope ML, Mallet J, Martin C, Parent CE, Patton AH, Pfennig KS, Rubinoff D, Schluter D, Seehausen O, Shaw KL, Stacy E, Stervander M, Stroud JT, Wagner C, Wogan GOU (2020) **Comparing Adaptive Radiations Across Space, Time, and Taxa** *J Hered* **111**:1–20 <https://doi.org/10.1093/jhered/esz064>
40. Gonçalves S, Haelewaters D, Furci G, Mueller GM (2021) **Include all fungi in biodiversity goals** *Science* <https://doi.org/10.1126/science.abk1312>

41. Guedes JJM, Moura MR (2023) **Species out of sight: elucidating the determinants of research effort in global reptiles** *Ecography (Cop) n/a:e* **6491** <https://doi.org/10.1111/ecog.06491>
42. Hall CM, James M, Baird T (2011) **Forests and trees as charismatic mega-flora: implications for heritage tourism and conservation** *J Herit Tour* **6**:309–323 <https://doi.org/10.1080/1743873X.2011.620116>
43. Hardin JW, Hilbe J (2007) **Generalized linear models and extensions**
44. Haukka A, Lehtikainen A, Mammola S, Morris W, Santangeli A (2023) **The iratebirds Citizen Science Project: a Dataset on Birds' Visual Aesthetic Attractiveness to Humans** *Sci Data* **10**:297 <https://doi.org/10.1038/s41597-023-02169-0>
45. Hedges SB (2002) **The origin and evolution of model organisms** *Nat Rev Genet* **3**:838–849 <https://doi.org/10.1038/nrg929>
46. Hicks JR, Stewart WP (2018) **Exploring potential components of wildlife-inspired awe** *Hum Dimens Wildl* **23**:293–295 <https://doi.org/10.1080/10871209.2018.1419518>
47. Hortal J, de Bello F, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, Ladle RJ (2015) **Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity** *Annu Rev Ecol Evol Syst* **46**:523–549 <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
48. Hoyer WD, Stokburger-Sauer NE (2012) **The role of aesthetic taste in consumer behavior** *J Acad Mark Sci* **40**:167–180 <https://doi.org/10.1007/s11747-011-0269-y>
49. Hughes AC, Orr MC, Ma K, Costello MJ, Waller J, Provoost P, Yang Q, Zhu C, Qiao H (2021) **Sampling biases shape our view of the natural world** *Ecography (Cop)* **44**:1259–1269 <https://doi.org/10.1111/ecog.05926>
50. Jacobs MH (2012) **Human Emotions Toward Wildlife** *Hum Dimens Wildl* **17**:1–3 <https://doi.org/10.1080/10871209.2012.653674>
51. Jarić I, Courchamp F, Gessner J, Roberts DL (2016) **Data mining in conservation research using Latin and vernacular species names** *PeerJ* **4** <https://doi.org/10.7717/peerj.2202>
52. Jarić I, Roll U, Bonaiuto M, Brook BW, Courchamp F, Firth JA, Gaston KJ, Heger T, Jeschke JM, Ladle RJ, Meinard Y, Roberts DL, Sherren K, Soga M, Soriano-Redondo A, Veríssimo D, Correia RA (2022) **Societal extinction of species** *Trends Ecol Evol* **37**:411–419 <https://doi.org/10.1016/j.tree.2021.12.011>
53. Johnston A, Newson SE, Risely K, Musgrove AJ, Massimino D, Baillie SR, Pearce-Higgins JW (2014) **Species traits explain variation in detectability of UK birds** *Bird Study* **61**:340–350 <https://doi.org/10.1080/00063657.2014.941787>
54. Kéry M, Gregg KB (2003) **Effects of life-state on detectability in a demographic study of the terrestrial orchid *Cleistes bifaria*** *J Ecol* **91**:265–273 <https://doi.org/10.1046/j.1365-2745.2003.00759.x>
55. Keyes O, Lewis J. (2020) **pageviews: An API Client for Wikimedia Traffic Data**

56. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB (2022) **TimeTree 5: An Expanded Resource for Species Divergence Times** *Mol Biol Evol* **39** <https://doi.org/10.1093/molbev/msac174>
57. Ladle RJ, Correia RA, Do Y, Joo GJ, Malhado ACM, Proulx R, Roberge JM, Jepson P (2016) **Conservation culturomics** *Front Ecol Environ* **14**:269–275 <https://doi.org/10.1002/fee.1260>
58. Ladle RJ, Jepson P, Correia RA, Malhado ACM (2019) **A culturomics approach to quantifying the salience of species on the global internet** *People Nat* **1**:524–532 <https://doi.org/10.1002/pan3.10053>
59. Ladle RJ, Alves-Martins F, Malhado AC, Reyes-García V, Courchamp F, Di Minin E, Roll U, Jaric I, Correia RA. (2023) **Biocultural aspects of species extinctions** *Cambridge Prisms: Extinction* :1–21 <https://doi.org/10.1017/ext.2023.20>
60. Langlois J, Guilhaumon F, Baletaud F, Casajus N, De Almeida Braga C, Fleuré V, Kulbicki M, Loiseau N, Mouillot D, Renoult JP, Stahl A, Stuart Smith RD, Tribot A-S, Mouquet N (2022) **The aesthetic value of reef fishes is globally mismatched to their conservation priorities** *PLOS Biol* **20** <https://doi.org/10.1371/journal.pbio.3001640>
61. Leather SR (2013) **Institutional vertebratism hampers insect conservation generally; not just saproxylic beetle conservation** *Anim Conserv* **16**:379–380 <https://doi.org/10.1111/acv.12068>
62. Loreau M, Barbier M, Filotas E, Gravel D, Isbell F, Miller SJ, Montoya JM, Wang S, Aussenac R, Germain R, Thompson PL, Gonzalez A, Dee LE (2021) **Biodiversity as insurance: from concept to measurement and application** *Biol Rev* **96**:2333–2354 <https://doi.org/10.1111/brv.12756>
63. Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D (2020) **performance: An R Package for Assessment Comparison and Testing of Statistical Models**. *J Open Source Softw* **6** <https://doi.org/10.21105/joss.03139>
64. Mammola S, Fukushima CS, Biondo G, Bongiorno L, Cianferoni F, Domenici P, Fruciano C, Lo Giudice A, Macías-Hernández N, Malumbres-Olarte J, Miličić M, Morganti M, Mori E, Munévar A, Pollegioni P, Rosati I, Tenan S, Urbano-Tenorio F, Fontaneto D, Cardoso P (2023) **How much biodiversity is concealed in the word ‘biodiversity’?** *Curr Biol* **33**:R59–R60 <https://doi.org/10.1016/j.cub.2022.12.003>
65. Mammola S, Michalik P, Hebets EA, Isaia M (2017) **Record breaking achievements by spiders and the scientists who study them** *PeerJ* **5** <https://doi.org/10.7717/peerj.3972>
66. Mammola S, Riccardi N, Prié V, Correia R, Cardoso P, Lopes-Lima M, Sousa R (2020) **Towards a taxonomically unbiased European Union biodiversity strategy for 2030** *Proc R Soc B Biol Sci* **287** <https://doi.org/10.1098/rspb.2020.2166>
67. McClain CR, Balk MA, Benfield MC, Branch TA, Chen C, Cosgrove J, Dove ADM, Gaskins L, Helm RR, Hochberg FG, Lee FB, Marshall A, McMurray SE, Schanche C, Stone SN, Thaler AD (2015) **Sizing ocean giants: patterns of intraspecific size variation in marine megafauna** *PeerJ* **3** <https://doi.org/10.7717/peerj.715>
68. Miralles A, Raymond M, Lecointre G (2019) **Empathy and compassion toward other species decrease with evolutionary divergence time** *Sci Rep* **9** <https://doi.org/10.1038/s41598-019-56006-9>

69. Muff S, Nilsen EB, O'Hara RB, Nater CR (2021) **Rewriting results sections in the language of evidence** *Trends Ecol Evol* **37**:203–210 <https://doi.org/10.1016/j.tree.2021.10.009>
70. Nyhus PJ (2016) **Human–Wildlife Conflict and Coexistence** *Annu Rev Environ Resour* **41**:143–171 <https://doi.org/10.1146/annurev-environ-110615-085634>
71. Oyanedel R, Hinsley A, Dentinger BTM, Milner-Gulland EJ, Furci G (2022) **A way forward for wild fungi in international sustainability policy** *Conserv Lett* **22** <https://doi.org/10.1111/conl.12882>
72. Palacio FX, Callaghan CT, Cardoso P, Hudgins EJ, Jarzyna MA, Ottaviani G, Riva F, Graco Roza C, Shirey V, Mammola S (2022) **A protocol for reproducible functional diversity analyses** *Ecography* **11** <https://doi.org/10.1111/ecog.06287>
73. Robert H Peters, Henry Peters Robert (1986) **The ecological implications of body size**
74. Popov M. (2020) **WikidataQueryServiceR: API Client Library for 'Wikidata Query Service. R Core Team. 2021.R: A Language and Environment for Statistical Computing**
75. Reyes-García V, Cámara-Leret R, Halpern BS, O'Hara C, Renard D, Zafra-Calvo N, Díaz S (2023) **Biocultural vulnerability exposes threats of culturally important species** *Proc Natl Acad Sci* **120** <https://doi.org/10.1073/pnas.2217303120>
76. Schuetz JG, Johnston A (2019) **Characterizing the cultural niches of North American birds** *Proc Natl Acad Sci* **116**:10868–10873 <https://doi.org/10.1073/pnas.1820670116>
77. Senior RA, Oliveira BF, Dale J, Scheffers BR (2022) **Wildlife trade targets colorful birds and threatens the aesthetic value of nature** *Curr Biol* **32**:4299–4305 <https://doi.org/10.1016/j.cub.2022.07.066>
78. Servais V (2018) **Anthropomorphism in human–animal interactions: A pragmatist view** *Front Psychol* **9** <https://doi.org/10.3389/fpsyg.2018.02590>
79. Šmíd J (2022) **Geographic and taxonomic biases in the vertebrate tree of life** *J Biogeogr* **49**:2120–2129 <https://doi.org/10.1111/jbi.14491>
80. Soga M, Gaston KJ (2016) **Extinction of experience: the loss of human–nature interactions** *Front Ecol Environ* **14**:94–101 <https://doi.org/10.1002/fee.1225>
81. Sumner S, Law G, Cini A (2018) **Why we love bees and hate wasps** *Ecol Entomol* **43**:836–845 <https://doi.org/10.1111/een.12676>
82. Tam J, Lagisz M, Cornwell W, Nakagawa S (2022) **Quantifying research interests in 7,521 mammalian species with h-index: a case study** *Gigascience* **11** <https://doi.org/10.1093/gigascience/giac074>
83. Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F (2017) **Taxonomic bias in biodiversity data and societal preferences** *Sci Rep* **7** <https://doi.org/10.1038/s41598-017-09084-6>
84. Vardi R, Mittermeier JC, Roll U (2021) **Combining culturomic sources to uncover trends in popularity and seasonal interest in plants** *Conserv Biol* **35**:460–471 <https://doi.org/10.1111/cobi.13705>

85. Ward PI, Mosberger N, Kistler C, Fischer O (1998) **The Relationship between Popularity and Body Size in Zoo Animals** *Conserv Biol* **12**:1408–1411 <https://doi.org/10.1111/j.1523-1739.1998.97402.x>
86. Weiss KCB, Ray CA (2019) **Unifying functional trait approaches to understand the assemblage of ecological communities: synthesizing taxonomic divides** *Ecography (Cop* **42**:2012–2020 <https://doi.org/10.1111/ecog.04387>
87. Wickham H (2016) **ggplot2: Elegant Graphics for Data Analysis**
88. Wilson Edward O (1993) **Biophilia and the conservation ethic**. In: Kellert S, Wilson E O, editors. **The Biophilia Hypothesis**. Washington, DC
89. Wilson JRU, Proches S, Braschler B, Dixon ES, Richardson DM (2007) **The (Bio)diversity of Science Reflects the Interests of Society** *Front Ecol Environ* **5**:409–414
90. Yorzinski JL, Penkunas MJ, Platt ML, Coss RG (2014) **Dangerous Animals Capture and Maintain Attention in Humans** *Evol Psychol* **12** <https://doi.org/10.1177/147470491401200304>
91. Żmihorski M, Dziarska-Pałac J, Sparks TH, Tryjanowski P (2013) **Ecological correlates of the popularity of birds and butterflies in Internet information resources** *Oikos* **122**:183–190
92. Zuur AF, Ieno EN (2016) **A protocol for conducting and presenting results of regression-type analyses** *Methods Ecol Evol* **7**:636–645 <https://doi.org/10.1111/2041-210X.12577>
93. Zuur AF, Ieno EN, Elphick CS (2009) **A protocol for data exploration to avoid common statistical problems** *Methods Ecol Evol* **1**:3–14 <https://doi.org/10.1111/j.2041-210x.2009.00001.x>
94. Zvaríková M, Prokop P, Zvarík M, Ježová Z, Medina-Jerez W, Fedor P (2021) **What makes spiders frightening and disgusting to people?** *Front Ecol Evol* **9** <https://doi.org/10.3389/fevo.2021.694569>

Author information

Stefano Mammola

Molecular Ecology Group (MEG), Water Research Institute (CNR-IRSA), National Research Council, Verbania Pallanza, Italy, Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History (LUOMUS), University of Helsinki, Helsinki, Finland, National Biodiversity Future Center, Palermo, Italy

For correspondence: stefano.mammola@cnr.it

ORCID iD: [0000-0002-4471-9055](https://orcid.org/0000-0002-4471-9055)

Martino Adamo

National Biodiversity Future Center, Palermo, Italy, Department of Life Sciences and Systems Biology, University of Turin, Torino, Italy

Dragan Antić

University of Belgrade - Faculty of Biology, Belgrade, Serbia

Jacopo Calevo

Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK, School of Molecular and Life Sciences, Curtin University, Bentley, WA 6102, Australia

Tommaso Cancellario

Molecular Ecology Group (MEG), Water Research Institute (CNR-IRSA), National Research Council, Verbania Pallanza, Italy

Pedro Cardoso

Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History (LUOMUS), University of Helsinki, Helsinki, Finland
ORCID iD: [0000-0001-8119-9960](https://orcid.org/0000-0001-8119-9960)

Dan Chamberlain

Department of Life Sciences and Systems Biology, University of Turin, Torino, Italy

Matteo Chialva

National Biodiversity Future Center, Palermo, Italy, Department of Life Sciences and Systems Biology, University of Turin, Torino, Italy

Furkan Durucan

Department of Aquaculture, Isparta University of Applied Sciences, Isparta, Türkiye

Diego Fontaneto

Molecular Ecology Group (MEG), Water Research Institute (CNR-IRSA), National Research Council, Verbania Pallanza, Italy, National Biodiversity Future Center, Palermo, Italy

Duarte V. Gonçalves

CIIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Matosinhos, Portugal

Alejandro Martínez

Molecular Ecology Group (MEG), Water Research Institute (CNR-IRSA), National Research Council, Verbania Pallanza, Italy
ORCID iD: [0000-0003-0073-3688](https://orcid.org/0000-0003-0073-3688)

Luca Santini

Department of Biology and Biotechnologies “Charles Darwin”, Sapienza University of Rome, Rome, Italy

Iñigo Rubio-López

Molecular Ecology Group (MEG), Water Research Institute (CNR-IRSA), National Research Council, Verbania Pallanza, Italy

Ronaldo Sousa

CBMA - Centre of Molecular and Environmental Biology, Department of Biology, University of Minho, Campus Gualtar, Braga, Portugal

David Villegas-Ríos

Instituto de Investigaciones Marinas, CSIC, Eduardo Cabello, 6. 36208. Vigo, Pontevedra, Spain

Aida Verdes

Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales, Madrid, Spain

Ricardo A. Correia

Helsinki Lab of Interdisciplinary Conservation Science (HELICS), Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland, Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland, CESAM – Centre for Environmental and Marine Studies, University of Aveiro, Aveiro, Portugal, Biodiversity Unit, University of Turku, 20014 Turku, Finland

Editors

Reviewing Editor

David Donoso

Escuela Politécnica Nacional, Ecuador

Senior Editor

George Perry

Pennsylvania State University, United States of America

Reviewer #1 (Public Review):

In the "Drivers of species knowledge across the Tree of Life", Mammola and collaborators explore the determinants of scientific and societal interest across very wide taxonomic and spatial scales. Their work highlights our uneven knowledge of biodiversity and its potential causes.

Reviewer #2 (Public Review):

Using standard and widely used tools, the author revealed the factors (cultural, phenotypic, phylogenetic, etc.) shaping societal and scientific interest in natural species around the globe. The strength of this ms (and the authors) lies in its command of the available literature, database and variable management and analysis, and its solid discussion. The authors thus achieved a manuscript that was pleasant to read.

While I agree that doing a global study requires losing details of local patterns, maybe this is exactly the biggest shortcoming of the manuscript, oblivious to how different cultures (compare USA to PNG, for example) are reflected in these global patterns.

Author Response

The following is the authors' response to the original reviews.

Reviewer #1 (Public Review)

Overall, I quite enjoyed reading the manuscript and found it very well-structured and organized. I congratulate the authors for building this nice research. I do have a few

major points to raise, but probably they would not affect the general message of the manuscript.

Thank you for taking the time to review our manuscript and the positive feedback. Following your suggestions, we have corrected some mistakes and added clarifications and a few of the suggested quality checks on the models. However, we decided not to run new analyses as: i) we believe there would be minor changes to the general message of the manuscript; and ii) while some suggested analyses are compelling, they are difficult to implement for different reasons or are outside the scope of the paper (clarified below).

I was confused about how IUCN data were used. The IUCN predictors are not mentioned in the model equations presented in the manuscript, but their effect size is reported in Figure 2.

Thank you for highlighting this issue. This was a typo: we forgot to mention the variable in both equations 1 and 2. Changed accordingly.

In the manuscript Methods, it is said that IUCN data was classified into 3 categories. I believe there was a mix of mechanisms in measuring it this way since at least two processes might be underlying IUCN data. First, one can inspect whether there is an effect on "scientific/societal interest" for assessed vs non-assessed species. This would not have any relationship with the assessed status itself. Assessed species are any with LC, NT, VU, EN, CR, EW, EX statuses, whereas non-assessed species might include DD and NE. Second, one may observe an effect of threat status itself, with threatened species being more researched than non-threatened species, this would only be possible for assessed species, although there are methods out there to impute missing statuses. By inspecting Figure 2, I got the feeling that only the second option was explored, but this would need to be confirmed.

We couldn't test the effect of single categories (LC, NT, VU, EN, CR, EW, EX) because observations within factor levels were unbalanced. So, we re-grouped the different categories into three levels: "Threatened" (EX, EW, CR, EN and VU), "Non-Threatened" (NT and LC), and "Unknown" (DD and NE) and only tested this variable (your second option). Note that the effect size of the level Unknown is not shown in Figure 2 as this is the reference category. This is clarified in the caption of Figure 2.

In Figure 2, I was confused about the presence of three categories of domain. In the text, it states that four categories have been used. I believe these domains are non-mutually exclusive, that's why there is a fourth category. Would it not be better to assess the influence of domain through three dummy variables (terrestrial, marine, freshwater), where multiple presences (1's) would indicate the "multiple" category?

We opted for a categorical variable (rather than a dummy) to have the same number of variables in the two groups ('species' vs 'culture'). This is needed for the variance partitioning analysis (VPA), because an unbalanced number of variables in one group of a VPA can artificially inflate R² (see, e.g., this source: <https://www.davidzeleny.net/anadat-r/doku.php/en:varpart>). As for Figure 2, the level "Multiple", being the reference category, is not shown. This is clarified in the caption: "Baseline levels for multilevel factor variables are: Domain [Multiple]".

At present, I felt that the spatial components of your data were unexplored. Since you have centroids representing species distribution, it could be interesting to explore the presence of the species within protected areas or biodiversity hotspots. That might be something triggering at least scientific interest. Also, one can derive information about

the major habitat of species occurrence (either using IUCN Major Habitat classification) or extracting overlap of species centroids with WWF biomes (e.g., simplified to just forested vs non-forested habitats; <https://ecoregions.appspot.com/>). Another point very common to research exploring biodiversity shortfalls is the proximity to research institutions (<https://doi.org/10.1111/2041-210X.13152>). And since societal interest is also being explored, what about the proximity to major cities ([doi:10.1038/nature25181](https://doi.org/10.1038/nature25181)). Finally, other metrics derived from species centroids could inform "tropicality", if the species is tropical or not. Most often, the tropics species are neglected in comparison with those occurring in temperate regions.

We thank the reviewer for this suggestion, and we are aware that there are important spatial drivers of interest as highlighted in earlier research. Indeed, the spatial aspects of the data were somewhat underexplored as a deliberate choice because we hope to carry out additional work to explore these aspects in more detail. Nevertheless, we included the centroid of each species range as a broad proxy of its distribution, to help explore, for example, the role of species latitudinal distribution in driving interest metrics. We have also considered the suggestions provided as additional analyses, but we find these may be challenging to implement with the current data for a few reasons. First, each species centroid was calculated based on GBIF occurrences and therefore represents the midpoint of all locations, but not necessarily an area that is known to be occupied by the species. Using the centroid to assess whether a species is located in a given biome or within protected areas using this approach would therefore be potentially misleading (for example, for some terrestrial species it may fall in the sea, and vice versa). Also, for the same reasons, taking the centroid to estimate the species accessibility or proximity to research institutions may be misleading. We find that while important, these spatial aspects require a more nuanced approach to be explored in detail.

I was also thinking about the influence of time on the models. Species described long ago are often more known to people and scientists and had more "time" to be researched. Although metrics of societal interest were restricted to the last decade here, that does not necessarily mean that peoples' interest is not affected by their accumulated experiences. Similar reasoning applies to scientific interests, which have a lengthier time frame (~80 years). That said, the year of description or time since description could be added to capture some metric of time.

This is a good point, which we discussed prior to running the analysis. Indeed, there is evidence that such accumulated experiences can drive species interest as our own research has also previously highlighted (e.g. see Ladle et al. 2017 doi: 10.1002/pan3.10053). However, we felt that comparing the date of description as a proxy of accumulated human experiences with species was only fair within major biological groups and not between them. This is because taxonomic practices, definitions, and methods vary widely between biological groups. We therefore decided not to include time since description as a variable driving the measures of scientific and societal interest in this work. Nevertheless, we recognize the importance of the history of such experiences in driving human interest in species, and the consequences emerging from the loss of such links, and have thus included a brief discussion of this topic in the manuscript (see lines 177-182).

Model residuals could be checked for phylogenetic or spatial autocorrelation. I am aware there is no phylogenetic tree used, but the hierarchical taxonomy could be used (Phylum / Class / Order / Family / Genus) as a proxy for phylogenetic relationship.

We agree. Indeed, the hierarchical taxonomy was already included as a random factor (Phylum / Class / Order) in eq. 1. Note that we excluded Family and Genus from the random

structure because in most Phyla a single genus and family has been sampled (as well as due to model convergence problems).

Concerning the spatial autocorrelation, one could check whether model residuals and their respective coordinate centroids of each species range. It is stated that GLMM has been used to avoid these non-independence issues, but it would be interesting to check whether residuals remained free of them.

Good suggestion, although the use of centroids may not be the most appropriate since it is only a rough approximation of each species distribution (see previous answer). Still, out of curiosity, we checked whether the random factor on biogeography was enough to capture residual spatial autocorrelation in the models. For this, we used the R package DHARMA, which performs a Moran's I test for distance-based autocorrelation. Given that some coordinates were duplicated, we grouped residuals by biogeographic regions (DHARMA requires all coordinates to be unique). Neither the Web of Science nor the Wikipedia models had spatial autocorrelation in the residuals:

Web of Science model: observed = -0.20482, expected = -0.14286, sd = 0.10682, p-value = 0.561

Wikipedia model: observed = -0.180820, expected = -0.142857, sd = 0.055513, p-value = 0.4941

A last point, it would be interesting to provide some sort of inset plots, such as barplots or donut plots (within the current plots), showing the proportion of species with respect to major clades and biogeographical regions.

This is a good suggestion, but we couldn't find a good way to show this as an inset. We added a barchart showing the number of species in each Phyla/Division in the supplementary materials (Figures S2C). As for the proportion of species in each region, we thought it would be redundant with Figure S1 (summarizing spatial information in sampled species).

Reviewer #2 (Public Review):

Using standard and widely used tools, the authors revealed the factors (cultural, phenotypic, phylogenetic, etc.) shaping societal and scientific interest in natural species around the globe. The strength of this manuscript (and the authors') lies in its command of the available literature, database and variable management and analysis, and its solid discussion. The authors thus achieved a manuscript that was pleasant to read.

Thank you for taking the time to review our manuscript and the positive feedback.

While I agree that doing a global study requires losing details of local patterns, maybe this is exactly the biggest shortcoming of the manuscript, oblivious to how different cultures (compare USA to PNG, for example) are reflected in these global patterns.

Related to this previous point, my only other comment is about using English as a reference of societal interest (i.e., the presence of a common name in English). While English may be widespread in Academia, it is still not that common in other societal circles, especially those not using Wikipedia for lack of internet access.

We acknowledge the limitation of this choice, as well as our limited capacity to represent specific cultural contexts with our approach. Our decision to consider only the existence of English common names as a variable was partly driven by practical reasons, and partly by the very factors the reviewer highlights. Indeed, many cultures, communities and social circles do not use English frequently and also do not use the internet frequently. One consequence of this is also that the information compiled for species in other languages is

more restricted than that available in English, including the existence of vernacular names. In languages other than English, it may even be the case that several common language names exist in reference to the same species, and this number may be an even better reflection of their cultural importance, but sadly this information is not comprehensively indexed across languages and biological groups which prevented us from considering it. On the other hand, most species have been attributed English common names as part of legislative, scientific and other societal processes, and it is therefore likely that if they are important in any specific cultural setting, they will probably also have a vernacular English language name. Ultimately, while we recognize the potential limitations of this decision, we felt that considering English common names was the simplest and less biased approach to represent the degree with which a species is individually recognized nowadays. We now better expose the reasons for the decision to consider only English common names, and the limitations associated with it in the manuscript (lines 178-193).