Drivers of species knowledge across the Tree of Life

Stefano Mammola[1,2,3,*], Martino Adamo[4], Dragan Antić[5], Jacopo Calevo[6,7], Tommaso Cancellario[1], Pedro Cardoso[2], Dan Chamberlain[4], Matteo Chialva[4], Furkan Durucan[8], Diego Fontaneto[1,3], Duarte V. Gonçalves[9], Alejandro Martínez[1], Luca Santini[10], Iñigo Rubio-López[1], Ronaldo Sousa[11], David Villegas-Ríos[12], Aida Verdes[13], Ricardo A. Correia[14,15,16]

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

2. Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History (LUOMUS), University of Helsinki, Helsinki, Finland

3. National Biodiversity Future Center, Palermo, Italy

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

5. University of Belgrade - Faculty of Biology, Belgrade, Serbia

6. Curtin University, Perth, Australia

7. Jodrell Laboratory, Kew Gardens, London, UK

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

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

10. Department of Biology and Biotechnologies "Charles Darwin", Sapienza University of Rome, Rome, Italy

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

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

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

14. Helsinki Lab of Interdisciplinary Conservation Science (HELICS), Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

15. Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland

16. CESAM – Centre for Environmental and Marine Studies, University of Aveiro, Aveiro, Portugal

* corresponding author: stefano.mammola@cnr.it ; stefano.mammola@helsinki.fi

Short title: Knowledge biases across the Tree of Life

Keywords: Biodiversity, Biological conservation, Culturomics, Research bias, Wildlife

Email & ORCID

Stefano Mammola: stefano.mammola@cnr.it ; ORCID: 0000-0002-4471-9055

Martino Adamo: martino.adamo@unito.it; ORCID: 0000-0001-7571-3505

Dragan Antić: dragan.antic@bio.bg.ac.rs; ORCID: 0000-0002-1231-4213

Jacopo Calevo: j.calevo@kew.org; ORCID: 0000-0002-1717-2365

Tommaso Cancellario: tommaso.cancellario@gmail.com; ORCID: 0000-0002-6637-4764

Pedro Cardoso: pedro.cardoso@helsinki.fi ; ORCID: 0000-0001-8119-9960

Dan Chamberlain: dan.chamberlain99@gmail.com; ORCID: 0000-0002-5381-2024

Matteo Chialva: matteo.chialva@unito.it; ORCID: 0000-0002-6996-6642

Furkan Durucan: f_durucan@hotmail.com; ORCID: 0000-0002-6168-2135

Diego Fontaneto: diego.fontaneto@cnr.it; ORCID: 0000-0002-5770-0353

Duarte Gonçalves: duartenvg@gmail.com; ORCID 0000-0003-4299-0375:

Alejandro Martínez: alejandro.martinezgarcia@cnr.it; ORCID: 0000-0003-0073-3688

Luca Santini: luca.santini@uniroma1.it; ORCID: 0000-0002-5418-3688

Iñigo Rubio-López: inirubio@ucm.es

Ronaldo Sousa: rg.eco.sousa@gmail.com; ORCID: 0000-0002-5961-5515

David Villegas-Ríos: dvillegas@iim.csic.es; ORCID: 0000-0001-5660-5322

Aida Verdes: aida.verdes@gmail.com; ORCID: 0000-0002-9193-9253

Ricardo A. Correia: ricardo.correia@helsinki.fi; ORCID: 0000-0001-7359-9091

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 3,007 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.

INTRODUCTION

Human relationships with biodiversity trace back to our dawn as a species[1]. 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[2–4]. 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[5]. Whether for utilitarian reasons or due to conflictual emotional stimuli[6], we have better knowledge of some species than others[7].

Widespread evidence indicates that biodiversity research has concentrated on certain lineages, habitats and geographic regions over others[8–12]. At the species level, for example, research interests and conservation efforts are often skewed toward vertebrates rather than other animals[13–15], plants[16,17], or fungi[18,19]. Furthermore, scientific and societal attention towards species may correlate, to some degree, with aesthetic features[20–22], online popularity[23,24] and phylogenetic proximity to

humans[25], 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, genome size, ploidy level[26]) and instead may be driven by economic, affective, cultural, or other subjective attributes[27].

Importantly, most attempts to quantify which features make species attractive to humans have focused on vertebrates—typically mammals and birds[25]. This means we now possess a growing understanding of research biases for selected taxa[10,28–30], 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 3,007 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, and the number of views of the Wikipedia page of each species as a measure of societal interest. 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 ± 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 ± 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 congenerics) 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. (ref. [31]) showed that in Israel, the most popular plants in terms of

online representation often have some use for humans. Similarly, Ladle et al. (ref. [32]) found that bird representation online is strongly associated with long histories of human interactions, for example in the form of hunting 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[33]. 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.

Species with a common name also attracted more scientific and popular interest, matching previous studies[31]. It must 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[34]. Interestingly, this speaks about the possible existence of interactions among different cultural traits, and within different cultures as we considered only English common names, that 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[35,36]. Previous studies documented positive relationships between human interest and body size, e.g., in different vertebrate groups[20,30,37,38] and flowering plants[22], while others observed negative relationships, e.g., in passerine birds[39] and butterflies[37]. 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[22].

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[23,37] or scientific interest[22]. Furthermore, 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[40] 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[41,42]. 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. (ref. [25]), 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[25]*. 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[17,24,43] 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[44]. 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., ref. [24] for a concrete agenda).

Global biodiversity is disappearing at an accelerating pace, not only from the physical world[45,46] but also from our minds[3,7,47]. 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[7,48,49]. 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; 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[50,51] 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*"[15]). 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 3007 species. 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[5,22,52,53]. We collected data using the R package 'wosr' version 0.3.0[54]. 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[55,56] 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[31]. 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[57]. 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[58] 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[59–61], 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[62,63]. Furthermore, we expected an innate preference for large-sized species among scientists, the media and the public alike[38,64–66]. 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[67]), 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[68] we hypothesized colorfulness to be a strong driver of attention toward biodiversity[40]. Furthermore, we scored red and blue patterns because these colors are known to impact people's affection, cognition and behavior[69]. Recent studies on European plants, for example, have highlighted that species with blue/purple flowers are more frequently studied in the scientific literature[22] and receive more conservation funds[17].

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[70]) 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[71] to extract distribution points for each species. As in Adamo et al. (ref. [22]), 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[72]. Geographical range size is not only a measure of ecological commonness[73], 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[74], e.g., common species that are available to us through direct experience[22,75]. 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[76] (available at www.marineregions.org; accessed on 1 November 2022), we

extracted the biogeographic region in which each species occurs (Afrotropical, Antarctic, Australasian, Nearctic, Neotropical, Oriental, Pacific, and Palaearctic) 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[77]. For this, we used a modified version of the *timetree()* function in the R package 'timetree' version 1.0 (https://github.com/FranzKrah/timetree; 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[1], relating to empathy and compassion towards species[25] and the degree of anthropomorphism in human-organism interactions[78].

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[79]. 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[80] 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[81]. We used the package ‘glmmTMB’ version 1.1.1 for modeling [82] and ‘ggplot2’ version 3.3.4[83] for visualizations. In all analyses, we followed the general approach by Zuur & Ieno (ref. [80]) 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[84]. For model validation, we used the suite of functions of the package ‘performance’ version 0.0.0.6[85] 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[86]. Furthermore, in interpreting and discussing results, we gave more relevance to explained variance and effect sizes rather than significance[87].

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 ~ Organism Size + Colorful + Range size + Domain + Taxonomic uniqueness + Common name + Human use + Harmful to human + Phylogenetic distance to humans + (1 | Phylum / Class / Order) + (1 | 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 | Phylum / Class / 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 | 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’[88]. 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[84].

Once the models were fitted and validated, we used variance partitioning

analysis[89] 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[90].

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 ~ Organism Size + Colorful + Range size + Domain + Genus uniqueness + Common name + Human use + Harmful to human + (1 | Class / Order) + (1 | 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.

ACKNOWLEDGEMENTS

Thanks to Caio Graco-Roza for helping with ggplot2. Filipe Chichorro kindly compiled traits for ants.

CONFLICT OF INTEREST

None declared

DATA AND CODE AVAILABILITY

The database used in the analyses is available in Figshare [doi: XXX]. The R code to generate analyses and figures is available on GitHub (XXX). (All data and scripts will be shared upon acceptance in a peer-reviewed journal).

SUPPORTING INFORMATION

Table S1–S3

Figure S1–S3

Supplementary text S1

LITERATURE CITED

1. Wilson EO. Biophilia and the conservation ethic. Island pre. In: Kellert S, Wilson EO, editors. The biophilia hypothesis. Island pre. Washington, DC; 1993.
2. Jacobs MH. Human Emotions Toward Wildlife. Hum Dimens Wildl. 2012;17: 1–3. doi:10.1080/10871209.2012.653674
3. Soga M, Gaston KJ. Extinction of experience: the loss of human–nature interactions. Front Ecol Environ. 2016;14: 94–101. doi:https://doi.org/10.1002/fee.1225
4. Hicks JR, Stewart WP. Exploring potential components of wildlife-inspired awe. Hum Dimens Wildl. 2018;23: 293–295. doi:10.1080/10871209.2018.1419518
5. Wilson JRU, Proches S, Braschler B, Dixon ES, Richardson DM. The (Bio)diversity of Science Reflects the Interests of Society. Front Ecol Environ. 2007;5: 409–414. Available: http://www.jstor.org/stable/20440727
6. Nyhus PJ. Human–Wildlife Conflict and Coexistence. Annu Rev Environ Resour. 2016;41: 143–171. doi:10.1146/annurev-environ-110615-085634

7. Jarić I, Roll U, Bonaiuto M, Brook BW, Courchamp F, Firth JA, et al. Societal extinction of species. Trends Ecol Evol. 2022;37: 411–419. doi:https://doi.org/10.1016/j.tree.2021.12.011

8. Clark JA, May RM. Taxonomic Bias in Conservation Research. Science (80-). 2002;297: 191 LP – 192. doi:10.1126/science.297.5579.191b

9. Hortal J, de Bello F, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, Ladle RJ. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annu Rev Ecol Evol Syst. 2015;46: 523–549. doi:10.1146/annurev-ecolsys-112414-054400

10. Šmíd J. Geographic and taxonomic biases in the vertebrate tree of life. J Biogeogr. 2022;49: 2120–2129. doi:https://doi.org/10.1111/jbi.14491

11. Mammola S, Fukushima CS, Biondo G, Bongiorni L, Cianferoni F, Domenici P, et al. How much biodiversity is concealed in the word 'biodiversity'? Curr Biol. 2023;33: R59–R60. doi:https://doi.org/10.1016/j.cub.2022.12.003

12. Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F. Taxonomic bias in biodiversity data and societal preferences. Sci Rep. 2017;7: 9132. doi:10.1038/s41598-017-09084-6

13. Cardoso P, Erwin TL, Borges PAV, New TR. The seven impediments in invertebrate conservation and how to overcome them. Biol Conserv. 2011;144: 2647–2655. doi:10.1016/j.biocon.2011.07.024

14. Cardoso P, Borges PA V, Triantis KA, Ferrandez MA, Martin JL. Adapting the IUCN Red List criteria for invertebrates. Biol Conserv. 2011;144: 2432–2440. doi:10.1016/j.biocon.2011.06.020

15. Leather SR. Institutional vertebratism hampers insect conservation generally; not just saproxylic beetle conservation. Anim Conserv. 2013;16: 379–380. doi:10.1111/acv.12068

16. Balding M, Williams KJH. Plant blindness and the implications for plant conservation. Conserv Biol. 2016;30: 1192–1199. doi:https://doi.org/10.1111/cobi.12738

17. Adamo M, Sousa R, Wipf S, Correia RA, Lumia A, Mucciarelli M, et al. Dimension and impact of biases in funding for species and habitat conservation. Biol Conserv. 2022;272: 109636. doi:https://doi.org/10.1016/j.biocon.2022.109636

18. Gonçalves S, Haelewaters D, Furci G, Mueller GM. Include all fungi in biodiversity goals. Science (80-). 2021;373: 403. doi:10.1126/science.abk1312

19. Oyanedel R, Hinsley A, Dentinger BTM, Milner-Gulland EJ, Furci G. A way forward for wild fungi in international sustainability policy. Conserv Lett. 2022;22: e12882. doi:https://doi.org/10.1111/conl.12882

20. Ward PI, Mosberger N, Kistler C, Fischer O. The Relationship between Popularity

and Body Size in Zoo Animals. Conserv Biol. 1998;12: 1408–1411. doi:https://doi.org/10.1111/j.1523-1739.1998.97402.x

21. Borgi M, Cogliati-Dezza I, Brelsford V, Meints K, Cirulli F. Baby schema in human and animal faces induces cuteness perception and gaze allocation in children . Frontiers in Psychology . 2014. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00411

22. Adamo M, Chialva M, Calevo J, Bertoni F, Dixon K, Mammola S. Plant scientists' research attention is skewed towards colourful, conspicuous and broadly distributed flowers. Nat Plants. 2021;7: 574–578. doi:10.1038/s41477-021-00912-2

23. Correia RA, Jepson PR, Malhado ACM, Ladle RJ. Familiarity breeds content: assessing bird species popularity with culturomics. PeerJ. 2016;4: e1728. doi:10.7717/peerj.1728

24. Mammola S, Riccardi N, Prié V, Correia R, Cardoso P, Lopes-Lima M, et al. Towards a taxonomically unbiased European Union biodiversity strategy for 2030. Proc R Soc B Biol Sci. 2020;287: 20202166. doi:10.1098/rspb.2020.2166

25. Miralles A, Raymond M, Lecointre G. Empathy and compassion toward other species decrease with evolutionary divergence time. Sci Rep. 2019;9: 19555. doi:10.1038/s41598-019-56006-9

26. Hedges SB. The origin and evolution of model organisms. Nat Rev Genet. 2002;3: 838–849. doi:10.1038/nrg929

27. Dietrich MR, Ankeny RA, Crowe N, Green S, Leonelli S. How to choose your research organism. Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci. 2020;80: 101227. doi:https://doi.org/10.1016/j.shpsc.2019.101227

28. Sumner S, Law G, Cini A. Why we love bees and hate wasps. Ecol Entomol. 2018;43: 836–845. doi:https://doi.org/10.1111/een.12676

29. Zvaríková M, Prokop P, Zvarík M, Ježová Z, Medina-Jerez W, Fedor P. What makes spiders frightening and disgusting to people? Frontiers in Ecology and Evolution. 2021. p. 424. Available: https://www.frontiersin.org/article/10.3389/fevo.2021.694569

30. Guedes JJM, Moura MR, Alexandre F. Diniz-Filho J. Species out of sight: elucidating the determinants of research effort in global reptiles. Ecography (Cop). 2023;n/a: e06491. doi:https://doi.org/10.1111/ecog.06491

31. Vardi R, Mittermeier JC, Roll U. Combining culturomic sources to uncover trends in popularity and seasonal interest in plants. Conserv Biol. 2021;35: 460–471. doi:https://doi.org/10.1111/cobi.13705

32. Ladle RJ, Jepson P, Correia RA, Malhado ACM. A culturomics approach to quantifying the salience of species on the global internet. People Nat. 2019;1: 524–532. doi:https://doi.org/10.1002/pan3.10053

33. Yorzinski JL, Penkunas MJ, Platt ML, Coss RG. Dangerous Animals Capture and Maintain Attention in Humans. Evol Psychol. 2014;12: 147470491401200320. doi:10.1177/147470491401200304

34. Farooq H, Bero C, Guilengue Y, Elias C, Massingue Y, Mucopote I, et al. Species perceived to be dangerous are more likely to have distinctive local names. J Ethnobiol Ethnomed. 2021;17: 69. doi:10.1186/s13002-021-00493-6

35. Kéry M, Gregg KB. Effects of life-state on detectability in a demographic study of the terrestrial orchid Cleistes bifaria. J Ecol. 2003;91: 265–273. doi:https://doi.org/10.1046/j.1365-2745.2003.00759.x

36. Johnston A, Newson SE, Risely K, Musgrove AJ, Massimino D, Baillie SR, et al. Species traits explain variation in detectability of UK birds. Bird Study. 2014;61: 340–350. doi:10.1080/00063657.2014.941787

37. Żmihorski M, Dziarska-Pałac J, Sparks TH, Tryjanowski P. Ecological correlates of the popularity of birds and butterflies in Internet information resources. Oikos. 2013;122: 183–190. doi:https://doi.org/10.1111/j.1600-0706.2012.20486.x

38. Berti E, Monsarrat S, Munk M, Jarvie S, Svenning J-C. Body size is a good proxy for vertebrate charisma. Biol Conserv. 2020;251: 108790. doi:https://doi.org/10.1016/j.biocon.2020.108790

39. Garrett KL, Shuford WD, Gill RE, Handel CM. Introducing change: A current look at naturalized bird species in western North America. Trends Tradit Avifaunal Chang West North Am. 2018; 116–130.

40. Langlois J, Guilhaumon F, Baletaud F, Casajus N, De Almeida Braga C, Fleuré V, et al. The aesthetic value of reef fishes is globally mismatched to their conservation priorities. PLOS Biol. 2022;20: e3001640. Available: https://doi.org/10.1371/journal.pbio.3001640

41. Borges AKM, Oliveira TPR, Alves RRN. Marine or freshwater: the role of ornamental fish keeper's preferences in the conservation of aquatic organisms in Brazil. Reimer J, editor. PeerJ. 2022;10: e14387. doi:10.7717/peerj.14387

42. Senior RA, Oliveira BF, Dale J, Scheffers BR. Wildlife trade targets colorful birds and threatens the aesthetic value of nature. Curr Biol. 2022;32: 4299-4305.e4. doi:https://doi.org/10.1016/j.cub.2022.07.066

43. Davies T, Cowley A, Bennie J, Leyshon C, Inger R, Carter H, et al. Popular interest in vertebrates does not reflect extinction risk and is associated with bias in conservation investment. PLoS One. 2018;14: e0212101. doi:10.1371/journal.pone.0203694

44. Chowdhury S, Zalucki MP, Hanson JO, Tiatragul S, Green D, Watson JEM, et al. Three-quarters of insect species are insufficiently represented by protected areas.

One Earth. 2023. doi:https://doi.org/10.1016/j.oneear.2022.12.003

45. Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, et al. Has the Earth's sixth mass extinction already arrived? Nature. 2011;471: 51. Available: https://doi.org/10.1038/nature09678

46. Cowie RH, Bouchet P, Fontaine B. The Sixth Mass Extinction: fact, fiction or speculation? Biol Rev. 2022;97: 640–663. doi:https://doi.org/10.1111/brv.12816

47. Donadio Linares LM. 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. 2022;135: 137–146. doi:https://doi.org/10.1016/j.envsci.2022.04.019

48. Loreau M, Barbier M, Filotas E, Gravel D, Isbell F, Miller SJ, et al. Biodiversity as insurance: from concept to measurement and application. Biol Rev. 2021;96: 2333–2354. doi:https://doi.org/10.1111/brv.12756

49. Díaz S, Malhi Y. Biodiversity: Concepts, Patterns, Trends, and Perspectives. Annu Rev Environ Resour. 2022;47: 31–63. doi:10.1146/annurev-environ-120120-054300

50. Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, et al. The Revised Classification of Eukaryotes. J Eukaryot Microbiol. 2012;59: 429–514. doi:https://doi.org/10.1111/j.1550-7408.2012.00644.x

51. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, et al. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. J Eukaryot Microbiol. 2019;66: 4–119. doi:https://doi.org/10.1111/jeu.12691

52. dos Santos JW, Correia RA, Malhado ACM, Campos-Silva J V, Teles D, Jepson P, et al. Drivers of taxonomic bias in conservation research: a global analysis of terrestrial mammals. Anim Conserv. 2020;23: 679–688. doi:https://doi.org/10.1111/acv.12586

53. Tam J, Lagisz M, Cornwell W, Nakagawa S. Quantifying research interests in 7,521 mammalian species with h-index: a case study. Gigascience. 2022;11: giac074. doi:10.1093/gigascience/giac074

54. Baker C. wosr: Clients to the "Web of Science" and "InCites" APIs. R package; 2018.

55. Correia RA, Jepson P, Malhado ACM, Ladle RJ. Internet scientific name frequency as an indicator of cultural salience of biodiversity. Ecol Indic. 2017;78: 549–555. doi:10.1016/j.ecolind.2017.03.052

56. Jarić I, Courchamp F, Gessner J, Roberts DL. Data mining in conservation research using Latin and vernacular species names. McKeon S, editor. PeerJ. 2016;4: e2202. doi:10.7717/peerj.2202

57. Popov M. WikidataQueryServiceR: API Client Library for 'Wikidata Query Service. R

package version 1.0.0; 2020.

58. Keyes O, Lewis J. pageviews: An API Client for Wikimedia Traffic Data. R package version 0.5.0; 2020. Available: https://cran.r-project.org/package=pageviews

59. Weiss KCB, Ray CA. Unifying functional trait approaches to understand the assemblage of ecological communities: synthesizing taxonomic divides. Ecography (Cop). 2019;42: 2012–2020. doi:10.1111/ecog.04387

60. Chichorro F, Urbano F, Teixeira D, Väre H, Pinto T, Brummitt N, et al. Trait-based prediction of extinction risk across terrestrial taxa. Biol Conserv. 2022;274: 109738. doi:https://doi.org/10.1016/j.biocon.2022.109738

61. Palacio FX, Callaghan CT, Cardoso P, Hudgins EJ, Jarzyna MA, Ottaviani G, et al. A protocol for reproducible functional diversity analyses. Ecography (Cop). 2022; e06287. doi:10.1111/ecog.06287

62. Peters RH, Peters RH. The ecological implications of body size. Cambridge university press; 1986.

63. Calder WA. Size, function, and life history. Courier Corporation; 1996.

64. McClain CR, Balk MA, Benfield MC, Branch TA, Chen C, Cosgrove J, et al. Sizing ocean giants: patterns of intraspecific size variation in marine megafauna. PeerJ. 2015;3: e715. doi:10.7717/peerj.715

65. Mammola S, Michalik P, Hebets EA, Isaia M. Record breaking achievements by spiders and the scientists who study them. PeerJ. 2017;5: e3972. doi:10.7717/peerj.3972

66. Hall CM, James M, Baird T. Forests and trees as charismatic mega-flora: implications for heritage tourism and conservation. J Herit Tour. 2011;6: 309–323. doi:10.1080/1743873X.2011.620116

67. Delhey K, Dale J, Valcu M, Kempenaers B. Migratory birds are lighter coloured. Curr Biol. 2021;31: R1511–R1512. doi:https://doi.org/10.1016/j.cub.2021.10.048

68. Hoyer WD, Stokburger-Sauer NE. The role of aesthetic taste in consumer behavior. J Acad Mark Sci. 2012;40: 167–180.

69. Elliot AJ, Maier MA. Color Psychology: Effects of Perceiving Color on Psychological Functioning in Humans. Annu Rev Psychol. 2014;65: 95–120. doi:10.1146/annurev-psych-010213-115035

70. Gillespie RG, Bennett GM, De Meester L, Feder JL, Fleischer RC, Harmon LJ, et al. Comparing Adaptive Radiations Across Space, Time, and Taxa. J Hered. 2020;111: 1–20. doi:10.1093/jhered/esz064

71. Chamberlain S, Oldoni D, Waller J. rgbif: Interface to the Global Biodiversity Information Facility API. 2022. Available: https://cran.r-project.org/package=rgbif

72. Hughes AC, Orr MC, Ma K, Costello MJ, Waller J, Provoost P, et al. Sampling biases shape our view of the natural world. Ecography (Cop). 2021;44: 1259–1269. doi:https://doi.org/10.1111/ecog.05926

73. Gaston KJ. Common Ecology. Bioscience. 2011;61: 354–362. doi:10.1525/bio.2011.61.5.4

74. Ladle RJ, Correia RA, Do Y, Joo GJ, Malhado ACM, Proulx R, et al. Conservation culturomics. Front Ecol Environ. 2016;14: 269–275. doi:10.1002/fee.1260

75. Schuetz JG, Johnston A. Characterizing the cultural niches of North American birds. Proc Natl Acad Sci. 2019;116: 10868–10873. doi:10.1073/pnas.1820670116

76. Balian E V, Segers H, Martens K, Lévêque C. An introduction to the freshwater animal diversity assessment (FADA) project. Hydrobiologia. 2008;600: 311.

77. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, et al. TimeTree 5: An Expanded Resource for Species Divergence Times. Mol Biol Evol. 2022;39: msac174. doi:10.1093/molbev/msac174

78. Servais V. Anthropomorphism in human–animal interactions: A pragmatist view. Front Psychol. 2018;9: 2590.

79. Chamberlain S. rredlist: “IUCN” Red List Client. 2022.

80. Zuur AF, Ieno EN. A protocol for conducting and presenting results of regression-type analyses. Methods Ecol Evol. 2016;7: 636–645. doi:10.1111/2041-210X.12577

81. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.

82. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. R J. 2017;9: 378–400. doi:10.32614/RJ-2017-066

83. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

84. Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol. 2009;1: 3–14. doi:10.1111/j.2041-210x.2009.00001.x

85. Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. J Open Source Softw. 2020;6: 3139. doi:10.21105/joss.03139

86. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, et al. Redefine statistical significance. Nat Hum Behav. 2018;2: 6–10. doi:10.1038/s41562-017-0189-z

87. Muff S, Nilsen EB, O'Hara RB, Nater CR. Rewriting results sections in the language of evidence. Trends Ecol Evol. 2021;37: 203–210. doi:10.1016/j.tree.2021.10.009

88. Hardin JW, Hilbe J. Generalized linear models and extensions. Stata Press; 2007.

89. Borcard D, Legendre P, Drapeau P. Partialling out the Spatial Component of Ecological Variation. Ecology. 1992;73: 1045–1055. doi:10.2307/1940179

90. Barbosa AM, Brown JA, Jimenez-Valverde A, Real A. modEvA: Model Evaluation and Analysis. 2015.

FIGURES

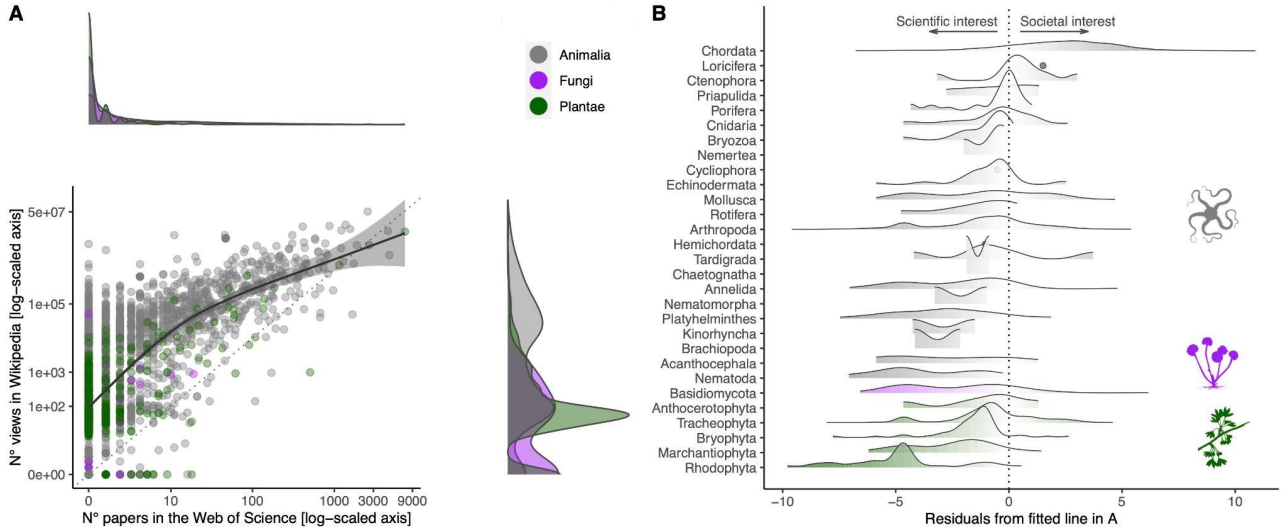

Figure 1. **Relationship between societal and scientific interest across the eukaryotic Tree of Life. A**) Relationship between number of views in Wikipedia (popular interest) and number of papers in the Web of Science (scientific interest) for each species. Both axes are log-scaled to ease visualization. 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.

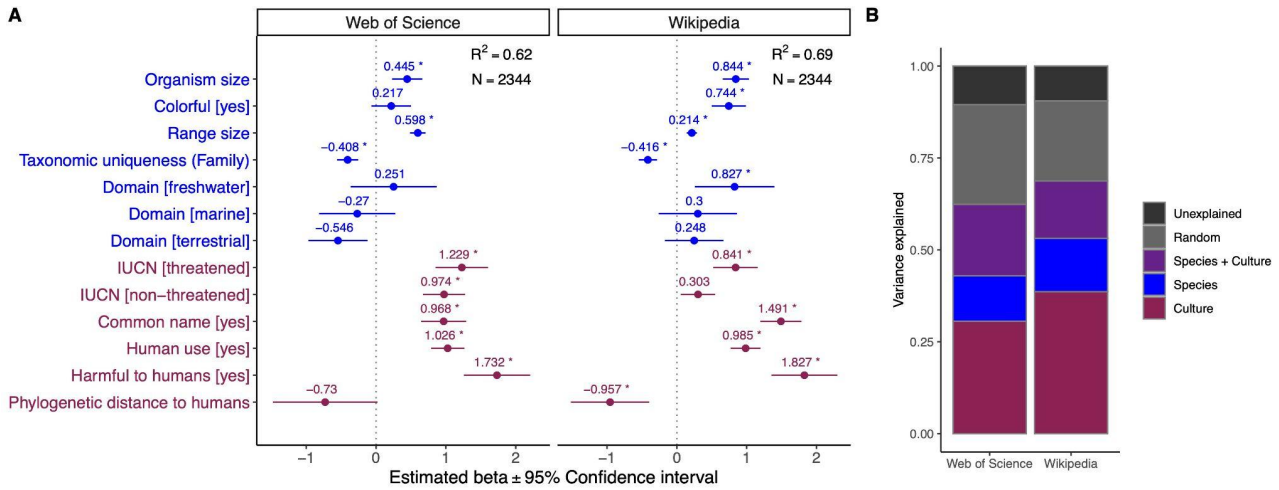

Figure 2. Influence of species-level traits (blue) and cultural factors (red) on the scientific and societal interest across the eukaryotic Tree of Life. A) Forest plots summarize the estimated parameters based on Negative binomial generalized linear mixed models (Eq. 1). Baseline levels for multilevel factor variables are: Domain [Multiple] and IUCN [Not Evaluated]. 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).

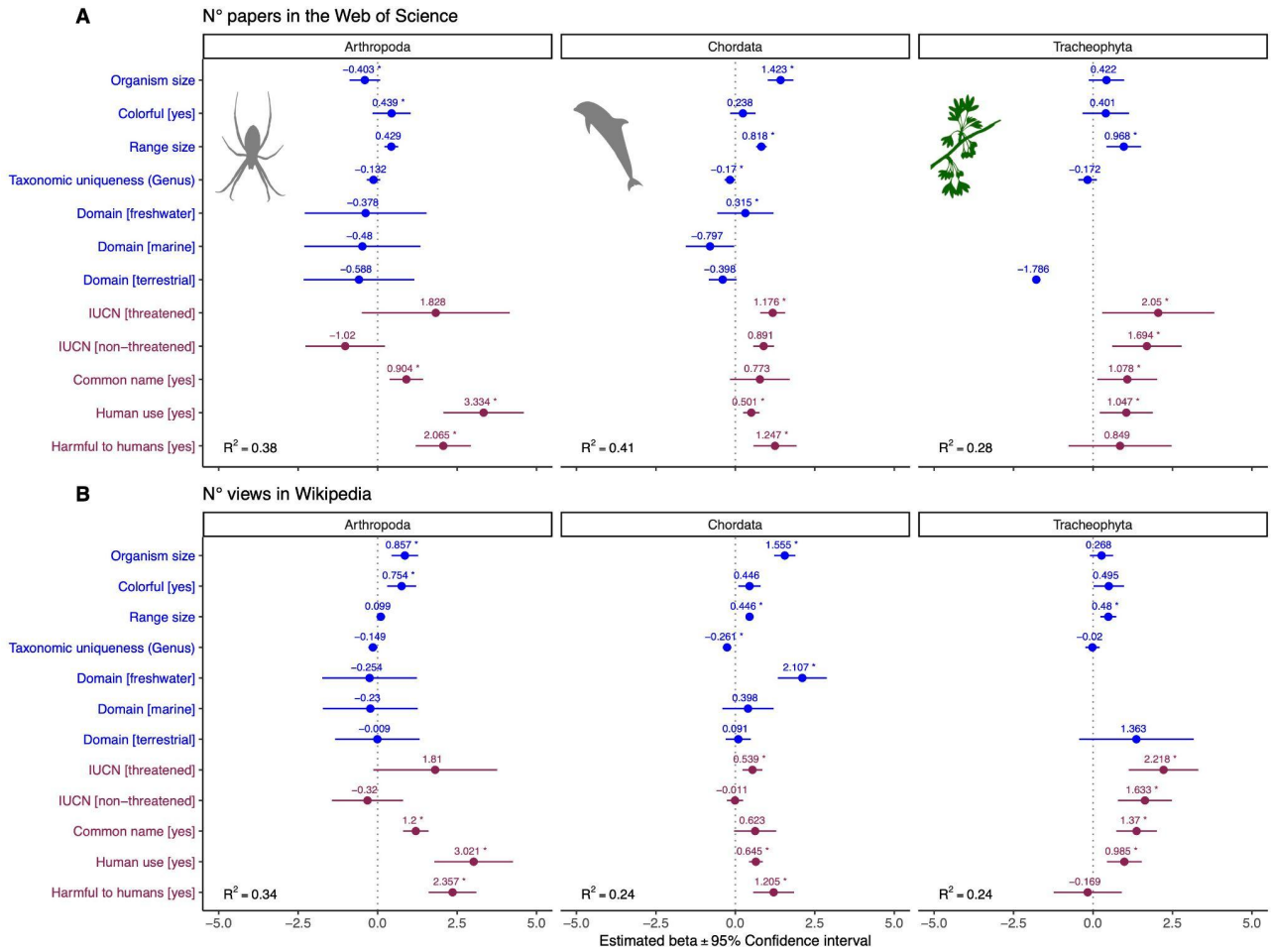

Figure 3. Influence of species-level traits (blue) and cultural factors (red) on scientific (A) and societal (B) interest for Arthopoda, Chordata, and Tracheophyta. Forest plots summarize the estimated parameters based on Negative binomial generalized linear mixed models (Eq. 2). Baseline levels for multilevel factor variables are: Domain [Multiple] and IUCN [Not Evaluated]. 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**).