Drivers of species knowledge across the Tree of Life

Stefano Mammola[1,2,3,*], Martino Adamo[3,4], Dragan Antić[5], Jacopo Calevo[6,7], Tommaso Cancellario[1], Pedro Cardoso[2], Dan Chamberlain[4], Matteo Chialva[3,4], 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[14,15,16,17]

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. Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK

7. School of Molecular and Life Sciences, Curtin University, Bentley, WA 6102, Australia

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

17. Biodiversity Unit, University of Turku, 20014 Turku, Finland

* corresponding author: stefano.mammola@cnr.it

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 3,019 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.

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 1; Figure 1—figure supplement 1). We sourced the number of scientific papers focusing on each species as a measure of scientific interest (Figure 1—figure supplement 2A), and the number of views of the Wikipedia page of each species as a measure of societal interest (Figure 1—figure supplement 2B). 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 1—figure supplement 2A). 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 1—figure supplement 2B). 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 2).

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 3A; 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 3B).

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 3A), 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 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 4). 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 4). 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, 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 3A 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; 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.

This random sampling procedure yielded a database consisting of 3,019 species (Figure 1—figure supplement 2C). 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 Appendix 1).

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), 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) 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). 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) and receive more conservation

funds (Adamo et al., 2022).

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), 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). Geographical range size is not only a measure of ecological commonness (Gaston, 2011), 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), e.g., common species that are available to us through direct experience (Adamo et al., 2021; Schuetz and Johnston, 2019). 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 1—figure supplement 1). Using the FADA Faunistic Regions database (Balian et al., 2008) (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 (Kumar et al., 2022). 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 (Wilson, 1993), relating to empathy and compassion towards species (Miralles et al., 2019) and the degree of anthropomorphism in human-organism interactions (Servais, 2018).

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). 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) 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) for visualizations. In all analyses, we followed the general approach by Zuur & Ieno (2016) 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). For model validation, we used the suite of functions of the package 'performance' version 0.0.0.6 (Lüdecke et al., 2020) 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). Furthermore, in interpreting and discussing results, we gave more relevance to explained variance and effect sizes rather than significance (Muff et al., 2021).

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 + IUCN + 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' (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 ~ Organism Size + Colorful + Range size + Domain + Genus uniqueness + Common name + IUCN + 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.

We also 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. 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. RAC acknowledges funding from the Academy of Finland (Grant agreement #348352) and the KONE Foundation (Grant agreement #202101976).

CONFLICT OF INTEREST

None declared

DATA AND CODE AVAILABILITY

The database used in the analyses is available in Figshare (https://doi.org/10.6084/m9.figshare.22731440.v1). The R code to generate analyses and figures is available on GitHub (https://github.com/StefanoMammola/Mammola_et_al_ToL_research_interest).

LITERATURE CITED

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. doi:10.1038/s41477-021-00912-2

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**:109636. doi:10.1016/j.biocon.2022.109636

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. doi:10.1111/jeu.12691

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. doi:10.1111/j.1550-7408.2012.00644.x

Baker C. 2018. wosr: Clients to the “Web of Science” and “InCites” APIs.

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

Balian E V, Segers H, Martens K, Lévêque C. 2008. An introduction to the freshwater animal diversity assessment (FADA) project. *Hydrobiologia* **600**:311. doi:10.1007/s10750-007-9235-6

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

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**:51. doi:10.1038/nature09678

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, Rouder 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. doi:10.1038/s41562-017-0189-z

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

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

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**:e14387. doi:10.7717/peerj.14387

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**:411. doi:10.3389/fpsyg.2014.00411

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. doi:10.32614/RJ-2017-066

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

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

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

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

Chamberlain S, Oldoni D, Waller J. 2022. rgbif: Interface to the Global Biodiversity Information Facility API.

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**:109738. doi:10.1016/j.biocon.2022.109738

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*. doi:10.1016/j.oneear.2022.12.003

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

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. doi:10.1016/j.ecolind.2017.03.052

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

Correia RA, Mammola S. 2023. The searchscape of fear: A global analysis of internet search trends for biophobias. *People Nat*. doi:10.1002/pan3.10497

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

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**:e0212101. doi:10.1371/journal.pone.0203694

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

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

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**:101227. doi:10.1016/j.shpsc.2019.101227

Donadio Linares LM. 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. doi:10.1016/j.envsci.2022.04.019

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. doi:10.1111/acv.12586

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

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**:69. doi:10.1186/s13002-021-00493-6

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**: 110118. doi:10.1016/j.biocon.2023.110118

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.

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

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. doi:10.1093/jhered/esz064

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

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

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.

doi:10.1080/1743873X.2011.620116

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

Haukka A, Lehikoinen 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. doi:10.1038/s41597-023-02169-0

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

Hicks JR, Stewart WP. 2018. Exploring potential components of wildlife-inspired awe. *Hum Dimens Wildl* **23**:293–295. doi:10.1080/10871209.2018.1419518

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. doi:10.1146/annurev-ecolsys-112414-054400

Hoyer WD, Stokburger-Sauer NE. 2012. The role of aesthetic taste in consumer behavior. *J Acad Mark Sci* **40**:167–180. doi:10.1007/s11747-011-0269-y

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. doi:10.1111/ecog.05926

Jacobs MH. 2012. Human Emotions Toward Wildlife. *Hum Dimens Wildl* **17**:1–3. doi:10.1080/10871209.2012.653674

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

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. doi:10.1016/j.tree.2021.12.011

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. doi:10.1080/00063657.2014.941787

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. doi:10.1046/j.1365-2745.2003.00759.x

Keyes O, Lewis J. 2020. pageviews: An API Client for Wikimedia Traffic Data.

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**:msac174. doi:10.1093/molbev/msac174

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. doi:10.1002/fee.1260

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. doi:10.1002/pan3.10053

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. doi:10.1017/ext.2023.20

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**:e3001640. doi:10.1371/journal.pbio.3001640

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

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. doi:10.1111/brv.12756

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**:3139. doi:10.21105/joss.03139

Mammola S, Fukushima CS, Biondo G, Bongiorni 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. doi:10.1016/j.cub.2022.12.003

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

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**:20202166. doi:10.1098/rspb.2020.2166

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**:e715. doi:10.7717/peerj.715

Miralles A, Raymond M, Lecointre G. 2019. Empathy and compassion toward other

species decrease with evolutionary divergence time. *Sci Rep* **9**:19555. doi:10.1038/s41598-019-56006-9

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

Nyhus PJ. 2016. Human–Wildlife Conflict and Coexistence. *Annu Rev Environ Resour* **41**:143–171. doi:10.1146/annurev-environ-110615-085634

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**:e12882. doi:10.1111/conl.12882

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**:e06287. doi:10.1111/ecog.06287

Peters Robert H, Peters Robert Henry. 1986. The ecological implications of body size. Cambridge university press.

Popov M. 2020. WikidataQueryServiceR: API Client Library for 'Wikidata Query Service.

R Core Team. 2021. R: A Language and Environment for Statistical Computing.

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**:e2217303120. doi:10.1073/pnas.2217303120

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

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.e4. doi:10.1016/j.cub.2022.07.066

Servais V. 2018. Anthropomorphism in human–animal interactions: A pragmatist view. *Front Psychol* **9**:2590. doi:10.3389/fpsyg.2018.02590

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

Soga M, Gaston KJ. 2016. Extinction of experience: the loss of human–nature interactions. *Front Ecol Environ* **14**:94–101. doi:10.1002/fee.1225

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

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**:giac074. doi:10.1093/gigascience/giac074

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

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. doi:10.1111/cobi.13705

Ward PI, Mosberger N, Kistler C, Fischer O. 1998. The Relationship between Popularity and Body Size in Zoo Animals. *Conserv Biol* **12**:1408–1411. doi:10.1111/j.1523-1739.1998.97402.x

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

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

Wilson Edward O. 1993. Biophilia and the conservation ethic. In: Kellert S, Wilson E O, editors. The Biophilia Hypothesis. Washington, DC.

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.

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

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. doi:0.1111/j.1600-0706.2012.20486.x

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

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

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**:694569. doi:10.3389/fevo.2021.694569

CAPTIONS

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 1—figure supplement 1. Global distribution of the sampled species. Maps are based on the centroid of the distribution of each species, hence only species for which distribution points were available in the Global Biodiversity Information Facility are depicted (89%, n = 2687). **A**) Density of distribution centroids across all groups; **B**) Distribution of centroids for animals; **C**) Distribution of centroids for plants; **D**) Distribution of centroids for fungi.

Figure 1—figure supplement 2. Breakdown of scientific (A) and societal (B) interest by phylum, as well as the number of sampled species (C). Jittered points are the actual values; boxplots summarize median and quantiles. Both axes are log-scaled to ease visualization.

Figure 2. Influence of species-level traits (blue) and cultural factors (red) on relative scientific and societal interest for different taxa. Forest plots summarize the estimated parameters based on Gaussian linear mixed models testing the relationship between residuals from the regression line in Figure 1A and species-level traits. Positive residuals indicate species with a greater popular than scientific interest, residuals close to zero indicate species with a balanced scientific and societal interest, and negative residuals

indicate species with a greater scientific than societal interest (Figure 1B). 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 (α = 0.01). Estimated regression parameters are in Supplementary File 1a**.**

Figure 3. 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 (α = 0.01). Exact estimated regression parameters and p-values are in Supplementary File 1a. **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).

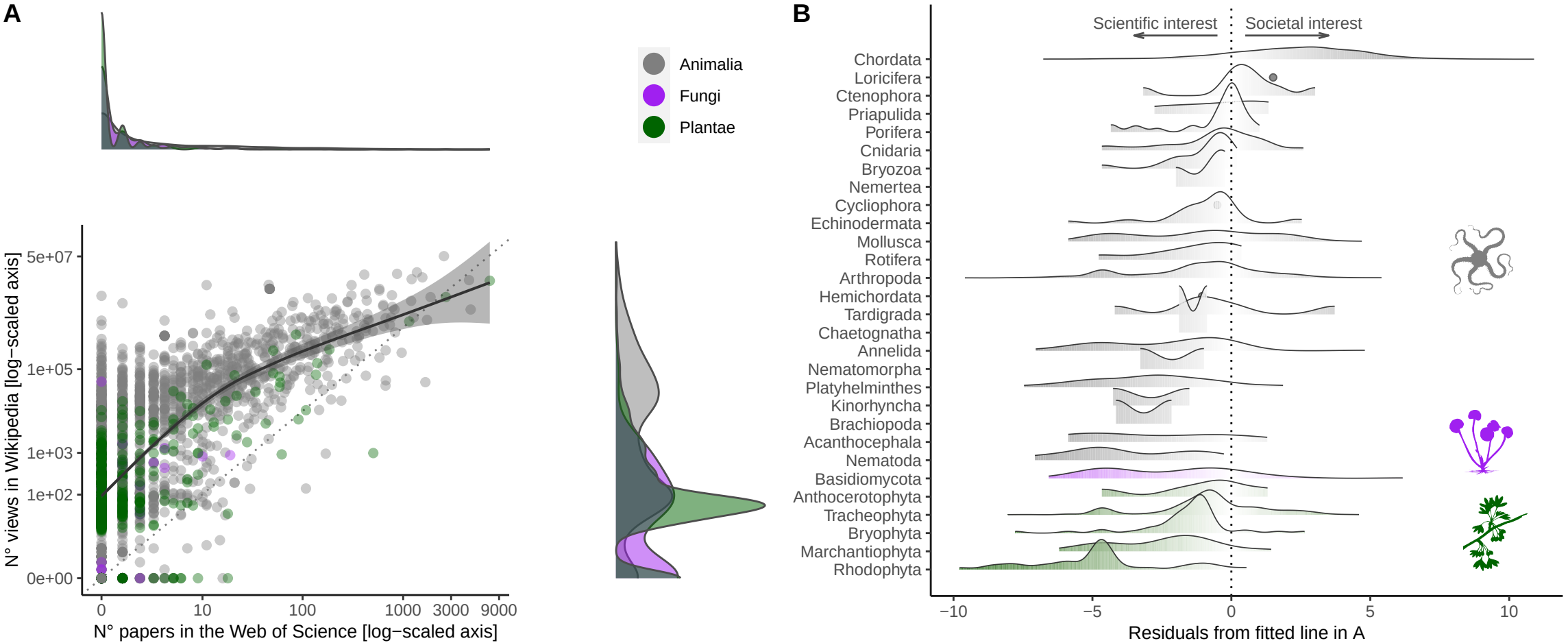
Figure 4. Influence of species-level traits (blue) and cultural factors (red) on scientific (A) and societal (B) interest for Arthopoda, Chordata, and Tracheophyta. Forest plots summarize the estimated parameters based on Negative binomial generalized linear mixed models (Eq. 2). Baseline levels for multilevel factor variables are: Domain [Multiple] and IUCN [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 (α = 0.01). Estimated regression parameters and p-values are in Supplementary File 1b (**A**) and Supplementary File 1c (**B**).

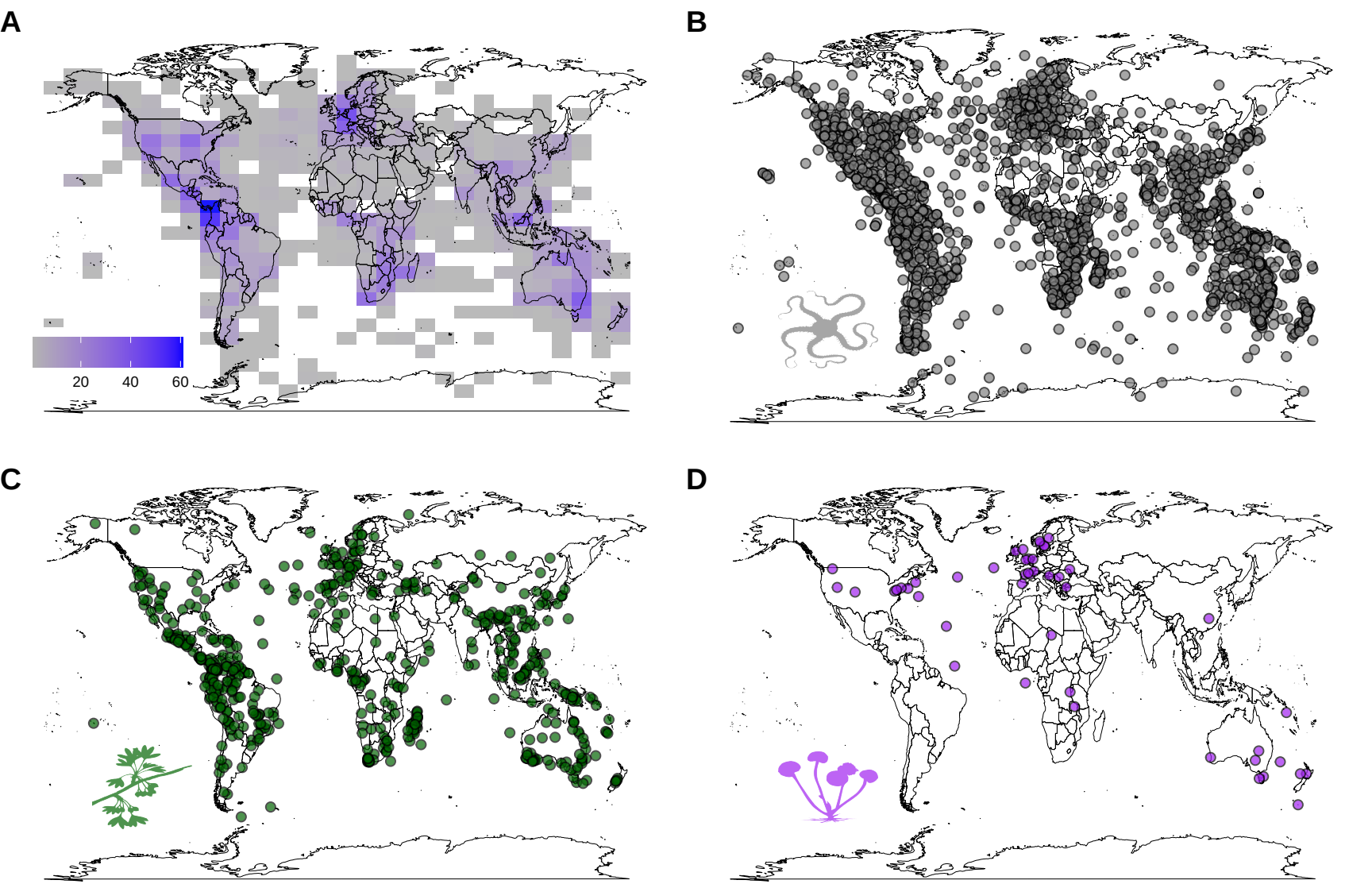
Supplementary File 1a. Estimated regression parameters for the full models.

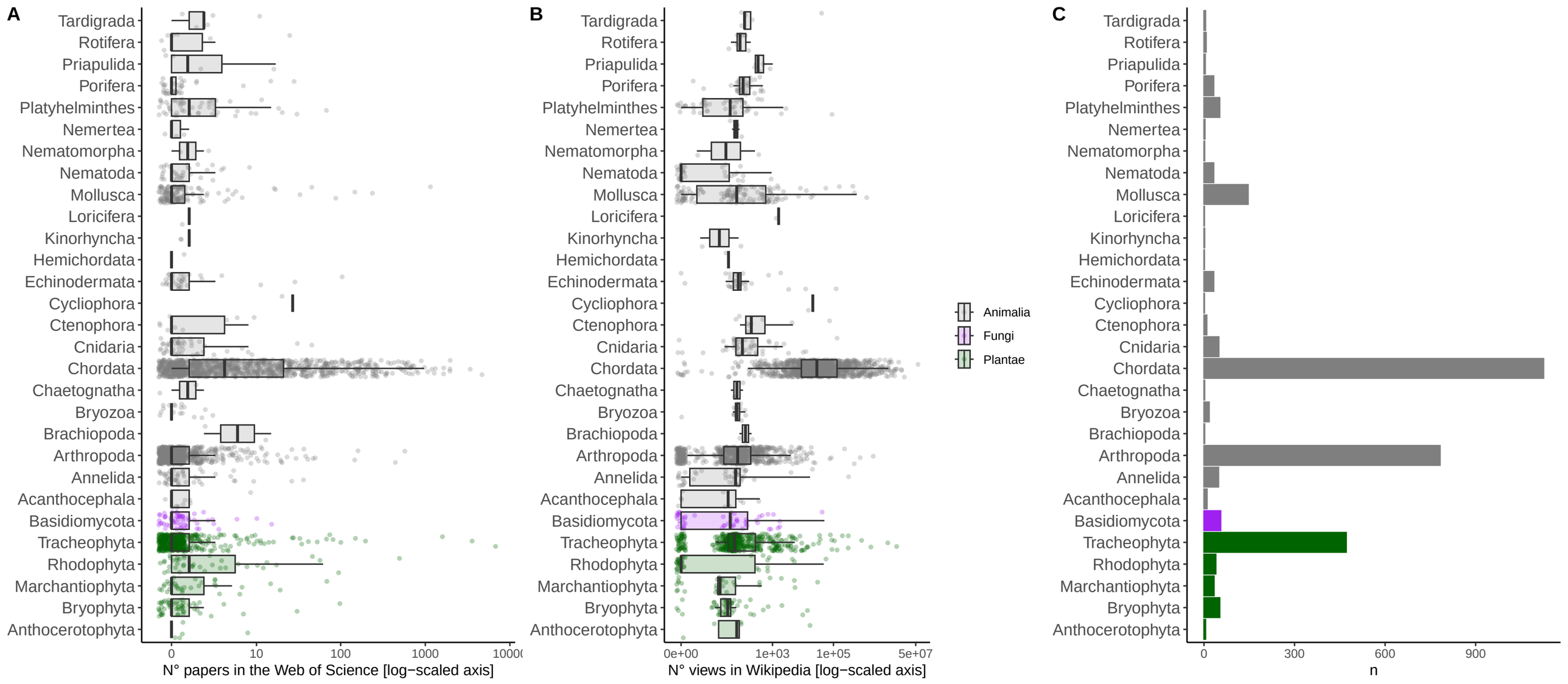
Supplementary File 1b. Estimated regression parameters for the subset models modeling

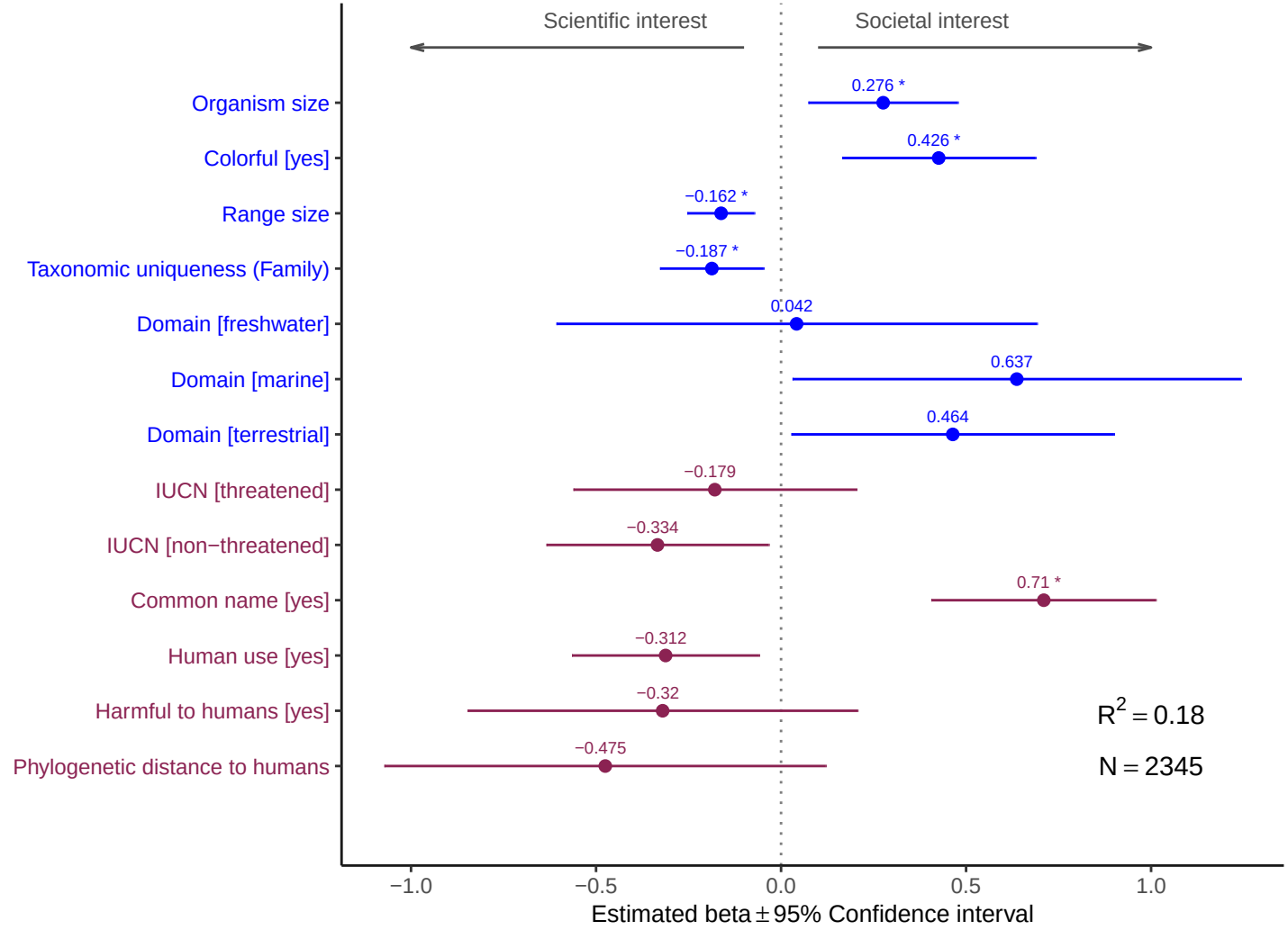
the N° of papers in the Web of Science.

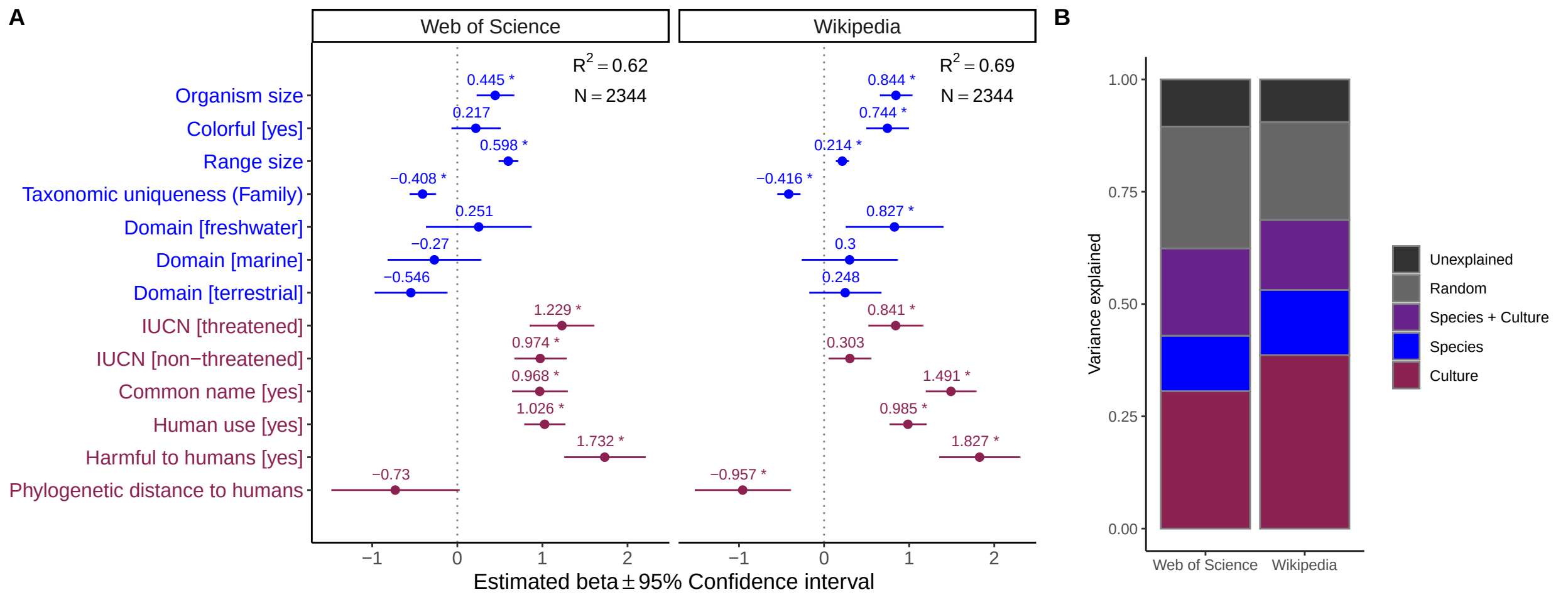
Supplementary File 1c. Estimated regression parameters for the subset models modeling the N° of views in Wikipedia.

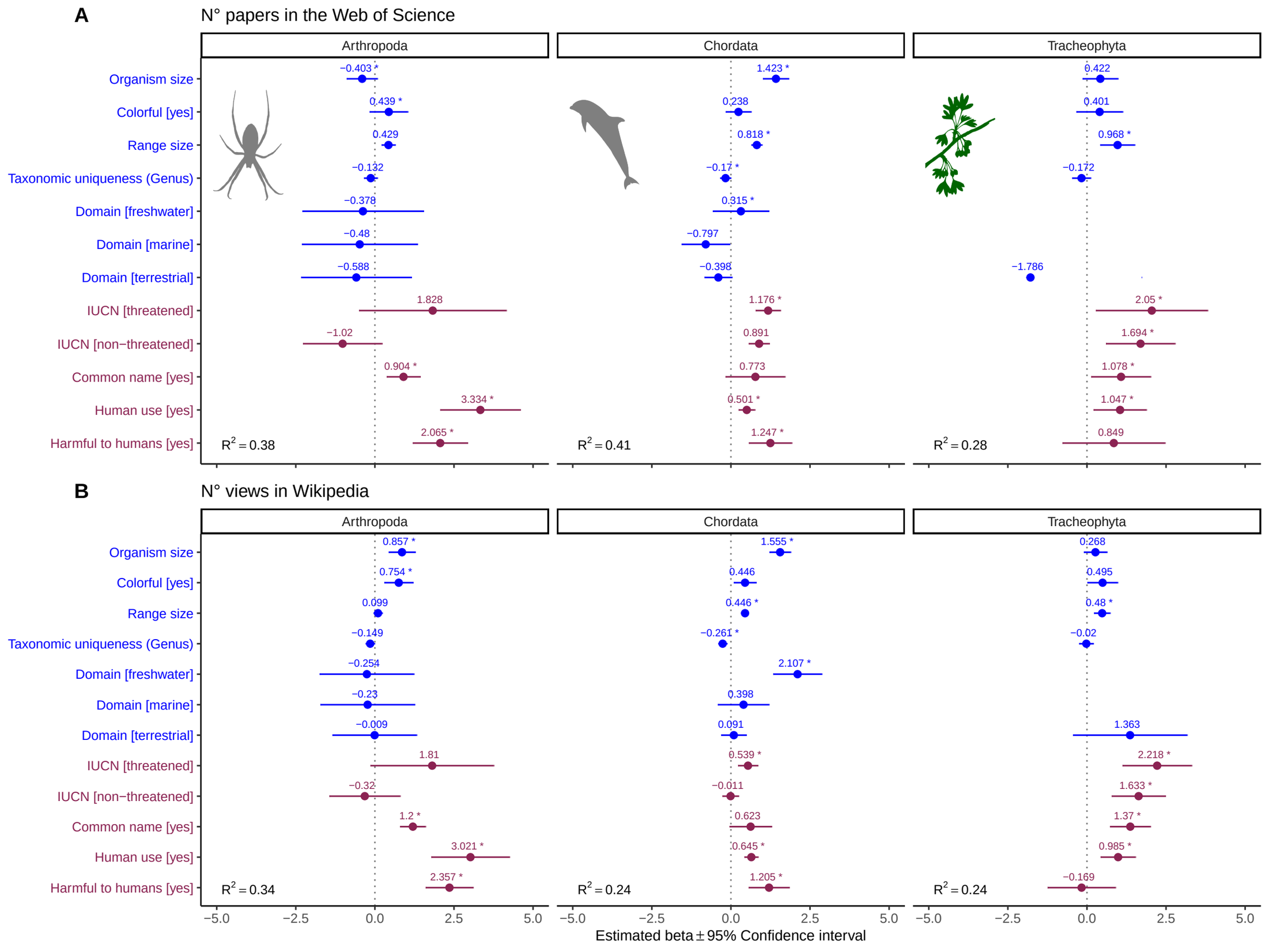

A
Animalia
Fungi
Plantae
N° views in Wikipedia [log−scaled axis]
5e+07
1e+05
1e+03
1e+02
0e+00
0
10
100
1000
3000
9000
N° papers in the Web of Science [log−scaled axis]
B
Scientific interest
Societal interest
Chordata
Loricifera
Ctenophora
Priapulida
Porifera
Cnidaria
Bryozoa
Nemertea
Cycliophora
Echinodermata
Mollusca
Rotifera
Arthropoda
Hemichordata
Tardigrada
Chaetognatha
Annelida
Nematomorpha
Platyhelminthes
Kinorhyncha
Brachiopoda
Acanthocephala
Nematoda
Basidiomycota
Anthocerotophyta
Tracheophyta
Bryophyta
Marchantiophyta
Rhodophyta
−10
−5
0
5
10
Residuals from fitted line in A


A
20
40
60
B
C
D


A
B
C
Tardigrada
Rotifera
Priapulida
Porifera
Platyhelminthes
Nemertea
Nematomorpha
Nematoda
Mollusca
Loricifera
Kinorhyncha
Hemichordata
Echinodermata
Cycliophora
Ctenophora
Cnidaria
Chordata
Chaetognatha
Bryozoa
Brachiopoda
Arthropoda
Annelida
Acanthocephala
Basidiomycota
Tracheophyta
Rhodophyta
Marchantiophyta
Bryophyta
Anthocerotophyta
0
10
100
1000
10000
N° papers in the Web of Science [log−scaled axis]
0e+00
1e+03
1e+05
5e+07
N° views in Wikipedia [log−scaled axis]
Animalia
Fungi
Plantae
0
300
600
900
n


Scientific interest
Societal interest
Organism size
Colorful [yes]
Range size
Taxonomic uniqueness (Family)
Domain [freshwater]
Domain [marine]
Domain [terrestrial]
IUCN [threatened]
IUCN [non−threatened]
Common name [yes]
Human use [yes]
Harmful to humans [yes]
Phylogenetic distance to humans
0.276 *
0.426 *
−0.162 *
−0.187 *
0.042
0.637
0.464
−0.179
−0.334
0.71 *
−0.312
−0.32
−0.475
R2 = 0.18
N = 2345
−1.0
−0.5
0.0
0.5
1.0
Estimated beta ± 95% Confidence interval


A
Web of Science
Wikipedia
$R^2 = 0.62$
N = 2344
$R^2 = 0.69$
N = 2344
Organism size
Colorful [yes]
Range size
Taxonomic uniqueness (Family)
Domain [freshwater]
Domain [marine]
Domain [terrestrial]
IUCN [threatened]
IUCN [non−threatened]
Common name [yes]
Human use [yes]
Harmful to humans [yes]
Phylogenetic distance to humans
0.445 *
0.217
0.598 *
−0.408 *
0.251
−0.27
−0.546
1.229 *
0.974 *
0.968 *
1.026 *
1.732 *
−0.73
0.844 *
0.744 *
0.214 *
−0.416 *
0.827 *
0.3
0.248
0.841 *
0.303
1.491 *
0.985 *
1.827 *
−0.957 *
−1
0
1
2
Estimated beta ± 95% Confidence interval
B
Variance explained
1.00
0.75
0.50
0.25
0.00
Web of Science
Wikipedia
Unexplained
Random
Species + Culture
Species
Culture


A
N° papers in the Web of Science
Arthropoda
Chordata
Tracheophyta
Organism size
Colorful [yes]
Range size
Taxonomic uniqueness (Genus)
Domain [freshwater]
Domain [marine]
Domain [terrestrial]
IUCN [threatened]
IUCN [non−threatened]
Common name [yes]
Human use [yes]
Harmful to humans [yes]
−0.403 *
0.439 *
0.429
−0.132
−0.378
−0.48
−0.588
1.828
−1.02
0.904 *
3.334 *
2.065 *
R2 = 0.38
1.423 *
0.238
0.818 *
−0.17 *
0.315 *
−0.797
−0.398
1.176 *
0.891
0.773
0.501 *
1.247 *
R2 = 0.41
0.422
0.401
0.968 *
−0.172
−1.786
2.05 *
1.694 *
1.078 *
1.047 *
0.849
R2 = 0.28
B
N° views in Wikipedia
Arthropoda
Chordata
Tracheophyta
Organism size
Colorful [yes]
Range size
Taxonomic uniqueness (Genus)
Domain [freshwater]
Domain [marine]
Domain [terrestrial]
IUCN [threatened]
IUCN [non−threatened]
Common name [yes]
Human use [yes]
Harmful to humans [yes]
0.857 *
0.754 *
0.099
−0.149
−0.254
−0.23
−0.009
1.81
−0.32
1.2 *
3.021 *
2.357 *
R2 = 0.34
1.555 *
0.446
0.446 *
−0.261 *
2.107 *
0.398
0.091
0.539 *
−0.011
0.623
0.645 *
1.205 *
R2 = 0.24
0.268
0.495
0.48 *
−0.02
1.363
2.218 *
1.633 *
1.37 *
0.985 *
−0.169
R2 = 0.24
−5.0
−2.5
0.0
2.5
5.0
Estimated beta ± 95% Confidence interval