Insect Conservation and Diversity

Insect Conservation and Diversity (2021) **14,** 700–708 doi: 10.1111/icad.12499

SHORT COMMUNICATION

Out of sight, out of mind: public and research interest in insects is negatively correlated with their conservation status

ZHENGYANG WANG,[1] JIMMY ZENG,[2] WEILIN MENG,[3] DAVID J. LOHMAN[4,5,6] and NAOMI E. PIERCE[1] [1]Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA, [2]32 River St., Cambridge, MA, USA, [3]300 Allston St., Brighton, MA, USA, [4]Biology Department, City College of New York, City University of New York, New York, NY, USA, [5]Ph.D. Program in Biology, Graduate Center, City University of New York, New York, NY, USA and [6]Entomology Section, National Museum of Natural History, Philippines, Manila, Philippines

Abstract. 1. Conservation 'culturomics' analyses changes in word frequencies within large digital sources to gain conservation insights. Studies of mammals and birds have found little correlation between conservation need and research effort or popular interest. In this study, we examine research and public interest in 9355 potentially endangered insect species from the Union for Conservation of Nature (IUCN) Red List.

2. Using search counts from Google search, Google Scholar and Ngram as indicators of public and research interest (Internet salience), we found that Internet attention is negatively correlated with the level of threat faced by an insect species. All measures of Internet salience and research effort were highly correlated, and searches for an insect's scientific name were more likely if it had a common name.

3. Since few past studies have incorporated phylogenetic information into their analyses, we used search counts and Twitter text content to study the phylogenetic signal of research interest across 870 insect families. Phylogenetic regressions demonstrate that species-rich families receive more searches, but research interest is not proportional to the size of a family. Phylogenetically distinct insect families receive fewer searches per family and per species.

4. Our results suggest that many endangered insects, unlike vertebrates, have been largely ignored on the World Wide Web.

Key words. Conservation culturomics, endangered insects, insect conservation, Internet salience, phylogenetic signal, taxonomic bias.

Introduction

Insects are currently experiencing drastic declines in diversity and abundance (Dunn, 2005; Hallmann *et al*., 2017; Sánchez-Bayo & Wyckhuys, 2019; van Strien *et al*., 2019; Cardoso *et al*., 2020), but entomological knowledge and efforts to conserve insects are taxonomically biased (Clark & May, 2002). Compared with birds and mammals, insects are studied disproportionately less often (Leather *et al*., 2008, Troudet *et al*., 2017) and are perceived as unfamiliar by the public (Simaika & Samways, 2018; Leandro & Jay-Robert, 2019). They are often ignored in conservation policies (Berenbaum, 2008). This relative dearth of research effort and public interest may reflect the intrinsic difficulties of studying insects, which are small, megadiverse, difficult to track, and often have immature stages that are difficult to find (Pawar, 2003; Cardoso *et al*., 2011). However, considering the immense ecosystem services that insects provide (Cardoso *et al*., 2020), this lack of attention is alarming.

Taxonomic bias exists among insects, too. Cardoso (2012) summarised the types of bias in a list of 122 arthropods in the European Habitats Directive as 'taxonomic, geographic, range, size, and aesthetic'. Similarly, Leandro *et al*. (2017) demonstrated through analysis of 15 traits that protected species in

Correspondence: Zhengyang Wang, Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA. E-mail: zhengyangwang@g.harvard.edu

Europe are 'significantly larger, better known, more widespread and more multi-coloured' than unprotected insects. Since human perception of insects influences their pattern of interaction in natural settings (Shipley & Bixler, 2016), it is valuable to investigate which insect taxa garner the most attention. One such method is through analysis of word frequencies in large databases.

Research that gleans conservation insight through analysis of text corpora has been called 'conservation culturomics' (Ladel *et al*., 2016). The most common way of quantifying taxonomic bias in 'public attention' for a focal taxon (also called 'Internet salience' or 'societal preference') is to quantify the number of Internet searches made for that taxon (Żmihorski *et al*., 2013; Kim *et al*., 2014; Murray et *al*., 2015; Correia *et al*., 2016; Troudet *et al*., 2017; Kim *et al*., 2018). Similarly, number of journal publications can be used as a proxy for research effort (Fisher *et al*., 2011; Fleming *et al*., 2016; Tensen, 2018; Dos Santos *et al*., 2020). Number of Internet searches is likely a reliable measure of public interest (Wilson *et al*., 2007) and is strongly correlated with number of published scientific studies (Martín-López *et al*., 2009). Results from different search engines and social media searches are usually similar to one another (Jaric *et al*., 2016).

This nascent field of conservation culturomics is not without caveats. Critics question drawing conclusions from frequency-based trends that do not reflect the abundance of certain words (Pechenick *et al*., 2015; Correia *et al*., 2019). Furthermore, relying on simple word counts to measure 'interest' leaves the rest of the social media content, such as written texts or posts, untapped. These online text corpora reflect individual belief and attitude beyond simple numeric indicators of interest (Vosoughi *et al*., 2018) and can be processed with text sentiment analysis to reveal hotspots of public interest (Li & Wu, 2010)). Much of conservation culturomics has yet to realise the field's true potential with this information.

It might also be possible to detect cultural differences in attitudes towards particular insect taxa (Ressurreição *et al*., 2012; Kanagavel *et al*., 2014). This question is most pertinent to the case of species common names, which differ among languages and cultures. Although Correia *et al*. (2017) found consistent correlations between common and scientific names of birds in six countries with five different languages, it is uncertain whether many searches for the common name of an insect species correlate with more searches for its scientific name. Confirming this relationship could provide an argument for assigning common names to insect species in need of protection.

To the best of our knowledge, explicit evolutionary relationships among insect taxa have not been considered in previous conservation culturomics studies on public and scientific appeal of insects. Despite the observation that research attention is not proportional to the number of species per taxon (Leather *et al*., 2008; Troudet *et al*., 2017), it is possible that patterns of interest may be more nuanced. For example, research effort and public interest could be disproportionately concentrated in phylogenetically distinctive taxa, or research effort and public interest could be disproportionately focused on highly diverse, species-rich taxonomic groups. Since insects have both evolutionarily distinct and species-rich taxa, they are an ideal group to test hypotheses that relate human preferences to a taxon's evolutionary history.

We further hypothesised that scientific research and public attention are not correlated with an insect taxon's conservation status (i.e., endangered species do not receive more attention). This is because the scale of insect conservation is fundamentally different from the task of conserving well-described vertebrate taxa. There are fewer than 30,000 species of birds and mammals; we might reasonably expect the endangered species among them to receive attention in both scientific literature and public discourse (Troudet *et al*., 2017). In contrast, there are over 1 million described insect species, and an estimated 5 million species yet to be described (Stork, 2018). Endangered insects are by definition among the least-commonly encountered species in this megadiverse fauna. The immense diversity confronting insect ecologists, taxonomists, and evolutionary biologists means that the allocation of conservation resources per endangered insect species is miniscule compared to their vertebrate counterparts. Even known insect species are under-studied, and funds to study and conserve them are scarce (Cardoso *et al*., 2011). Despite a few famous exceptions, such as the Lord Howe Island stick insect (Priddel *et al*., 2003), it seems unlikely that the majority of endangered insect species will garner much research attention. Many insects could go extinct without ever leaving a dent in the World Wide Web.

Material and methods

We used the tools of culturomics to test hypotheses of taxonomic bias within public and scientific interest in insects (Supporting Information Fig. S1). All data were collected between 2 and 9 May 2020. We used several measures of internet salience, social network salience, and research effort to explore public and scientific interest in two datasets: (i) 9355 species of insects listed in the IUCN Red List and (ii) 870 insect taxa sampled across a family-level insect phylogeny. Both datasets are used to study whether various measures of internet salience and research effort are consistent with each other. Further, the first dataset is used to investigate whether the conservation status of an insect and whether it has a common name contributes to its scientific or public interest. The second dataset is used to explore phylogenetic signal in the scientific community's choice of research taxon and social media's perception of insect families.

The IUCN Red List dataset

We downloaded data pertaining to all 9355 potentially threatened insect species named in the IUCN Red List (global list, version 2020.1) from https://www.iucnredlist.org/. The IUCN data for each insect species include the scientific name of the species, some English common names (3673), its population trend (increasing: 46, decreasing: 1236, stable: 1208 or unknown: 6410) plus seven levels of assessment as defined by IUCN (data deficient: 2505; extinct: 61; critically endangered: 322; endangered: 641; near threatened: 535; vulnerable: 796; least concern: 4495) (Supporting Information Fig. S2).

For each of the scientific names, we measured the number of relevant results returned from a Google search (number of search results from https://www.google.com) and a Google Scholar search (number of Google Scholar search results, total number of citations for the first 20 results from https://scholar.google.com) using SerpApi (2020), and relevant results from an Ngram search (number of books published since the year 1500 containing the name, relative frequency of the name in the Ngram corpus from https://books.google.com/ngrams) using PhraseFinder API (Trenkmann, 2020). We applied the same methods to obtain the number of relevant results from a Google or Ngram search for the common name of each species.

Insect family-level phylogenetic dataset

We used the family-level insect phylogeny of Rainford *et al*. (2014), which includes relationships among 870 families inferred from nuclear (CAD, Ef1α, PGD,18S, 28S), and mitochondrial (COI, COII, 16S) sequences. While the IUCN list data are organised by species, this tree includes exemplars of 870 insect families. We cross-referenced the Catalogue of Life Check List (Roskov *et al*., 2019) and obtained species tallies for 821 of the 870 (94%) families. We measured how 'distinct' a family is by its average distance from all other branch tips. For each family name, we then measured the number of relevant results returned from a Google, Google Scholar, and Ngram search as described above. We used the Twitter premium API (2020) to obtain the top 500 unique tweets containing the family name from 2 April to 2 May 2020 (the maximum timespan allowed for the chosen API), as well as the total number of retweets. If the content of a tweet was in English (detected by the pycld2 package in Python), we conducted two sets of text sentiment analysis, one with the 'Transformers' package in Python (Wolf *et al*., 2020; trained on the Stanford Sentiment Treebank Dataset: https://deepai.org/dataset/stanford-sentiment-treebank), and one with the "TextBlob" package in Python (Loria, 2018), to assign a value between −1 and 1 to each tweet. High absolute values in the 'Transformers' analysis indicates high confidence in assigning a sentiment value to the text; positive and negative values in "TextBlob" analysis indicate levels of positive and negative sentiments in the text.

Analysis

All analyses were conducted in R v.3.6.2. We log-transformed the number of Internet searches to use as a standard measure of 'Internet salience' (Sitas *et al*., 2009; Żmihorski *et al*., 2013; Correia *et al*., 2016). We then constructed a Pearson correlation matrix of all search results to verify consistency among search results that are proxies of academic interest (e.g., Google Scholar), public interest (Google search), and social media interest (e.g., Twitter). By 'consistency,' we mean that we expect search results of the same taxon from different platforms not to be significantly negatively correlated. We also verified consistency among search results that are count-based (e.g., Google Scholar search counts, tweet counts) and impact-based (Google scholar citation counts, retweet counts). Lastly, we verified consistency between search results based on scientific names and common names. We conducted principal component analysis (PCA) using Google search counts and Google Scholar search counts and citation counts for each species' scientific and common names as variables to visualise our result (R; https://cran.r-project.org).

Using the IUCN Red List dataset, we ran generalised linear models with Google search counts by scientific name as the dependent variable (log-linked, Gaussian distribution, R 'stats' package), and the following potential explanatory variables: (i) level of endangerment (only critically endangered, endangered, near threatened, vulnerable, least concern as five numeric levels, excluding the extinct and data deficient categories); (ii) estimated population trends (four categories, increasing, decreasing, stable, unknown); and (3) whether the taxon has a common name. AIC values were used to select the best model when differences in residual sums between two models were not significant. The correlation between level of endangerment and search counts of a species' scientific name were then verified by running the same regressions with search counts of the species' common name as the dependent variable. To avoid confounding high search volume common names that do not only refer to an endangered species (e.g., 'Apollo' for *Parnassius apollo*), we repeated our common name analysis with only common names longer than three words (e.g., 'Lord Howe Island stick insect') ($n = 1612$) to increase the specificity of the common name reference.

Using the family-level dataset, we examined whether research and public interest are randomly distributed across the phylogeny. We calculated Pagel's λ (Pagel, 1999) to assess the phylogenetic signal of the following results per family as continuous traits: Google search counts, Google Scholar search counts; Google Scholar citation counts; number of unique tweets in the past 30 days; average sentiment value of tweets for the family. We also calculated the phylogenetic signal of species richness and the phylogenetic distinctiveness of each family – commonly known as the phylogenetic signal of clade size and clade age (Rabosky *et al*., 2012). We then calculated the phylogenetic signal of the number of family searches divided by species per family (i.e., the number of searches for each species in a family).

We ran generalised linear models with Google search counts of family names as dependent variables (log-linked, Gaussian distribution, R 'stats' package) and phylogenetic distinctiveness of the family as explanatory variables. Since phylogenetic distinctiveness has a strong phylogenetic signal, we further conducted a phylogenetic generalised least squares analysis (Ives *et al*., 2007) to examine whether any signal is detectable after accounting for phylogenetic structure. In the phylogenetic regression models, we used either Google search count per family or per species as the dependent variable with one of four explanatory variables: (i) the results of Google search, Ngram search, Twitter search; (ii) mean sentiment polarity for each family; (iii) species count for each family; and (iv) phylogenetic distinctiveness of each family.

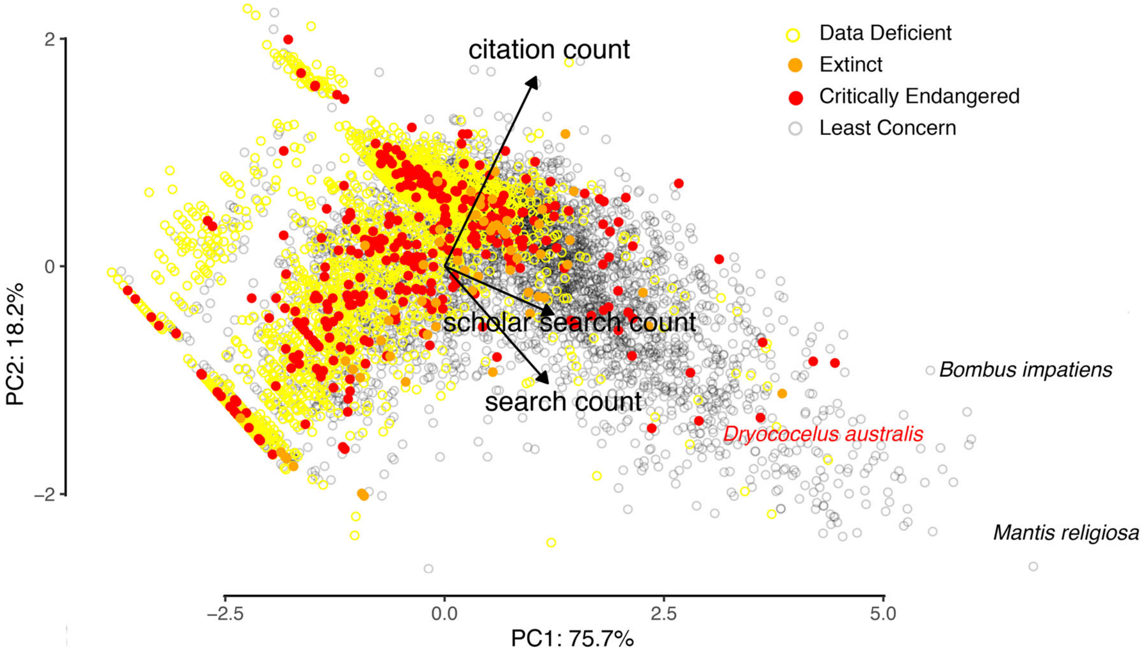

Fig 1. Critically endangered insect species (red dots) from the 9355 species listed on the IUCN Red List are under-researched (left side of principal component analysis plot). Arrows indicate increased Google search counts, Google Scholar search counts, and Google Scholar citation counts on a log scale. Only a few flagship species such as the Lord Howe Island stick insect (*Dryococelus australis*) are the subject of much public interest and research attention, while most critically endangered insect species (red dots) are under-researched. Species of least conservation concern (grey circles) are more evenly distributed.

Results and discussion

Endangered insects receive less attention than common ones

The number of results from searching a species' scientific name in Google, Google Scholar, and Ngram were significantly positively correlated (Supporting Information Table S1; Fig. S3). In our PCA, the first axis of variation, representing quantity of search counts (on a log scale), explains 75.7% of the variation in our data (Fig. 1). Our measures of public interest and research effort were highly positively correlated ($P < 0.001$), which is consistent with the result of public perception driving scientific funding (Martín-López *et al*., 2009), or effectively the converse, with new scientific knowledge driving public discourse. Further, the positive correlation between search counts of common and scientific names was highly significant ($P < 0.001$), which is consistent with results of Correia *et al*. (2017) and Jaric *et al*. (2016), who found positive correlations between search counts of common and scientific names for species of birds and mammals. The model that best explains the Google search count for species scientific names includes (i) level of endangerment, (ii) population trend, and (iii) presence of common names, with no interactions between variables (pseudo-$R^2 = 0.14$; Supporting Information Table S4). Species with increasing population trends have more searches ($P < 0.001$), while those of unknown population trends have fewer searches ($P = 0.03$) (Supporting Information Fig. S4). Species with common names received more searches for their scientific names ($P = 0.006$), supporting the idea that common names increase the recognisability of the insect (Berenbaum, 2008; New, 2008). Alternatively, there is little impetus to give an insect a common name if it only has locally constrained ranges or if it is rarely encountered by the non-scientific public.

The more endangered an insect is, the less often people search for it. A species' level of endangerment is negatively correlated with number of Google searches ($P = 0.015$). The same is true of Google Scholar searches, and of both Google and Google Scholar that search by common name instead of scientific name (Supporting Information Fig. S6). In fact, across all measures, public and scientific attention to *extinct* insects is higher than to *critically endangered* insects (Fig. 2; Fig S6)! A similarly negative relationship between number of searches and degree of threat was found in a conservation culturomics study of 236 bird species in Brazil (Corrêa *et al*., 2016), while a weakly positive correlation was found by Dos Santos *et al*. (2020) studying 4108 non-marine mammal species. Kim *et al*. (2014, 2018) also found a positive correlation in lists of endangered species (all taxa) in Korea and Japan. Other researchers found no such pattern in European birds (Murray *et al*., 2015), coral reefs (Fisher *et al*., 2011), felids, and canids (Tensen, 2018).

If one goal of IUCN threat evaluation is to raise awareness about endangered species, this evidence suggests that while larger terrestrial mammals enjoy public and scientific recognition (Kim *et al*., 2014, 2018; Dos Santos *et al*., 2020), threatened insects decline towards extinction in relative anonymity. There

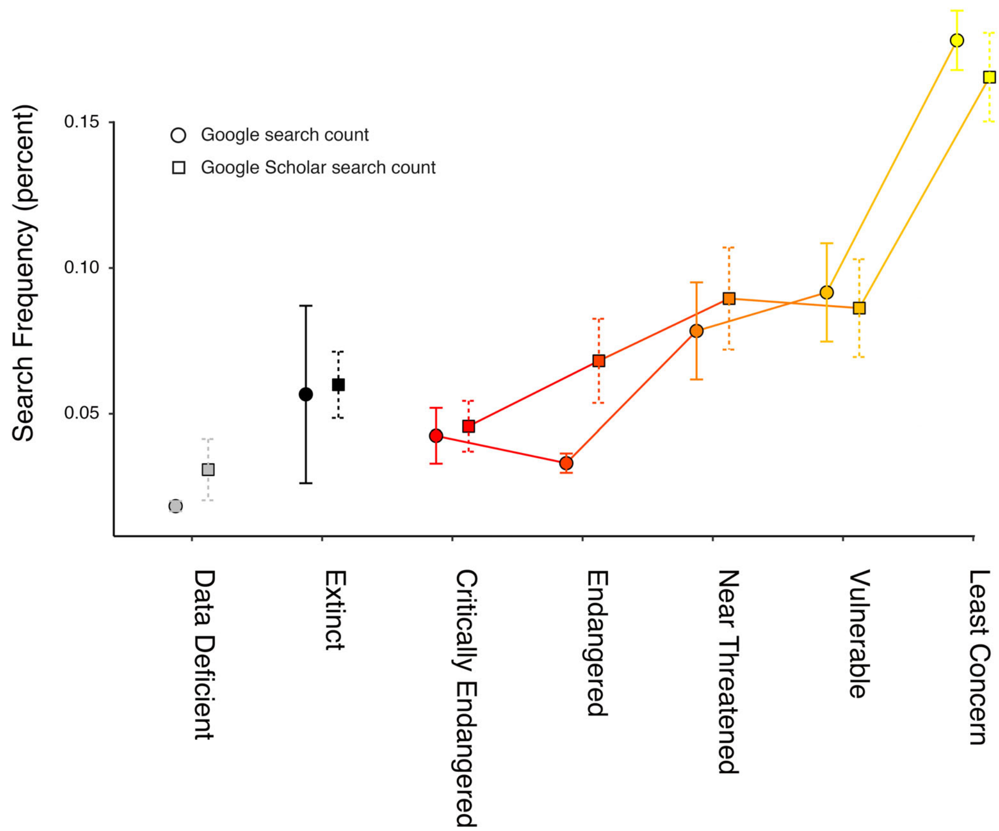

Fig 2. Critically endangered species receive fewer Google and Google Scholar searches than insect species of least concern. Circles and squares indicate the mean frequency of Google and Google Scholar search counts, respectively, examined for each IUCN Red List category in this study (bars indicate standard errors).

are several reasons for this omission. First, compared with vertebrates, we know relatively little about each of the 8–10 million potential insect species on earth (Stork, 2018). Endangered or threatened insects are usually less abundant or only regionally abundant and therefore infrequently encountered in comparison with species in less dire IUCN categories. Fewer searches for these rare insects may reflect the intrinsic difficulty of finding and studying them (Cardoso *et al.*, 2011). Secondly, most applied entomological research funding and associated publications are aimed at killing insect pests or disease vectors, not protecting them (Kegan & Prokopy, 2009; Smith & Kennedy, 2009). Insect conservation for the sake of their ecosystem services has only recently gained traction given the significant economic implications of the subject (Losey & Vaughan, 2006; Dangles & Casas, 2019; Cardoso *et al.*, 2020). Finally, insect conservation is primarily achieved through the lens of habitat conservation (Hebel *et al.*, 2019), which aims to preserve insect (meta)communities and their ecological functions. Despite alarm calls about global insect decline, the fate of endangered insects has not received tailored attention.

In our study, conservation success stories for endangered insect species tend to be outliers among the critically endangered and data deficient species in the PCA plots (Fig. 1; Supporting Information Fig. S5), along with a few charismatic insects frequently seen on the black market (such as the Golden Kaiser I-Hind butterfly, *Teinopalpus aureus*, Wang *et al.*, 2018; Xing *et al.*, 2019). Most endangered species lie on the left side of the figure, indicating low to negligible public interest and research effort. Outside a short list of well-known species, most endangered insects seem to reside in a forgotten corner of the World Wide Web.

Species-rich, phylogenetically distinct insect clades receive less research attention

All measures of Internet salience and research effort measured on Google, Google Scholar, Ngram, and Twitter are positively correlated (Fig. S7), mirroring the IUCN search results (-Table S2). In the PCA plot, the first axis, indicating research output and public interest, accounts for 71.3% of the variation (Fig. S8). Although our data represent only 1 month of Twitter posts, we found that insect family names with more Google and Google Scholar searches are indeed tweeted and re-tweeted more often. The difference between Tweet counts and Google search counts explains 11.9% of the variation on the second PCA axis (Fig. S8), which suggests an incomplete overlap between the two.

The results of our text sentiment analysis are limited to the English language. Among a total of 8723 tweets, 4855 were in English (followed by 683 in Spanish, 664 in Japanese, and 568 in Portuguese). Most tweets in English convey a single sentiment polarity (i.e., bimodal distribution in 'Transformers' analysis, Fig. S9) and the average skews towards slightly positive sentiments ("TextBlob" individual mean = 0.12, Sc.D. = 0.25, family mean = 0.11, s.d. = 0.15). We note that tweets are trendy, and this study analyses tweets from only a single month (April) in 2020. These tweets could carry seasonal or geographical bias because users in the northern and southern hemispheres might observe and tweet about different insects in April. However, if tweets more closely reflected academic publications, seasonal biases should be less prominent. In this study, we are also limited by the number of query words we could access; we might have missed more nuanced patterns beyond mere mention of insect family names, such as hashtags of insect common names in different languages. As Twitter begins to open up its dataset to researchers (Twitter Developers Forums, 2021), we expect more temporally, spatially, and linguistically fine-scaled culturomics studies to emerge.

Family names of species-rich families receive more searches, even accounting for clade size (phylogenetic regression, $R^2 = 0.295$, $P < 0.001$; Fig. 3a linear regression), suggesting heightened research interest in hyper-diverse families such as Curculionidae, Carabidae, and Noctuidae (Fig. 4). However, this increased interest is not proportional to the species richness of a family. The number of searches for a family per species (i.e., number of searches for a family divided by its size in terms of number of species) is negatively correlated with species richness of the family (phylogenetic regression, $R^2 = 0.470$, $P < 0.001$)(Fig. 3b). In other words, although families that are more species-rich receive more searches per family overall, species in those families receive fewer searches.

We measured the phylogenetic distinctiveness of each insect family by calculating its average distance from all other branch tips. The results of this measurement might be numerically different from the formal definition of 'evolutionary distinctiveness' ED (Isaac *et al.*, 2007). We found a negative correlation between the number of searches and the phylogenetic distinctiveness of a family ($R^2 = 0.070$, $P < 0.001$) (Fig. 3c); this negative correlation still holds when the number of searches is normalised by species richness per family ($R^2 = 0.014$, $P < 0.001$); although in both cases the variance in searches explained by phylogenetic distinctiveness is relatively low. These results provide some evidence that phylogenetically distinct families receive fewer searches per family and per species. In contrast, this bias against phylogenetically distinct taxa is not found in terrestrial mammals (Tensen, 2018; dos Santos *et al.*, 2020).

Phylogenetic signal was not significant in any of the search results, nor in sentiments towards families (Table S3). This suggests that interest in insect families is more likely to relate to their perceived importance or negative impact on humankind (Lounibos, 2002; Cardoso *et al.*, 2020). As shown in this study, insect families garnering the highest interest are those that interact with crop plants or potentially vector disease (e.g., Apidae see Owen, 2017, Culicidae see Killeen *et al.*, 2006). Other high-interest families are viewed as pretty or charismatic (e.g., butterfly families, Fleishman and Murphy, 2009), or as pests or popular research taxa (e.g., Formicidae see Leach *et al.*, 2013, Carabidae see Allen, 1979) (Fig. 4). There is no intrinsic reason why insects of interest to humans might exhibit phylogenetic signal. On the other hand, the dual contributions of evolutionary history and research interest to species richness are difficult to disentangle. Some taxonomic property that has a strong phylogenetic signal, such as species richness itself, could be correlated with human interest (Poe *et al.*, 2021). An insect family could be species-rich because it occurs in parts of the world with longer traditions of entomological research and thus is more intensively studied.

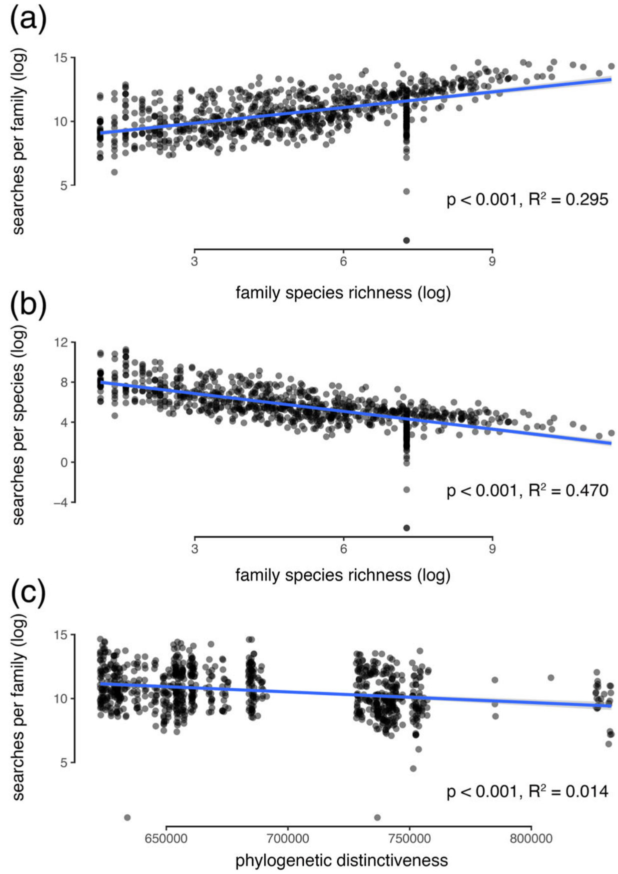

Fig 3. Google search counts are positively correlated with species richness and negatively correlated with phylogenetic distinctiveness of insect families (*sensu* Rainford *et al.*, 2014) ($n = 821$). (a) Google search counts for insect families are positively correlated with the number of species in the family, but (b) the average Google search count per species is negatively correlated with the number of species in its family. (c) Google search counts for an insect family are negatively correlated with its distinctiveness, measured as its average distance from all other branch tips on the phylogeny.

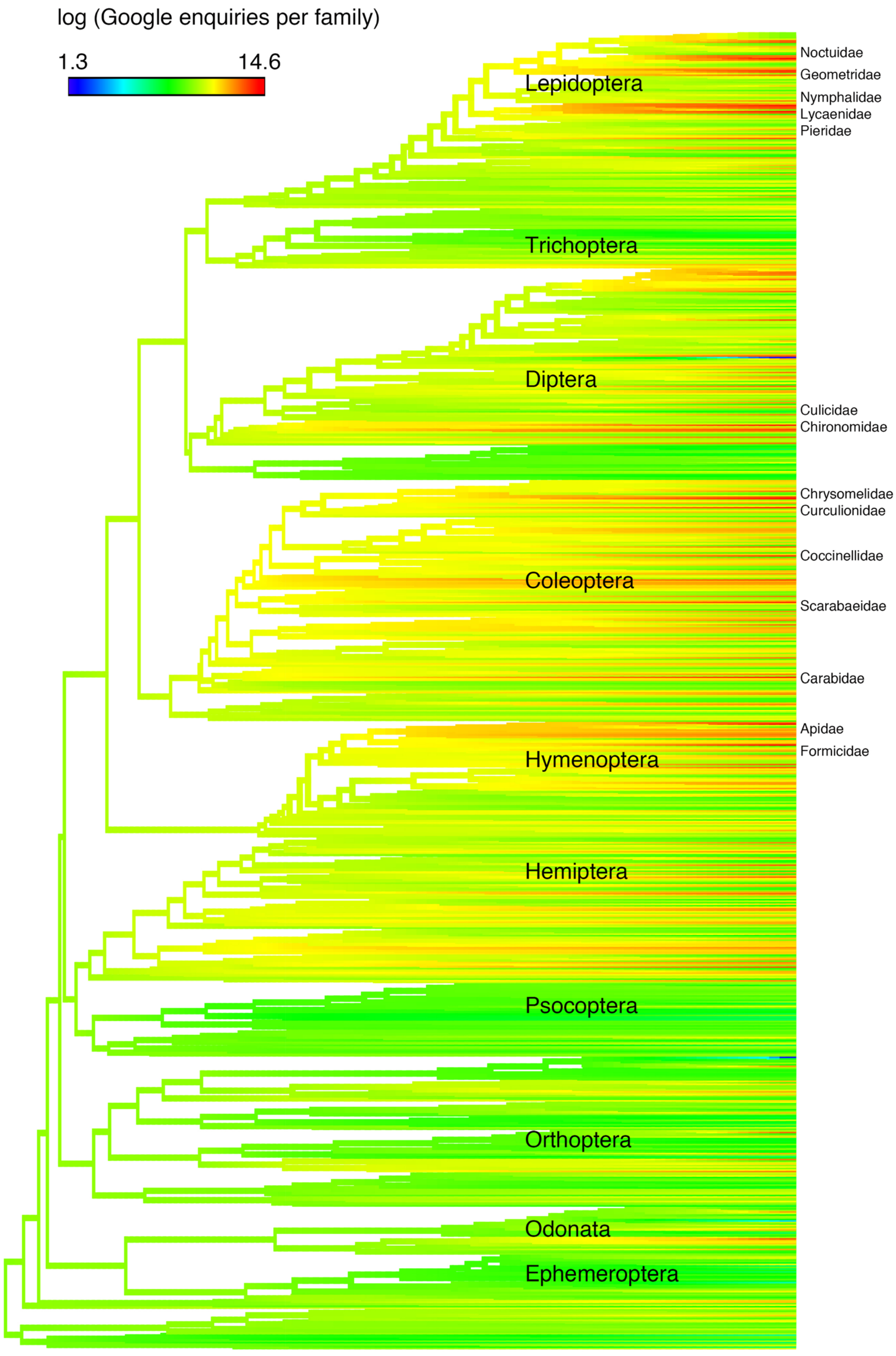

Fig 4. Google search counts for 870 insect families mapped as a continuous trait onto an insect phylogeny (Rainford *et al*., 2014 show no phylogenetic signal but are positively correlated with the species richness for each family. Large insect orders are labelled. The 14 insect families with the most searches are also labelled.

Acknowledgements

Z.W. and J.Z. thank Professors Sara Hiebert Burch and Rachel Merz from Swarthmore College for cultivating their interest in biology. Z.W. was supported by a graduate fellowship from the Department of Organismic and Evolutionary Biology at Harvard University. D.J.L. was supported by National Science Foundation (NSF) DEB-1541557 and N.E.P was supported by NSF DEB-1541560. The authors thank Allison King and Samuel Church for helpful discussions, Sofia Prado-Irwin for commenting on an earlier draft, and two anonymous reviewers for their constructive comments on the manuscript.

Author contributions

Z.W. conceived and designed the study. J.Z. and W.M collected the dataset. Z.W. and J.Z analysed the data. Z.W., D.J.L. and N.E.P. wrote the article.

Ethics Statement

Ethics approval was not required for this study.

Data availability statement

The data that support the findings of this study are openly available in at: https://github.com/jzengg/insect-conservation

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1: Supporting information

References

Allen, R.T. (1979) The occurrence and importance of ground beetles in agricultural and surrounding habitats. *Carabid Beetles*. (ed. by T.L. Erwin, G.E. Ball, D.R. Whitehead and A.L. Halpern). Springer, Dordrecht.

Berenbaum, M. (2008) Insect conservation and the Entomological Society of America. *American Entomologist*, **54**, 117–120.

Cardoso, P., Erwin, T.L., Borges, P.A.V. & New, T.R. (2011) The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, **144**, 2647–2655.

Cardoso, P. (2012) Habitats directive species lists: urgent need of revision. *Insect Conservation and Diversity*, **5**, 169–174.

Cardoso, P., Barton, P.S., Birkhofer, K., Chichorro, F., Deacon, C., Fartmann, T., Fukushima, C.S., Gaigher, R., Habel, J.C., Hallmann, C.A., Hill, M.J., Hochkirch, A., Kwak, M.L., Mammola, S., Ari Noriega, J., Orfinger, A.B., Pedraza, F., Pryke, J. S., Roque, F.O., Settele, J., Simaika, J.P., Stork, N.E., Suhling, F., Vorster, C. & Samways, M.J. (2020) Scientists' warning to humanity on insect extinctions. *Biological Conservation*, **242**, 108426.

Clark, J.A. & May, R.M. (2002) Taxonomic bias in conservation research. *Science*, **297**, 191–192.

Correia, R.A., Jepson, P.R., Malhado, A.C.M. & Ladle, R.J. (2016) Familiarity breeds content: assessing bird species popularity with culturomics. *PeerJ*, **4**, e1728.

Correia, R.A., Jepson, P., Malhado, A.C.M. & Ladle, R.J. (2017) Internet scientific name frequency as an indicator of cultural salience of biodiversity. *Ecological Indicators*, **78**, 549–555.

Correia, R.A., Di Minin, E., Jarić, I., Jepson, P., Ladle, R., Mittermeier, J., Roll, U., Soriano-Redondo, A. & Veríssimo, D. (2019) Inferring public interest from search engine data requires caution. *Frontiers in Ecology and the Environment*, **17**, 254–255.

Dangles, O. & Casas, J. (2019) Ecosystem services provided by insects for achieving sustainable development goals. *Ecosystem Services*, **35**, 109–115.

dos Santos, J.W., Correia, R.A., Malhado, A.C.M., Campos-Silva, J.V., Teles, D., Jepson, P. & Ladle, R.J. (2020) Drivers of taxonomic bias in conservation research: a global analysis of terrestrial mammals: Conservation research bias. *Animal Conservation*, **22**(6), 679–688. https://doi.org/10.1111/acv.12586.

Dunn, R.R. (2005) Modern insect extinctions, the neglected majority. *Conservation Biology*, **19**, 1030–1036.

Fleishman, E. & Murphy, D.D. (2009) A realistic assessment of the indicator potential of butterflies and other charismatic taxonomic groups. *Conservation Biology*, **23**, 1109–1116.

Fisher, R., Radford, B.T., Knowlton, N., Brainard, R.E., Michaelis, F. B. & Caley, M.J. (2011) Global mismatch between research effort and conservation needs of tropical coral reefs: global distribution of coral reef research. *Conservation Letters*, **4**, 64–72.

Fleming, P.A. & Bateman, P.W. (2016) The good, the bad, and the ugly: which Australian terrestrial mammal species attract most research? *Mammal Review*, **46**, 241–254.

Habel, J.C., Samways, M.J. & Schmitt, T. (2019) Mitigating the precipitous decline of terrestrial European insects: requirements for a new strategy. *Biodiversity and Conservation*, **28**, 1343–1360.

Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., Goulson, D. & de Kroon, H. (2017) More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS One*, **12**, e0185809.

Isaac, N.J., Turvey, S.T., Collen, B., Waterman, C. & Baillie, J.E. (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE*, **2**, e296.

Ives, A.R., Midford, P.E. & Garland, T. (2007) Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology*, **56**, 252–270.

Jarić, I., Courchamp, F., Gessner, J. & Roberts, D.L. (2016) Data mining in conservation research using Latin and vernacular species names. *PeerJ*, **4**, e2202.

Kanagavel, A., Raghavan, R. & Veríssimo, D. (2014) Beyond the "general public": implications of audience characteristics for promoting species conservation in the western Ghats hotspot, India. *AMBIO*, **43**, 138–148.

Killeen, G.F., Kihonda, J., Lyimo, E., Oketch, F.R., Kotas, M.E., Mathenge, E., Schellenberg, J.A., Lengeler, C., Smith, T.A. & Drakeley, C.J. (2006) Quantifying behavioural interactions between humans and mosquitoes: evaluating the protective efficacy of insecticidal nets against malaria transmission in rural Tanzania. *BMC Infectious Diseases*, **6**, 161.

Kim, J.Y., Do, Y., Im, R.-Y., Kim, G.-Y. & Joo, G.-J. (2014) Use of large web-based data to identify public interest and trends related to endangered species. *Biodiversity and Conservation*, **23**, 2961–2984.

Kim, J.Y., Noda, A., Im, R.-Y. & Nishihiro, J. (2018) Web search volume as a surrogate of public interest in biodiversity: a case study of Japanese red list species. *Ecosystem Health and Sustainability*, **4**, 289–298.

Kogan, M. & Prokopy, R. (2009) Agricultural entomology. *Encyclopedia of Insects*. (ed. by V.H. Resh and R.T. Cardé), Vol. **53**, 2nd Edn, pp. p. 4–8.Academic Press, Elsevier, London, UK.

Ladle, R.J., Correia, R.A., Do, Y., Joo, G.-J., Malhado, A.C., Proulx, R., Roberge, J.-M. & Jepson, P. (2016) Conservation culturomics. *Frontiers in Ecology and the Environment*, **14**, 269–275.

Leach, E., Nakamura, A., Turco, F., Burwell, C.J., Catterall, C.P. & Kitching, R.L. (2013) Potential of ants and beetles as indicators of rainforest restoration: characterising pasture and rainforest remnants as reference habitats. *Ecological Management & Restoration*, **14**, 202–209.

Li, N. & Wu, D.D. (2010) Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, **48**, 354–368.

Leandro, C., Jay-Robert, P. & Vergnes, A. (2017) Bias and perspectives in insect conservation: a European scale analysis. *Biological Conservation*, **215**, 213–224.

Leandro, C. & Jay-Robert, P. (2019) Perceptions and representations of animal diversity: where did the insects go? *Biological Conservation*, **237**, 400–408.

Leather, S.R., Basset, Y. & Hawkins, B.A. (2008) Insect conservation: finding the way forward. *Insect Conservation and Diversity*, **1**, 67–69.

Loria, S. (2018) textblob Documentation. Release 0.15, 2.

Losey, J.E. & Vaughan, M. (2006) The economic value of ecological services provided by insects. *BioScience*, **56**, 311–323.

Lounibos, L.P. (2002) Invasions by insect vectors of human disease. *Annual Review of Entomology*, **47**, 233–266.

Martín-López, B., Montes, C., Ramírez, L. & Benayas, J. (2009) What drives policy decision-making related to species conservation? *Biological Conservation*, **142**, 1370–1380.

Murray, H., Green, E., Williams, D., Burfield, I. & de Brooke, M. (2015) Is research effort associated with the conservation status of European bird species? *Endangered Species Research*, **27**, 193–206.

New, T.R. (2008) What's in common names: are they really valuable in insect conservation? *Journal of Insect Conservation*, **12**, 447–449.

Owen, R. (2017) Role of human action in the spread of honey bee (Hymenoptera: Apidae) pathogens. *Journal of Economic Entomology*, **110**, 797–801.

Pagel, M. (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

Pawar, S. (2003) Taxonomic chauvinism and the methodologically challenged. *BioScience*, **53**, 861.

Pechenick, E.A., Danforth, C.M. & Dodds, P.S. (2015) Characterizing the Google Books corpus: strong limits to inferences of socio-cultural and linguistic evolution. *PLoS ONE*, **10**, e0137041.

Poe, S., Anderson, C. & Barnett, J. (2021) On the selection and analysis of clades in comparative evolutionary studies. *Systematic Biology*, **70**, 190–196.

Priddel, D., Carlile, N., Humphrey, M., Fellenberg, S. & Hiscox, D. (2003) Rediscovery of the 'extinct' Lord Howe Island stick-insect (*Dryococelus australis* (Montrouzier)) (Phasmatodea) and recommendations for its conservation. *Biodiversity and Conservation*, **12**, 1391–1403.

Rabosky, D.L., Slater, G.J. & Alfaro, M.E. (2012) Clade age and species richness are decoupled across the eukaryotic tree of life. *PLoS Biology*, **10**, e1001381.

Rainford, J.L., Hofreiter, M., Nicholson, D.B. & Mayhew, P.J. (2014) Phylogenetic distribution of extant richness suggests metamorphosis is a key innovation driving diversification in insects. *PLoS ONE*, **9**, e109085.

Ressurreição, A., Gibbons, J., Kaiser, M., Dentinho, T.P., Zarzycki, T., Bentley, C., Austen, M., Burdon, D., Atkins, J., Santos, R.S. & Edwards-Jones, G. (2012) Different cultures, different values: The role of cultural variation in public's WTP for marine species conservation. *Biological Conservation*, **145**, 148–159.

Roskov, Y., Ower, G., Orrell, T., Nicolson, D., Bailly, N., Kirk, P.M., Bourgoin, T., DeWalt, R.E., Decock, W., Nieukerken, E., van Zarucchi, J. & Penev, L. (2019) *Species 2000 & ITIS Catalogue of Life, 2019 Annual Checklist*. Naturalis, Leiden, the Netherlands.

Sanchez-Bayo, F. & Wyckhuys, K.A. (2019) Worldwide decline of the entomofauna: a review of its drivers. *Biological Conservation*, **232**, 8–27.

SerpApi. (2020) <https://serpapi.com/> 2nd May 2020

Shipley, N.J. & Bixler, R.D. (2016) On the need to interpret insects: an always small but gargantuan opportunity. *Journal of Interpretation Research*, **21**, 65–72.

Sitas, N., Baillie, J.E.M. & Isaac, N.J.B. (2009) What are we saving? Developing a standardized approach for conservation action. *Animal Conservation*, **12**, 231–237.

Simaika, J.P. & Samways, M.J. (2018) Insect conservation psychology. *Journal of Insect Conservation*, **22**, 635–642.

Smith, E. & Kennedy, G. (2009) History of Entomology. *Encyclopedia of insects*. (ed. by V.H. Resh and R.T. Cardé), Vol. **53**, 2nd Edn, pp. p. 449–458.Academic Press, Elsevier, London, UK.

Stork, N.E. (2018) How many species of insects and other terrestrial arthropods are there on earth? *Annual Review of Entomology*, **63**, 31–45.

Tensen, L. (2018) Biases in wildlife and conservation research, using felids and canids as a case study. *Global Ecology and Conservation*, **15**, e00423.

Trenkmann, M. (2020) PhraseFinder – Search millions of books for language use. <https://phrasefinder.io> 2nd May 2020.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. (2017) Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, **7**, 9132.

Twitter API. (2020) <https://developer.twitter.com/en/docs/basics/getting-started> 2nd May 2020

Twitter Developers Forums. (2021) <https://twittercommunity.com/t/introducing-the-new-academic-research-product-track/148632> 26th Feb 2021

van Strien, A.J., van Swaay, C.A.M., van Strien-van Liempt, W.T.F.H., Poot, M.J.M. & WallisDeVries, M.F. (2019) Over a century of data reveal more than 80% decline in butterflies in The Netherlands. *Biological Conservation*, **234**, 116–122.

Vosoughi, S., Roy, D. & Aral, S. (2018) The spread of true and false news online. *Science*, **359**, 1146–1151.

Wang, Z., Huang, Y., Luo, X., Qin, K., Merz, R. & Zhou, S. (2018) Habitat monitoring of an endangered Asian butterfly, *Teinopalpus aureus* (Lepidoptera: Papilionidae) and change in local residents' conservation awareness. *Journal of Insect Conservation*, **22**, 721–729.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T.L., Gugger, S., Drame, M., Lhoest, Q. & Rush, A.M. (2020) HuggingFace's transformers: state-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, 38–45.

Wilson, J.R., Procheş, Ş., Braschler, B., Dixon, E.S. & Richardson, D.M. (2007) The (bio)diversity of science reflects the interests of society. *Frontiers in Ecology and the Environment*, **5**, 409–414.

Xing, S., Au, T.F., Dufour, P.C., Cheng, W., Landry Yuan, F., Jia, F., Vu, L.V., Wang, M. & Bonebrake, T.C. (2019) Conservation of data deficient species under multiple threats: lessons from an iconic tropical butterfly (*Teinopalpus aureus*). *Biological Conservation*, **234**, 154–164.

Żmihorski, M., Dziarska-Pałac, J., Sparks, T.H. & Tryjanowski, P. (2013) Ecological correlates of the popularity of birds and butterflies in Internet information resources. *Oikos*, **122**, 183–190.

Accepted 27 April 2021

Editor: Raphael Didham; Associate Editor: Catherine Matilda (Tilly) Collins