



HAL
open science

An expert-curated global database of online newspaper articles on spiders and spider bites

Stefano Mammola, Jagoba Malumbres-Olarte, Valeria Arabesky, Diego Alejandro Barrales-Alcalá, Aimee Lynn Barrion-Dupo, Marco Antonio Benamú, Tharina Bird, Maria Bogomolova, Pedro Cardoso, Maria Chatzaki, et al.

► To cite this version:

Stefano Mammola, Jagoba Malumbres-Olarte, Valeria Arabesky, Diego Alejandro Barrales-Alcalá, Aimee Lynn Barrion-Dupo, et al.. An expert-curated global database of online newspaper articles on spiders and spider bites. *Scientific Data*, 2022, 9 (1), pp.109. 10.1038/s41597-022-01197-6 . hal-03657025

HAL Id: hal-03657025

<https://univ-rennes.hal.science/hal-03657025>

Submitted on 6 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



OPEN

DATA DESCRIPTOR

An expert-curated global database of online newspaper articles on spiders and spider bites

Stefano Mammola *et al.*[#]

Mass media plays an important role in the construction and circulation of risk perception associated with animals. Widely feared groups such as spiders frequently end up in the spotlight of traditional and social media. We compiled an expert-curated global database on the online newspaper coverage of human-spider encounters over the past ten years (2010–2020). This database includes information about the location of each human-spider encounter reported in the news article and a quantitative characterisation of the content—location, presence of photographs of spiders and bites, number and type of errors, consultation of experts, and a subjective assessment of sensationalism. In total, we collected 5348 unique news articles from 81 countries in 40 languages. The database refers to 211 identified and unidentified spider species and 2644 unique human-spider encounters (1121 bites and 147 as deadly bites). To facilitate data reuse, we explain the main caveats that need to be made when analysing this database and discuss research ideas and questions that can be explored with it.

Background & Summary

Spiders have an unfortunate reputation. There are tales about massive infestations of false black widows shutting down entire schools; apocryphal stories of dangerous arachnids lurking under toilet seats of international airports; and urban myths of tiny spiders crawling into your mouth while you are asleep. Of course, these are just anecdotes, but they illustrate how, even today, arachnophobic sentiments permeate our society at all levels^{1–4}. This is nothing surprising: arachnophobia is likely the most widespread fear related to animals⁵, with an estimated prevalence between 3.5–11.4% of the world population^{6–9}. However, such a skewed perception towards the potential harm that spiders can cause humans contrasts with two facts. First, less than 0.5% of spider species can cause severe envenomation in humans¹⁰. Second, the habitat of these few potentially dangerous species rarely overlaps with that of humans, making dangerous human-spider encounters unlikely¹¹. Since a limited number of fatalities due to spider bites have occurred in the past few decades^{12–16}, the reasons behind our exaggerated perception of risk associated with spiders remain uncertain.

Despite gigantic leaps forward in cognitive science and neurology, we still do not know the exact reason why arachnophobia is so widespread. What we do know is that arachnophobic sentiments have a significant social and cultural component^{17–19}. For example, a recent study suggested that arachnophobic behaviours may be reduced following an exposure to the superhero movie *Spider-Man*²⁰. Likewise, it was suggested that ongoing urbanization is the key driver of the prevalence of disgust for insects and spiders, because exposure to animals in urban areas is less frequent²¹. Accordingly, it is reasonable to hypothesize that the public perception of spiders should be affected by how spider-related information is framed and circulated², given that traditional media are particularly effective in disseminating knowledge by conveying messages quickly and reaching a broad audience²². From the belief in the role of bats as disease spreaders^{23,24} to the fear of being attacked by large carnivores^{25,26}, the crucial role of traditional media in the construction and circulation of risk perception associated with wild animals is undisputed²⁷. However, media representation of spiders is still a poorly studied subject: as far as we are aware, the only two available studies are focused on a local selection of news in Australia²⁸ and Italy².

Intending to fill this gap, here we compiled an expert-curated global database on the coverage of human-spider encounters in online newspaper media and their accuracy and reliability over the past ten years

[#]A full list of authors and their affiliations appears at the end of the paper.



That day Gary understood the potential and the risks of news-related data

Fig. 1 The potential of news articles as a source of data. This database offers a quantitative baseline to pursue research on the human-dimensions of spiders and their representation in the media. This research may include answering questions related to the cultural component of spider conservation, evaluation of people's perceptions of spiders via opinion mining techniques, and generation of ecological insights, among others. Original illustration by Jagoba Malumbres-Olarte.

(2010–2020) (Fig. 1). This database includes detailed information about the location of each human-spider encounter reported by the media, year of publication, and a quantitative characterisation of the content of each piece of news (presence of photographs of spiders and bites, number and type of errors, whether experts were consulted, and whether the content is sensationalistic or neutral). With this database, we hope to stimulate further research on the human dimension of spiders and their representation in the media.

Methods

Geographical coverage of the database. We aimed to compile a comprehensive database of global coverage of countries, languages and online newspapers. To this end, we put together a large network of spider experts to mine data in as many languages and countries as possible. We searched for news in 40 unique languages and covered the online press in 81 countries and all six continents where spiders can be found. Due to an uneven availability of experts, however, there is a bias in the database towards temperate regions (Europe and North America). African countries are the least represented.

Temporal coverage of the database. We focused on newspaper articles published online between 2010 and 2020 (partial and uneven temporal coverage for 2020). Thus, the temporal span of our study mostly covered the advent of online journalism and the parallel diffusion of news through social media platforms²⁹.

Data mining protocol. We adapted the methodology of Mammola *et al.* (ref. ²) for retrieving news articles on human-spider encounters published in online newspapers in the target countries. To ensure that different authors in charge of different countries and languages adhered to an unequivocal data mining strategy, we began by preparing a general protocol for retrieving and extracting information from the news (Appendix S1). This protocol, shared with all authors, included: i) instructions for media report retrieval and data mining (see below); ii) a continuously updated list of Frequently Asked Questions discussing how to handle specific cases; iii) a description of the most common envenomation symptoms, which was used to standardize the assessment of the errors related to spider venom (see next section).

For each country and language, we carried out online searches in different languages with Google News (Appendix S1 – Figure S1), choosing multiple keyword combinations and the years between 2010 and

2020—this can be specified using the ‘Custom range’ tool in Google News (Appendix S1 – Figure S2). Note that for a few countries that were not available in Google News at the time of the search, such as Finland, Denmark, or Botswana, we performed searches directly in Google. We first searched for the words for ‘spider’ in each language, followed by the word ‘bite’ (e.g., “spider bite”). We repeated the search using the word ‘sting’ instead of ‘bite’, which is anatomically incorrect but often used³⁰. We then repeated the search by changing the general word ‘spider’ with scientific and vernacular names of the main species perceived as “dangerous” in each country. These include species that are considered medically important (e.g., *Latrodectus* spp., *Atrax* spp., *Loxosceles* spp., and *Phoneutria* spp.; list of genera in Appendix S1 – Table S1) and/or widely feared (e.g., *Cheiracanthium* spp., *Lampona* spp., and *Steatoda* spp.; list of genera in Appendix S1 – Table S2). The list of species names used in searches for each country was tailored by each author, based on their expertise and knowledge of the spider fauna and the scientific literature on spider bites in the assigned country. We noted, however, that including additional search terms besides ‘spider bite’ and ‘spider sting’ yielded diminishing returns since these two broad keywords usually covered the vast majority of relevant news articles.

For each unique keyword search combination, we manually inspected all news up to the final available page in Google News, systematically collecting news articles referring to one or more purported encounters between humans and spiders. We included i) all news articles referring to a human-spider encounter (e.g., a family interviewed by a local newspaper about the spider they found in their house; a farmer bitten by a spider while working in the field; a person who was hospitalized following a spider bite); and ii) events that occurred either in the searched country or abroad (e.g., an Indian newspaper talking about a biting event that occurred in England). Note that we included all reports of human-spider interactions regardless of the likelihood of a spider actually being involved (e.g., a person claiming to have a spider bite when they likely had a skin infection instead). We disregarded: i) media items reporting general facts about spiders, venomous spiders, arachnophobia, spider-related research findings, and doctors’ advice about what to do in case of a spider bite; and ii) blog posts.

Extraction of information for each news article. For each news article, we extracted the qualitative and quantitative information detailed in Table 1. We first reported basic information: a) URL; b) title; c) date of publication; d) newspaper name (or news outlet if not a traditional newspaper); e) newspaper type, broken down into “Traditional newspaper” (Official newspaper in a country, with both a printed and an online version), “Online newspaper” (online only newspaper), or “Magazine” (for magazines, tabloids, and similar); and f) newspaper circulation (regional, national, or international).

Then, we read the full article and scored the g) spider species identity based on the description in the news article, even if the attribution was incorrect based on our expert opinion. We reported species identity to the lowest taxonomic level possible based on the information in the article (typically species or genus level, but sometimes only family level). If the species identification was not provided in the article but it was possible to infer (e.g., referring to a species being identifiable from a picture or a report of a “widow spider” identifiable to species based on geography), we reported this identification in the database. To achieve standardization throughout the database, we converted all names to the closest valid scientific name, based on the most updated spider taxonomy³¹. We next recorded the h) type of event, broken down into “encounter”, “bite”, or “deadly bite”; i) year of the event; and j) location of the event (latitude and longitude in decimal degrees). We used Google Maps to obtain WGS84 coordinates of the approximate centre of the most precise geographic region named in the article (i.e., the country, province/state, or city/town where the encounter occurred). Finally, we recorded the k) presence/absence of any photograph of the spider [whether or not of the actual spider(s) involved in the encounter]; l) presence/absence of photographs of the bite (regardless of whether the bite is being reported); and m) presence/absence of an expert-opinion, broken down into the categories “arachnologist” (e.g., spiders experts, taxonomists, entomologists), “medical professional” (e.g., doctors, veterinarians), and “other expert” (e.g., pest controllers) (see ref. ³² for a discussion).

Since several news articles often covered the same event, we created an identifier for each unique event (ID_Event), by combining location, country, and year of the event (e.g., “London_UK_2018”).

We assessed the quality of each news article by recording the presence/absence of any of four types of errors in the text and figures:

- i) errors in photographs/figures, when the photograph(s) of the species in the news article (if any) did not correspond to the species mentioned in the text, or when the attribution was not possible (e.g., blurry photographs);
- ii) errors in systematics and taxonomy, like the common mistake of considering spiders insects³³ or inaccuracies in terms of species names and in higher Linnaean taxonomic ranks (e.g., referring to tarantulas as a single species or the genus *Latrodectus* as a family);
- iii) errors in the description of venom toxicity, symptoms of envenomation, and other physiological or medical aspects or terminology (e.g., stating only female black widows can be venomous or describing the venom of recluse spiders, which causes tissue necrosis, as “neurotoxic”; see Appendix S1 for more details); and
- iv) errors in morphology and anatomy, such as the frequent “spider sting” instead of “spider bite”³⁰, or errors in describing the number of legs or eyes.

Each error type was scored as present or absent, so we did not count cumulative errors of the same type in the same news article.

Finally, we evaluated the title, subheadings, main text, and photographs/video content of each news article and assessed it as overstated (sensationalistic) or not (neutral). Sensationalism in animal-related news articles is often associated with emotional words, expressions, and images^{2,25,26}. Examples of titles of sensationalistic

Variable	Description
ID	A unique identifier for each media report. Note that the ID can be repeated when a news item includes multiple species or events.
URL	The link to the online media report. Because this was collected at the time of data mining, some URLs may not be working anymore.
Language	The language in which the media report is written.
Country_search	The country where the newspaper is published / where the search was conducted.
Newspaper	The newspaper in which the media report is published.
Type_of_newspaper	A generic description of the type of Newspaper. Levels: “Traditional newspaper” (Official newspaper in expert’s country, with both a printed and an online version), “Online newspaper” (online-only newspapers), or “Magazine” (for magazine, tabloids, etc.).
Circulation	The circulation of the Newspaper. Levels: “Regional”, “National”, “International”.
d m y	day, month, and year of publication of the media report.
Title	Article title (in the original language).
ID_Event	A unique ID for the human-spider encounter described in the media report, constructed by combining the Country_event, Location_event, and Year_event. An ID_Event can be repeated through the database when the same event was taken up by multiple newspapers.
Year_event	The year in which the ID_Event took place.
Location_event	The location (name of city/town/region) in which the ID_Event took place.
Country_event	The country in which the ID_Event described in the media report took place.
Continent	Continent in which the ID_Event described in the media report took place.
lon lat	Coordinates (longitude, latitude) of the Location_event in decimal degrees (WGS84 reference system) (e.g., 7.47; 44.72). These were derived with Google Maps / Google Earth.
lon2 lat2	Coordinates (longitude, latitude) of the Country_event in decimal degrees (WGS84 reference system) (e.g., 7.47; 44.72). These were derived with Google Maps / Google Earth.
lon3 lat3	Coordinates (longitude, latitude) of the Country_search in decimal degrees (WGS84 reference system) (e.g., 7.47; 44.72). These were derived with Google Maps / Google Earth.
Species	The scientific name of the spider species involved in the ID_Event, as reported in the news item. If the species is not mentioned and/or impossible to infer from the text and figures, the notation “Gen sp” is used.
Genus	The genus of the spider involved in the ID_Event.
Family	The family of the spider involved in the ID_Event.
Order	The order “Araneae” (to which all spiders belong) is used unless the media report incorrectly assigned other organisms as spiders (e.g., harvestmen, camel spiders, insects).
Bite	Does the human-spider encounter result in a bite? 1 = yes; 0 = no.
Death	Does the human-spider encounter result in a deadly bite? 1 = yes; 0 = no.
Figure_species	Does the media report contain a photograph (or video content) of a spider species? 1 = yes; 0 = no.
Figure_bite	Does the media report contain a photograph (or video content) of a spider bite? 1 = yes; 0 = no.
Expert_arachnologist	Was an expert consulted/capable of identifying the spider involved in the ID_Event (arachnologist, entomologist, taxonomist, etc.)? 1 = yes; 0 = no.
Expert_doctor	Was a medical doctor or other similar medical professional consulted in the media report? 1 = yes; 0 = no.
Expert_others	Was any other ‘expert’ consulted in the news (e.g., a pest controller)? 1 = yes; 0 = no.
Sensationalism	Is the media report sensationalistic/overstated? 1 = yes; 0 = no.
Taxonomic_error	Does the article contain any taxonomic error? 1 = yes; 0 = no.
Venom_error	Does the article contain any error related to spider venom? 1 = yes; 0 = no.
Anatomy_error	Does the article contain any error related to the anatomy of spiders? 1 = yes; 0 = no.
Photo_error	Does the article contain any error in the photographs (or video content)? 1 = yes; 0 = no. Note that we used ‘NA’ if there was no photo present.
Quality_check	Was the article re-assessed (see section “Data accuracy and curation”)? Levels: “yes”, “yes*” (when a new entry was added as a result of the re-assessment), and “no”.
Contributor	The researcher(s) who collected the data associated with the specific media report.
Notes	Any other information related to the media report.

Table 1. Description of each column in the database. In the database, the R notation ‘NA’ is used for missing values. See main text for more information.

versus non-sensationalistic news articles focusing on the same event are: “Thousands of spiders ‘bleed out of the walls’ and force family from home” vs. “Home Infested With Brown Recluse Spiders in Missouri”. Throughout the database, frequent words associated with sensationalistic content were ‘alarm’, ‘agony’, ‘attack’, ‘boom’, ‘deadly’, ‘creepy crawly’, ‘devil’, ‘fear’, ‘hell’, ‘killer’, ‘murderer’, ‘nasty’, ‘nightmare’, ‘panic’, ‘terrible’, ‘terrifying’, and ‘terror’², as well as magnifying adjectives that exaggerated any features of the encounter (e.g., body size^{34–36}, hairiness³⁵). However, the presence of one of these words did not necessarily result in an article being scored as sensationalistic. For example, articles that referred to spider species whose venom can be fatal without medical intervention (e.g., *Latrodectus* and *Atrax* spp.) as ‘deadly’ could be overall non-sensationalistic, whereas articles describing non-medically important spiders as ‘deadly’ were more likely to be scored as sensational.

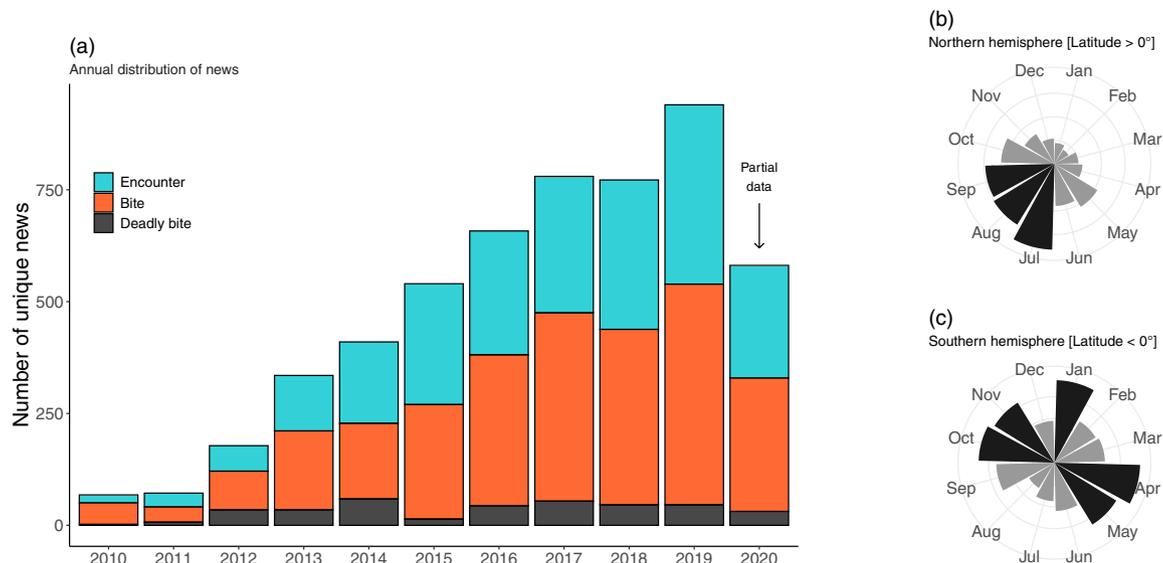


Fig. 2 Temporal distribution of unique news articles. **(a)** Annual distribution of news articles by type of event (partial data for 2020). **(b)** Monthly distribution of news articles (cumulative of all years) in the northern hemisphere. **(c)** Monthly distribution of news articles (cumulative of all years) in the southern hemisphere—darker colours highlight months with the highest numbers of news articles.

Language (Country)	Expert in charge	Details
Basque (France, Spain)	Jagoba Malumbres-Olarte	No news article was found, both using Google or by searching directly the websites of the few newspapers in Basque.
English and Setswana (Botswana)	Tharina L. Bird; Naledi T. Gonye	No news article was found, both using Google or by searching directly the websites of the local newspapers (n = 9). The lack of news was further confirmed by phone (see main text).
Galician (Spain)	Alejandro Martínez	No news article was found with direct search in Google and Google News.
Icelandic (Iceland)	Ingi Agnarsson (see section “Acknowledgements”)	No news article was found with direct search in Google and Google News. This lack of news is corroborated by personal communication with the most active entomologist in Iceland.
Montenegrin, Serbian, and Croatian (Montenegro)	Marija Miličić	No news article was found for Montenegro with direct search in Google and Google News.

Table 2. Countries and languages for which no spider-related news articles were found.

Data Records

Database availability. The database is freely available in Figshare³⁷. The database is provided both as a tab-delimited file (.csv) and as an excel file (.xlsx). Description of columns is in Table 1 but also in the metadata file uploaded alongside the database in Figshare. Code to access the database in R environment³⁸ and derive basic summary statistics and graphs shown in this paper is available in GitHub (see section “Code Availability”).

Description of the database. In total, we collected 5348 unique news articles from 81 countries in 40 languages. The database has an uneven temporal coverage, with most news articles concentrating in recent years (Fig. 2a). There is also a seasonal pattern in the distribution of news articles. In the northern hemisphere, most news articles occur throughout the summer season (Fig. 2b), whereas the pattern is less clear in the south (Fig. 2c). The number of news items by countries varies by at least three orders of magnitude, from hundreds (United Kingdom: 865; United States: 537; Italy: 412; Russia: 395; France: 319) to a handful of news articles, or none (Table 2).

The database includes 6204 reports of human-spider encounters (corresponding to 2644 unique events) and 211 identified and unidentified spider species—note that a single news article may report about multiple human-spider encounters. Of these unique events, 1121 were reported by the news articles as bites, and 147 as deadly bites (Fig. 3a). The majority of reported encounters is concentrated at northern latitudes in the northern hemisphere (median latitude = 46.9°), whereas the median latitude of reported bites and deadly bites is further south (41.3° and 26.1°, respectively) (Fig. 3b).

The presence of comments from experts consulted for articles about human-spider encounters varies substantially across countries and continents (Fig. 4a). Spider experts were only rarely consulted (Fig. 4b). One or more error types are present in 47% of news articles (Fig. 4c), although the frequency of different types of

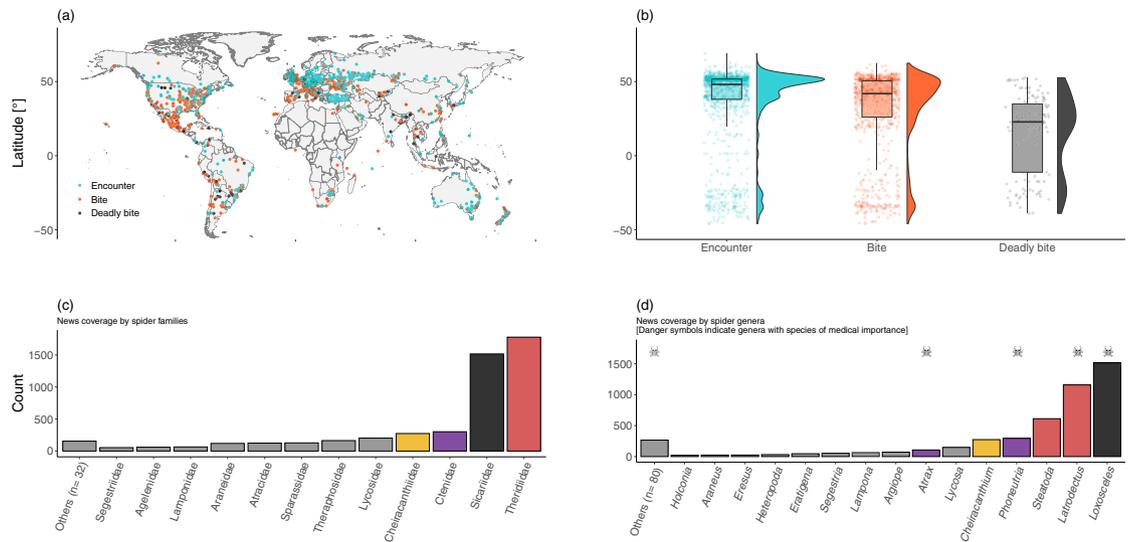


Fig. 3 Geographical coverage of the human-spider encounters in the database. **(a)** Global distribution of event localities reported in the media report; due to the proximity of several localities, most points appear superimposed. **(b)** Latitudinal distribution of events. **(c)** News coverage by spider families. **(d)** News coverage by spider genera. Danger symbol marks genera with species of medical importance. In c–d, for the four most abundant families, colours represent families.

errors is variable (data not shown). Also, 43% of news articles were assessed by experts as sensationalistic; the frequency of sensationalistic versus non-sensationalistic news varied substantially by continent (Fig. 4d).

Even zeros matter. For a few countries, we found no online news articles reporting human-spider encounters (Table 2). An informative case is Botswana, where the authors in charge performed an in-depth investigation to explain the lack of relevant results. Since Google News does not work for Botswana, they carried out the initial search in Google, which yielded no relevant results using any combination of keywords. To exclude the possibility this result was an artefact due to the searching tool, they repeated the search directly on the websites of the nine Botswana newspapers; once again, this search yielded no positive results. Finally, they phoned each of these nine newspapers individually. Six newspapers (Sunday Standard, The Midweek Sun, The Patriot, The Weekend Post, Mmegi, and Daily News) explained that their online presence is very recent and that the content placed online remains very selective. Thus, there were no human-spider encounters reported online. Conversely, they could not reach two newspapers by phone (Botswana Gazette and Botswana Guardian) and a third (The Voice Newspaper) declined to provide any information. This suggests that the search strategy is reliable in detecting absence of news articles, and these absences can be considered in analyses as ‘true zeros’³⁹.

Technical validation

Data accuracy and curation. To increase the accuracy and internal consistency of our database, and given subjectivity in the assignment of certain values (e.g., sensationalism, errors), we re-assessed news for most articles in English ($N = 1719$; 80%) and some of the other most common languages based on the availability of native speakers (French, 53%; Italian, 88%; Spanish, 48%). The column “Quality_check” in the database indicates whether this re-assessment was performed for a given article. We assigned a pair of authors to a subset of these language-based datasets so they could independently re-examine and score articles that were previously mined by the original contributor. The re-assessors scored the articles in the same manner as described previously (see section “Extraction of information for each media report”) and compared their individual datasets with each other. Discrepancies in scores were discussed and compared with the original dataset to reach a consensus on final scores.

We estimated the rate of agreement between two re-assessors via Cohen’s kappa statistic⁴⁰, calculated for all the scores of variables that may imply a degree of subjectivity in the assessment. We derived the confidence intervals using variance estimate⁴¹. We carried out this analysis only on English news (Table 3).

Limitations in using the database. Users of the database must be aware of the following limitations:

- i) The data collected here refer to online journalism only. The database does not cover the representation of spiders in the printed versions of traditional newspapers;
- ii) Because Google’s search algorithm varies by country and user, the relationship between the number of published news articles and the number of results returned is likely not consistent. For example, the total number of news articles in China is unreliable due to the restrictions imposed by the government on

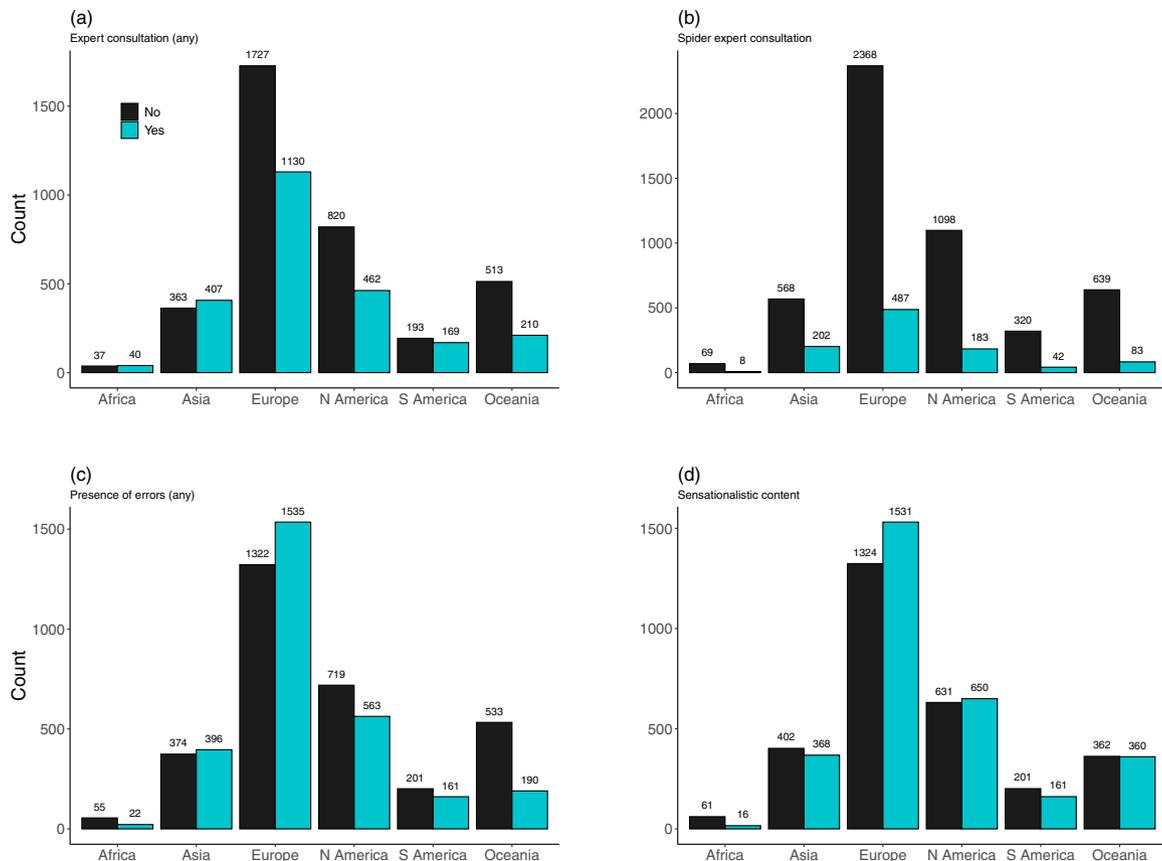


Fig. 4 Content of news articles by continent. **(a)** Frequency of expert consultation in news articles (any type of expert). **(b)** Frequency of spider expert consultation (arachnologists, entomologists and similar) in news articles. **(c)** Frequency of errors in news articles (any type of error). **(d)** Frequency of sensationalistic versus non-sensationalistic news articles.

Google. Consequently, we recommend against comparing absolute numbers of news items across countries, but rather always using relative numbers (e.g., proportion of errors, proportion of spider bites versus encounters, relative frequency of a given species in the press); and

- iii) Likewise, any temporal trend must be interpreted with caution because news publishers can occasionally remove pieces of news from the Google News index or simply delete old news. The probability of this happening increases with time and thus may be partly responsible for the apparent increase in the volume of news articles over time in this database (Fig. 2a).

Usage Notes

This database allows users to investigate questions related to the social dimension of spiders and the psychology of arachnophobia, but also contemporary problems in ecology^{42,43} and conservation biology⁴⁴. To stimulate the use of the database, we discuss what we believe are some important avenues of research—while being aware that many other questions and patterns await to be explored⁴⁵. Note that some of these research questions have already been briefly introduced in Mammola *et al.* (ref. 2), but here are expanded on and contextualized within the framework of a global-scale database.

Comparison among countries and through time. Some of the first questions that come to mind are about the reasons underlying the disparity in the quality of news across regions and countries², namely what are the main ecological, cultural, and/or social factors that explain the observed patterns? This dataset could be used to test the hypothesis that socio-economic factors, demographic features, level of education or literacy and/or cultural values affect the quality and taxonomic bias of spider-related news in a given country. For example, the relative number of reports of (presumed) bites in relation to encounters and quality of news articles may be greater in those countries/regions with either a high number of medically important species that can seriously harm humans (e.g., South America, Australia) or with a high species diversity (e.g., Brazil). Alternatively, an opposite pattern could emerge due to the paradoxically high prevalence of arachnophobia in areas with few or no dangerous spider species (e.g., the UK⁴⁶). All these questions can be directly answered, among other ways, by summarizing relative values by country (e.g., proportion of errors, proportion of spider bites versus encounters) and by relating these variables with country-level indicators⁴⁷.

Variable	Cohen's kappa (Confidence interval)	Possible reason for the discrepancy
Bite	0.99 (0.98–1.00)	In a few cases, it was not clear from the article description whether the biting event occurred or not.
Deadly_bite	0.94 (0.90–0.97)	In a few cases, it was not clear from the article description whether a fatality was attributable to the spider bite.
Figure_species	0.96 (0.94–0.97)	A photo may be overlooked for some articles filled with Ads or in presence of anti-spam filters. Also, some of the raters did not score the presence of photos in video link.
Figure_bite	0.98 (0.96–0.99)	
Expert_arachnologist	0.94 (0.91–0.96)	The assessment of all these variables implies a certain degree of subjectivity. See Appendix S1 for more details.
Expert_doctor	0.91 (0.87–0.94)	
Expert_others	0.87 (0.83–0.90)	
Sensationalism	0.89 (0.87–0.91)	
Taxonomic_error	0.88 (0.84–0.92)	
Venom_error	0.90 (0.87–0.92)	
Anatomy_error	0.87 (0.80–0.94)	
Photo_error	0.85 (0.80–0.90)	

Table 3. Cohen's kappa coefficients⁴⁰ and confidence intervals⁴¹ for independent scoring of the same news articles. Cohen's kappa statistic ranges from -1 to 1 ; values above 0.8 indicate very high to near perfect agreements among scorers. We performed this analysis only for rescored English news ($N = 1719$).

Comparison with other animal groups. The protocol for data mining discussed here is effective and inexpensive, and thus can be adapted to other cases. This would allow for comparisons of the types of errors and the levels of sensationalism across multiple taxa, including other venomous animals (e.g., bees, scorpions, wasps, snakes, and jellyfish) and answering questions such as:

- i) Is there a relationship between taxon-specific features (distribution, diversity, adaptations, dangerousness, interactions with humans) and the content and quality of the media articles referring to them?
- ii) Does a negative representation by the traditional media translate to a lower prioritization and fewer measures for conservation⁴⁸? Conversely, does such sensationalism heighten public interest in nature, biodiversity monitoring, and invasive species?

Link between traditional and social media. Social media have changed the way news information is framed and circulated⁴⁹, including biodiversity-related content. In a recent study set in Italy, we found that the volume of newspaper articles shared on Facebook has increased substantially in recent years and that sensationalistic and overstated news stories about human-spider encounters are more likely to be shared². Using the URL associated with each news article, one could directly analyse the shares on different social media platforms to answer questions about the factors that drive the popularity of news online^{22,50}. To better understand the opinions, sentiments, and subjectivity of people sharing these news reports, one could even use text mining algorithms to perform quantitative analyses (e.g., sentiment analysis) on the public comments posted in response to the news (see next section).

Linguistic analysis. The title of each news article and other bodies of texts that can be automatically extracted using the URL of each news item (e.g., comments on the online article, comments in response to the shares of the article on social media) offer a large source of data in the domain of Sentiment Analysis or Opinion Mining—defined as the computational study of people's emotions and attitudes toward a given topic⁵¹. A very simple example is provided in Fig. 5, where we used R text mining tools⁵² to compare the usage of words in sensationalistic versus non-sensationalistic titles of articles written in English. Inevitably, these kinds of analyses are more easily performed on news published in English because of the larger sample size.

Source of iEcological knowledge. This database has the potential to generate ecological insights. Similar applications broadly fall within the domain of iEcology, an emerging research field that “[...] seeks to quantify patterns and processes in the natural world using data accumulated in digital sources collected for other purposes”⁴². The database of spider news provides data on > 200 spider species (Fig. 3d); for most of these species, we have recorded the coordinates of the locality where the human-spider encounter supposedly took place. Some of these records are clearly wrong (i.e., the species reported is not correct), but the presence of the supposedly encountered species could be verified by experts. In other words, the news reports could be used by experts (spider experts, in this case) to inspect areas where a given species may be present. Once a reliable database for a given species is cleaned, one could explore different ecological patterns.

In Italy, for example, the seasonal distribution of news articles on the Mediterranean black widow [*Latrodectus tredecimguttatus* (Rossi, 1790)] overlaps almost perfectly with the known phenology of the species². Similar phenological insights may be checked for other species. Also, recent niche modelling studies have shown that internet-derived distribution data can be useful for mapping the predicted distribution of spiders (reviewed in ref. ⁵³), especially species that are easily identified in the field or by photos^{54–57}. For some of the most abundant species in our database (e.g., *Loxosceles* spp. and *Latrodectus* spp.), it is possible to compare whether

15. Hauke, T. J. & Herzig, V. Love bites – Do venomous arachnids make safe pets? *Toxicon* **190**, 65–72 (2021).
16. Chippaux, J.-P. Epidemiology of envenomations by terrestrial venomous animals in Brazil based on case reporting: from obvious facts to contingencies. *J. Venom. Anim. Toxins Incl. Trop. Dis.* **21**, 13 (2015).
17. Davey, G. C. L. The ‘disgusting’ spider: The role of disease and illness in the perpetuation of fear of spiders. *Soc. Anim.* **2**, 17–25 (1994).
18. Davey, G. C. L. *et al.* A cross-cultural study of animal fears. *Behav. Res. Ther.* **36**, 735–750 (1998).
19. Merckelbach, H., Muris, P. & Schouten, E. Pathways to fear in spider phobic children. *Behav. Res. Ther.* **34**, 935–938 (1996).
20. Hoffman, Y. S. G., Pitscho-Prelowitz, S., Ring, L. & Ben-Ezra, M. “Spidey Can”: Preliminary evidence showing arachnophobia symptom reduction due to superhero movie exposure. *Front. Psychiatry* **10**, 354 (2019).
21. Fukano, Y. & Soga, M. Why do so many modern people hate insects? The urbanization–disgust hypothesis. *Sci. Total Environ.* **777**, 146229 (2021).
22. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
23. López-Baucells, A., Rocha, R. & Fernández-Llamazares, Á. When bats go viral: negative framings in virological research imperil bat conservation. *Mamm. Rev.* **48**, 62–66 (2018).
24. MacFarlane, D. & Rocha, R. Guidelines for communicating about bats to prevent persecution in the time of COVID-19. *Biol. Conserv.* **248**, 108650 (2020).
25. Nanni, V. *et al.* Social media and large carnivores: Sharing biased news on attacks on humans. *Front. Ecol. Evol.* **8**, 71 (2020).
26. Bombieri, G. *et al.* Content analysis of media reports on predator attacks on humans: Toward an understanding of human risk perception and predator acceptance. *Bioscience* **68**, 577–584 (2018).
27. Gerber, D. L. J., Burton-Jeangros, C. & Dubied, A. Animals in the media: New boundaries of risk? *Health Risk Soc.* **13**, 17–30 (2011).
28. Cushing, N. & Markwell, K. ‘Watch out for these KILLERS!’: Newspaper coverage of the Sydney funnel web spider and its impact on antivenom research. *Health History* **12**, 79–96 (2010).
29. Ju, A., Jeong, S. H. & Chyi, H. I. Will social media save newspapers?: Examining the effectiveness of Facebook and Twitter as news platforms. *Journal. Pract.* **8**, 1–17 (2014).
30. Afshari, R. Bite like a spider, sting like a scorpion. *Nature* **537**, 167 (2016).
31. World Spider Catalog. Version 23.0. *Natural History Museum Bern* 10.24436/2 (2022).
32. Vetter, R. S. Arachnids misidentified as brown recluse spiders by medical personnel and other authorities in North America. *Toxicon* **54**, 545–547 (2009).
33. Jambriña, C. U., Vacas, J. M. & Sánchez-Barbudo, M. Preservice teachers’ conceptions about animals and particularly about spiders. *Electron. J. Res. Educ. Psychol.* **8**, 787–814 (2010).
34. Leibovich, T., Cohen, N. & Henik, A. Itsy bitsy spider?: Valence and self-relevance predict size estimation. *Biol. Psychol.* **121**, 138–145 (2016).
35. Zvaríková, M. *et al.* What makes spiders frightening and disgusting to people? *Frontiers in Ecology and Evolution* **9**, 424 (2021).
36. Frynta, D. *et al.* Emotions triggered by live arthropods shed light on spider phobia. *Sci. Rep.* **11**, 22268 (2021).
37. Mammola, S. *et al.* Global Spider News. *Database. Figshare* <https://doi.org/10.6084/m9.figshare.14822301> (2021).
38. R Core Team. R: A Language and Environment for Statistical Computing. (2021).
39. Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M. & Castells, E. What does a zero mean? Understanding false, random and structural zeros in ecology. *Methods Ecol. Evol.* **10**, 949–959 (2019).
40. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).
41. Fleiss, J. L., Cohen, J. & Everitt, B. S. Large sample standard errors of kappa and weighted kappa. *Psychol. Bull.* **72**, 323–327 (1969).
42. Jarić, I. *et al.* iEcology: Harnessing large online resources to generate ecological insights. *Trends Ecol. Evol.* **35**, 630–639 (2020).
43. Jarić, I. *et al.* Expanding conservation culturomics and iEcology from terrestrial to aquatic realms. *PLoS Biol.* **18**, e3000935 (2020).
44. Ladle, R. J. *et al.* Conservation culturomics. *Front. Ecol. Environ.* **14**, 269–275 (2016).
45. Yanai, I. & Lercher, M. A hypothesis is a liability. *Genome Biol.* **21**, 231 (2020).
46. Cavell, M. Arachnophobia and early english literature. In *New Medieval Literature* (eds Ashe, L., Knox, P., Lawton, D. & Scase, W.) 1–44 (D.S. Brewer, 2018).
47. Mammola, S. *et al.* The global spread of (mis)information on spiders. *Research Square*, <https://doi.org/10.21203/rs.3.rs-1383492/v1>.
48. Knight, A. J. ‘Bats, snakes and spiders, Oh my!’ How aesthetic and negativistic attitudes, and other concepts predict support for species protection. *J. Environ. Psychol.* **28**, 94–103 (2008).
49. Jones-Jang, S. M., Hart, P. S., Feldman, L. & Moon, W.-K. Diversifying or reinforcing science communication? Examining the flow of frame contagion across media platforms. *Journal. Mass Commun. Q.* **97**, 98–117 (2019).
50. Palazzi, M. J. *et al.* Resilience and elasticity of co-evolving information ecosystems. *arXiv* 2005.07005 (2020).
51. Medhat, W., Hassan, A. & Korashy, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng. J.* **5**, 1093–1113 (2014).
52. Silge, J. & Robinson, D. tidytext: Text mining and analysis using tidy data principles in R. *J. Open Source Softw.* **1**, 37 (2016).
53. Mammola, S. *et al.* Challenges and opportunities of species distribution modelling of terrestrial arthropod predators. *Divers. Distrib.* **27**, 2596–2614 (2021).
54. Wang, Y., Casajus, N., Buddle, C., Berteaux, D. & Larrivee, M. Predicting the distribution of poorly-documented species, Northern black widow (*Latrodectus variolus*) and Black purse-web spider (*Sphodros niger*), using museum specimens and citizen science data. *PLoS One* **13** (2018).
55. Jimenez-Valverde, A., Pena-Aguilera, P., Barve, V. & Burguillo-Madrid, L. Photo-sharing platforms key for characterising niche and distribution in poorly studied taxa. *Insect Conserv. Divers.* **12**, 389–403 (2019).
56. Hart, A. G., Nesbit, R. & Goodenough, A. E. Spatiotemporal variation in house spider phenology at a national scale using citizen science. *Arachnology* **17**, 331–334 (2018).
57. Sadir, M. & Marske, K. A. Urban environments aid invasion of Brown Widows (Theridiidae: *Latrodectus geometricus*) in North America, constraining regions of overlap and mitigating potential impact on native widows. *Front. Ecol. Evol.* **9**, 757902 (2021).

Acknowledgements

We are grateful to Dr. Ingi Agnarsson for checking the presence of spider-related news in Icelandic newspapers. Thanks to Dr. Aino K. Juslen and the Finnish Museum of Natural History (LUOMUS; University of Helsinki, Helsinki, Finland) for financially supporting the publication of this work.

Author contributions

Conceptualization: SM, JM-O, CS, AC; Data collection & validation: all authors; Data management: SM, VN, AC; Data analysis & visualization (Figs. 2–5): SM; Summary illustration (Fig. 1): JM-O; Writing (first draft): SM; Writing, contributions: JM-O, CS, AC; All authors read the text, provided comments, suggestions, and corrections, and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01197-6>.

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

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

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Stefano Mammola ^{1,2}✉, Jagoba Malumbres-Olarte ^{1,3,61}, Valeria Arabesky^{4,5}, Diego Alejandro Barrales-Alcalá ⁶, Aimee Lynn Barrion-Dupo ⁷, Marco Antonio Benamú ^{8,9,10}, Tharina L. Bird ^{11,12}, Maria Bogomolova¹³, Pedro Cardoso ¹, Maria Chatzaki ¹⁴, Ren-Chung Cheng ¹⁵, Tien-Ai Chu ¹⁵, Leticia M. Classen-Rodríguez ¹⁶, Iva Čupić¹⁷, Naufal Urfi Dhiya'ulhaq ¹⁸, André-Philippe Drapeau Picard ¹⁹, Hisham K. El-Hennawy²⁰, Mert Elverici ²¹, Caroline S. Fukushima ¹, Zeana Ganem^{22,23}, Efrat Gavish-Regev ²², Naledi T. Gonnye²⁴, Axel Hacala ²⁵, Charles R. Haddad ²⁶, Thomas Hesselberg ²⁷, Tammy Ai Tian Ho ²⁸, Thanakorn Into²⁹, Marco Isaia³⁰, Dharmaraj Jayaraman ³¹, Nanguel Karuaera³², Rajashree Khalap³³, Kiran Khalap³³, Dongyoung Kim ³⁴, Tuuli Korhonen¹, Simona Kralj-Fišer ³⁵, Heidi Land³⁶, Shou-Wang Lin ³⁶, Sarah Loboda ³⁷, Elizabeth Lowe ³⁸, Yael Lubin ^{5,39}, Alejandro Martínez ², Zingisile Mbo²⁶, Marija Miličić ^{1,40}, Grace Mwendé Kioko⁴¹, Veronica Nanni ^{30,42}, Yusoff Norma-Rashid ⁴³, Daniel Nwankwo⁴⁴, Christina J. Painting ⁴⁵, Aleck Pang⁴⁶, Paolo Pantini⁴⁷, Martina Pavlek^{17,48}, Richard Pearce⁴⁹, Booppa Petcharad²⁹, Julien Pétilion ^{25,50}, Onjasherizo Christian Raberahona⁵¹, Joni A. Saarinen¹, Laura Segura-Hernández ⁵², Lenka Sentenská ⁵³, Gabriele Uhl ³⁶, Leilani Walker^{54,55}, Charles M. Warui⁵⁶, Konrad Wiśniewski ⁵⁷, Alireza Zamani⁵⁸, Catherine Scott ^{37,61} & Angela Chuang ^{59,60,61}

¹Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History (LUOMUS), University of Helsinki, Helsinki, Finland. ²Molecular Ecology Group (MEG), Water Research Institute, National Research Council of Italy (CNR-IRSA), Largo Tonolli 50, 28922, Verbania Pallanza, Italy. ³CE3C – Centre for Ecology, Evolution and Environmental Changes / Azorean Biodiversity Group and Universidade dos Açores, Angra do Heroísmo, Azores, Portugal. ⁴Albert Katz International School for Desert Studies, Ben-Gurion University of the Negev, Sede Boqer Campus, Beersheba, Israel. ⁵Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boqer Campus, Beersheba, Israel. ⁶Colección Nacional de Arácnidos, Instituto de Biología, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico. ⁷Environmental Biology Division, Institute of Biological Sciences, College of Arts and Sciences and Museum of Natural History, University of the Philippines Los Baños, 4031, Los Baños, Philippines. ⁸Centro Universitario de Rivera, Universidad de la República, Montevideo, Uruguay. ⁹Lab. Ecotoxicología de Artrópodos Terrestres, Centro Universitario de Rivera, Universidad de la República, Montevideo, Uruguay. ¹⁰Laboratorio Ecología del Comportamiento, Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), Montevideo, Uruguay. ¹¹Ditsong National Museum of Natural History, PO Box 4197, Pretoria, 0001, South Africa. ¹²Department of Zoology and Entomology, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa. ¹³Freelance translator, Verbania Pallanza, Italy. ¹⁴Department of Molecular Biology and Genetics, Democritus University of Thrace, Komotini, Greece. ¹⁵Department of Life sciences, National Chung Hsing University, No.145 Xingda Rd., South Dist., Taichung City, 402204, Taiwan. ¹⁶Department of Biology, Macelwane Hall, 3507 Laclede Avenue, Saint Louis University, St. Louis, MO, 63103, USA. ¹⁷Croatian Biospeleological Society, Rooseveltov trg 6, Zagreb, Croatia. ¹⁸Program Sarjana, Fakultas Biologi, Universitas Gadjah Mada, Yogyakarta, Indonesia. ¹⁹Insectarium

de Montréal, Espace pour la vie, 4101, rue Sherbrooke Est, Montréal, Québec, H1X 2B2, Canada. ²⁰Serket, Arachnid Collection of Egypt (ACE), Cairo, Egypt. ²¹Erzincan Binali Yıldırım University, Faculty of Science and Arts, Biology Department, 24002, Erzincan, Turkey. ²²The National Natural History Collections, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, 9190401, Israel. ²³The Department of Ecology, Evolution and Behavior, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, 9190401, Israel. ²⁴Botswana International University of Science and Technology, Palapye, Botswana. ²⁵UMR CNRS 6553 Ecobio, Université de Rennes, 263 Avenue du Gal Leclerc, CS 74205, 35042, Rennes Cedex, France. ²⁶Department of Zoology and Entomology, University of the Free State, P.O. Box 339, Bloemfontein, 9300, South Africa. ²⁷Department of Zoology, University of Oxford, Oxford, OX1 3PS, United Kingdom. ²⁸Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore, 117543, Singapore. ²⁹Department of Biotechnology, Faculty of Science and Technology, Thammasat University, Rangsit, Pathum Thani, 12121, Thailand. ³⁰Dept. of Life Science and Systems Biology, University of Torino, Via Accademia Albertina, 13 - 10123, Torino, Italy. ³¹Unit of Conservation Biology, Department of Zoology, Bharathiar University, Coimbatore, 641046, Tamilnadu, India. ³²National Museum of Namibia, Windhoek, Namibia. ³³5A Sagar Sangeet, SBS Marg, Mumbai, 400005, India. ³⁴Department of Biological Sciences, Ajou University, Suwon, Republic of Korea. ³⁵Research Centre of the Slovenian Academy of Sciences and Arts, Jovan Hadži Institute of Biology, Ljubljana, Slovenia. ³⁶University of Greifswald, Zoological Institute and Museum, General and Systematic Zoology, Loitzerstrasse 26, 17489, Greifswald, Germany. ³⁷Department of Natural Resource Sciences, McGill University, 21 111 Lakeshore Road, Sainte-Anne-de-Bellevue, Quebec, H9X 3V9, Canada. ³⁸Department of Biological Science, Macquarie University, Sydney, NSW, 2122, Australia. ³⁹Mitrani Department of Desert Ecology, University in Midreshet Ben-Gurion, Midreshet Ben-Gurion, Israel. ⁴⁰BioSense Institute – Research Institute for Information Technologies in Biosystems, University of Novi Sad, Dr Zorana Đinđića 1, 21000, Novi Sad, Serbia. ⁴¹National Museums of Kenya, Museum Hill, P.O. BOX 40658- 00100, Nairobi, Kenya. ⁴²School for Advanced Studies IUSS, Science, Technology and Society Department, 25100, Pavia, Italy. ⁴³Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603, Kuala Lumpur, Malaysia. ⁴⁴Department of Animal and Environmental Biology, Federal University, Oye-Ekiti, Ekiti State, Nigeria. ⁴⁵Te Aka Mātua School of Science, University of Waikato, Private Bag 3105, Hamilton, 3240, New Zealand. ⁴⁶Independent researcher, Toronto, Canada. ⁴⁷Museo Civico di Scienze Naturali “E. Caffi”, Piazza Cittadella, 10, I-24129, Bergamo, Italy. ⁴⁸Ruđer Bošković Institute, Bijenička cesta 54, 10000, Zagreb, Croatia. ⁴⁹Biodiversity Research Laboratory, Moreton Morrell, Warwickshire College University Centre, Warwickshire, United Kingdom. ⁵⁰Institute for Coastal and Marine Research, Nelson Mandela University, Port Elizabeth, South Africa. ⁵¹Department of Entomology, University of Antananarivo, Antananarivo, Madagascar. ⁵²School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, Nebraska, 68588, United States. ⁵³Department of Biological Sciences, University of Toronto Scarborough, 1265 Military Trail, Scarborough, Ontario, M1C 1A4, Canada. ⁵⁴Natural Sciences, Auckland War Memorial Museum, Parnell, Auckland, 1010, New Zealand. ⁵⁵Te Pūnaha Matatini, University of Auckland, Auckland, New Zealand. ⁵⁶Murang’a University of Technology, Department of Physical & Biological Sciences, P.O.Box 75-10200, Murang’a, Kenya. ⁵⁷Institute of Biology and Earth Sciences, Pomeranian University in Słupsk, Arciszewskiego 22a, 76-200, Słupsk, Poland. ⁵⁸Zoological Museum, Biodiversity Unit, FI-20014, University of Turku, Turku, Finland. ⁵⁹Department of Psychology, University of Tennessee, Knoxville, Tennessee, USA. ⁶⁰Department of Entomology and Nematology, Citrus Research and Education Center, University of Florida, Lake Alfred, Florida, USA. ⁶¹These authors contributed equally: Jagoba Malumbres-Olarte, Catherine Scott, Angela Chuang. ✉e-mail: stefano.mammola@cnr.it