— uncorrected proof —

Zootaxa 0000 (0): 000–000
https://www.mapress.com/j/zt/

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

https://doi.org/10.11646/zootaxa.0000.0.0
http://zoobank.org/urn:lsid:zoobank.org:pub:00000000-0000-0000-0000-00000000000

An inventory of online reptile images

BENJAMIN MICHAEL MARSHALL[1*], PAUL FREED[2], LAURIE J. VITT[3], PEDRO BERNARDO[4], GERNOT VOGEL[5], SEBASTIAN LOTZKAT[6], MICHAEL FRANZEN[7], JAKOB HALLERMANN[8], RICHARD D. SAGE[9], BRIAN BUSH[10], MARCELO RIBEIRO DUARTE[11], LUCIANO JAVIER AVILA[12], DAVID JANDZIK[13], BORIS KLUSMEYER[14], BRAD MARYAN[15], JIŘÍ HOŠEK[16], PETER UETZ[17*]

[1]*Suranaree University of Technology, Nakhon Ratchasima, Thailand*
benjaminmichaelmarshall@gmail.com; https://orcid.org/0000-0001-9554-0605
[2]*Scotts Mills, Oregon 97375, USA. herps2apes@gmail.com*
[3]*Sam Noble Museum, Norman, Oklahoma 73072, USA. vitt@ou.edu*
[4]*Royal Ontario Museum, Toronto, ON, Canada. bernardoph@gmail.com*
[5]*Society for Southeast Asian Herpetology, Heidelberg, Germany*
gernot.vogel@t-online.de; https://orcid.org/0000-0002-4542-518X
[6]*Staatliches Museum für Naturkunde Stuttgart, Stuttgart, Germany*
lotzkat@yahoo.com; https://orcid.org/0000-0001-8171-9321
[7]*Zoologische Staatssammlung München (ZSM-SNSB), 81247 München, Germany. franzen@snsb.de*
[8]*Universität Hamburg, Centrum für Naturkunde, Hamburg, Germany*
hallermann@uni-hamburg.de; https://orcid.org/0000-0002-8835-9303
[9]*Museum of Vertebrate Zoology, University of California, Berkeley, California 94720, and Sociedad Naturalista Andino Patagónica (SNAP), S. C. de Bariloche, Río Negro, 8400, Argentina. rdsage1943@gmail.com; https://orcid.org/0000-0002-0897-6927*
[10]*Stoneville, WA, Australia. bush@iinet.net.au; https://orcid.org/0000-0002-3487-6620*
[11]*Instituto Butantan, Laboratório de Coleções Zoológicas, São Paulo, Brazil. marcelo.duarte@butantan.gov.br*
[12]*GHP-LASIBIBE, IPEEC-CONICET, Puerto Madryn, Chubut, Argentina*
avila@cenpat-conicet.gob.ar; https://orcid.org/0000-0003-4641-6581
[13]*Department of Zoology, Comenius University, Bratislava, Slovakia*
david.jandzik@uniba.sk; https://orcid.org/0000-0002-4239-0014
[14]*Bremen, Germany. boris.klusmeyer@gmail.com*
[15]*Department of Terrestrial Zoology, Western Australian Museum, Welshpool, Western Australia 6106, AUSTRALIA*
lerista.3@internode.on.net; https://orcid.org/0000-0003-0349-0255
[16]*Folvarek 19, 724 00 Ostrava, Czech Republic. jiri.hosek@gmail.com*
[17]*Center for Biological Data Science, Virginia Commonwealth University, Richmond, VA, USA*
peter@uetz.us; http://orcid.org/0000-0001-6194-4927
[*]*Corresponding authors*

Abstract

No central online repository exists for the collection of animal images; hence it remains unclear how extensively species have been illustrated in the published literature or online. Here we compiled a list of more than 8000 reptile species (out of 11,341) that have photos in one of six popular online repositories, namely iNaturalist (6,349 species), the Reptile Database (5,144), Flickr (4,386), CalPhotos (3,071), Wikimedia (2,952), and Herpmapper (2,571). These sites have compiled over one million reptile photos, with some species represented by tens of thousands of images. Despite the number of images, many species have only one or a few images. This suggests that a considerable fraction of morphological and geographic variation is under documented or difficult to access. We highlight prominent gaps in amphisbaenians, lizards, and snakes, with geographic hotspots for species without images in Central Africa, Pacific Islands, and the Andes Mountains. We present a list of ~3,000 species without photos in any of the six databases and ask the community to fill the gaps by depositing images on one of these sites (preferably with minimal copyright restrictions).

Key words: web 2.0, big data, community science, social media, photography, Reptilia, Sauria, Serpentes

Accepted by J. Smid: 3 Nov. 2020; published: xx xxx. 2020

Introduction

The internet has enabled a wealth of community created resources for the exploration of ecological and evolutionary questions. Recently referred to as i-ecology (Jarić *et al.* 2020), online species information has provided insights into taxonomy, macroecological patterns, and evolutionary processes. Species distributions are enriched by the inclusion of community generated data (Jiménez-Valverde *et al.* 2019), invasive species expansions are identified and tracked (Allain 2019), and species interactions are identified from social media posts (Maritz & Maritz 2020). Outside of strictly ecological questions, species online presences provide a mirror to society's interest and attitude towards species (Kidd *et al.* 2018; McClain 2019; Roll *et al.* 2016), while potentially highlighting conflicts, ranging from habitat destruction to nature tourism (Measey *et al.* 2019; Miranda *et al.* 2016, Otsuka & Yamakoshi 2020). Reptiles are also threatened by commercial overharvesting and some species possibly even by the pet trade (Auliya *et al* 2016, Marshall *et al*. 2020).

Most biological studies make use of photographs in one way or another, e.g., representing occurrence records (Jarić *et al.* 2020; Marshall & Strine 2019a), phenotypic variation (Drury *et al.* 2019), behaviour (Dylewski *et al.* 2017), or ecology (Maritz & Maritz 2020). Existing image databases are driven by community science projects (iNaturalist, HerpMapper, Wikimedia, Flickr, CalPhotos), but (semi-) professional sources exist as well (Reptile Database, Morphobank). Taxon-specific databases can be a considerable source of images, connecting species identification with images. For instance, Amphibiaweb holds 40,947 images covering over 57% of all 8,205 amphibian species (as of 2020-07-16; maintained by CalPhotos).

Community science projects such as iNaturalist or HerpMapper have made substantial contributions to mapping species ranges and to some extent their abundances. These sites show that volunteers can fill numerous gaps and thus productively complement the work of professional taxonomists.

To explore images and the questions they may answer, we must first discover and document image availability. Here we aim to document the quantity and location of reptile images available online, making use of the aforementioned community science projects complemented by searches of web 2.0 platforms (i.e., websites populated by user-generated content, namely iNaturalist, HerpMapper, Wikimedia, Flickr, CalPhotos). Our goal was to investigate how many of the ~11,300 reptile species have easily accessible photos and how they represent their taxonomic and geographic distribution.

Methods

We compiled metadata associated with reptile images from the online sources listed in **Table 1**. Most of these sources are also used by other major biodiversity platforms, such as the Encyclopedia of Life (https://eol.org/). For the following analysis we use the Reptile Database species list (accessed 2020-08-03) as a backbone taxonomy.

TABLE 1. List of all resources used in this study, number of species with photos, and their URLs.

Resource	Reptile species	URL
iNaturalist	6,349	http://inaturalist.org
Reptile Database	5,144	http://www.reptile-database.org
Flickr	4,386	https://www.flickr.com/
CalPhotos	3,071	https://calphotos.berkeley.edu/
Herpmapper	2,952	http://herpmapper.org
Wikimedia Commons	2,571	https://commons.wikimedia.org/

Summarised database results

Reptile Database. We used the 15,445 images stored in the August 2020 release of the Reptile Database, representing 5,144 species. In contrast to the other resources, the Reptile Database has photographs of preserved specimens from museum collections, although those represent only ~700 images and less than 300 species, respectively. Many of the latter are also represented by live photos. We included photos of preserved specimens as they often represent species for which no (or few) live photos are available or because they show characters not easily visible in photographs of live specimens.

At the time of this writing (2020-08-02), **CalPhotos** contained 29,619 reptile photos (tagged as such by CalPhotos), representing 3,071 species. Since CalPhotos also uses the Reptile Database taxonomy, names were mapped to the Reptile Database. However, CalPhotos also contains photos without scientific names or other media, such as images of skeletons deposited in the Museum of Vertebrate Zoology, Berkeley, among others, which we ignored. The CalPhoto summary only covered a single entry per species, so all counts of images are one.

iNaturalist. A list of 6,349 species represented in iNaturalist as "research grade" photos was kindly provided by Scott Loarie (iNaturalist). "Research grade" indicates that more than 2/3 of identifiers agree on a taxon identity.

Systematic web searches

For **Wikimedia**, HerpMapper and Flickr we made use of platform's APIs or systematically searched for species data.

We used R packages *jsonlite v.1.6.1* (Ooms 2014), *stringr v.1.4.0* (Wickham 2019b), *urltools v.1.7.3* (Keyes *et al.* 2019) to interact with the MediaWiki API (https://en.wikipedia.org/w/api.php?). Initially we queried the search endpoint with *action=query&list=search&srsearch=* using each Reptile Database species as a search term. If the search returned results, we used those results to query the API for pages alongside information held in the *image* and *title* categories (*prop=categories|images&titles=*). We discarded any page results that did not include the "*Articles with 'species' microformats*" category. For pages detected via the search and included species information, we further queried the API *&prop=imageinfo&&iiprop=timestamp|extmetadata|user|url* to extract details on the images present in the page (name, description, URL, user, artist, license and usage information).

We repeated the Wikimedia search using all Reptile Database synonyms because the page titles used by Wikimedia may not be taxonomically up to date. Once initial and synonym searches had been completed, we removed duplicate images by detected duplicated URLs.

We retrieved **HerpMapper** data (accessed: 2020-07-13–14) by using the *downloader v.0.4* package (Chang 2015) to retrieve pages listing all records for each of the reptile groups: "https://www.herpmapper.org/records?taxon=" + "Amphisbaenia", "Crocodylia", "Lacertilia", "Serpentes", "Sphenodon", "Testudines". Using the *dplyr v.0.8.4* (Wickham *et al.* 2019), *stringr v.1.4.0* (Wickham 2019b), *rvest v.0.3.5* (Wickham 2019a), and *xml2 v.1.2.2* (Wickham *et al.* 2018) packages, we extracted the number of pages the records were split across, and proceeded to cycle through the record pages retrieving details on species name, record ID, date recorded, and location.

We adapted code from previous work (Marshall & Strine 2019a; b) to query the **Flickr** API (Flickr Development Team 2019) for reptile images. The Flickr querying code made use of *jsonlite v.1.6.1* (Ooms 2014) and *RCurl v.1.95.4.12* (Lang & The CRAN Team 2018). We used binomial names listed on Reptile Database as search terms, and limited the search to images classed as *photographs*, and tagged with at least one of the following: "snake", "reptile", "tortoise", "turtle", "lizard", "crocodile", "alligator", "crocodilian", "amphisbaenian", "tuatara", "gharial", or "caiman". We initially queried the API to determine the number of results. If there were images detected by the search, we ran a second set of queries to retrieve image details. When the initial Flickr API query returned over 250 results, the results were split between multiple pages (each holding 250 results). To obtain information on all results we systematically cycled through each page, compiling image information for each set/page of 250 images. We retrieved information on the images: user, title, license, location, location accuracy, date taken, and URL. We slowed the pace of API queries by introducing five seconds delay between requests. In a few cases, special characters (usually in image titles) caused parsing errors when reading the Flickr API's json formatted results. We flagged instances of parsing error and those results are excluded from final species and image counts (27 instances of parsing caused by unrecognised characters).

Facebook (FB). FB offers a large number of groups and other opportunities to post and share reptile photos. Unfortunately, systematic access to FB via the FB API requires specific user access points. Use of user access points means searches are likely non-reproducible and impacted by user language, location, and group access. For more targeted searches that could be undertaken manually, FB undoubtably presents a useful resource but for a broad inventory, manual search of groups and posts was not feasible.

Data and software availability

We retrieved and summarised data from Wikimedia, Flickr and HerpMapper using *R v.3.5.3* (R Core Team 2020) and *R Studio v.1.2.1335* (R Studio Team 2020). We used *dplyr v.1.0.2* (Wickham *et al.* 2019) and *tidyr v.1.1.1*

(Wickham & Henry 2019) to manipulate data prior to plotting, we used *ggplot2 v.3.2.1* (Wickham 2016) to generate figures (Figs. 2, 3, 5-7) with *ggrepel v.0.8.2* (Slowikowski 2018) for labelling (Fig. 3) and *scico v.1.1.0* (Pedersen & Crameri 2018) for colourway generation (Fig. 7), we used *cowplot v.0.9.4* (Wilke 2019) and *ggpubr v.0.4.0* (Kassambara 2018) to arrange plots into final multi-panel figures (Figs. 2, 3 & 7). We also used *fasterize v.1.0.2* (Ross 2020), *sf v.0.8.2* (Pebesma 2018), and *raster v.3.0.12* (Hijmans 2020) packages to summarise reptile ranges from the Global Assessment of Reptile Distributions (GARD) database (Roll *et al.* 2017) and map reptile diversity without images (Fig. 7A and 7B). For Figure 6, we used tree data from TimeTree.org (Hedges *et al.* 2006; Kumar *et al.* 2017; Kumar & Hedges 2011), and used *ape v.5.4* to manipulate data ready for plotting with (Paradis & Schliep 2019), *ggplot2 v.3.2.1* (Wickham 2016), and *ggtree v.2.0.4* (Yu *et al.* 2017, 2018) packages. We combined tree figures into the final Figure 6 with Affinity Designer v.1.8.3.641 (Serif 2020). We created the Venn diagram using InteractiVenn (Heberle *et al.* 2015). We have provided code and data used as supplementary material and at Zenodo: https://doi.org/10.5281/zenodo.4010155. Photos from Flickr, Calphotos, and iNaturalist, and the Reptile Database are available in the species entries of the Reptile Database (if their scientific species names match).

Results

Across all data sources we find over 1,193,764 images of reptiles, representing 8,207 of 11,242 (73%) reptile species, each of which is documented in at least one image (Figs. 1, 2; Supplementary Table 1; Supplementary Table 2). Excluding species with no images, the average number of images per species was 48.78 ± 4.54, with a maximum of 49,128 images for *Sceloporus occidentalis* BAIRD & GIRARD (Fig. 3).

FIGURE 1. Photos in nature vs preserved material. (A) *Podarcis siculus* RAFINESQUE-SCHMALTZ in its natural environment. (B) Preserved *Lacerta agilis* LINNAEUS (ZSM 1648/2009), showing details, such as the ventral side, not visible in its natural environment.

The most fruitful source of reptile images was iNaturalist. iNaturalist provided the greatest number of species (Fig. 2), while also supplying the greatest mean number of images per species of any data source (132.24 ±15.96 images per species; Fig. 3). Overlap between the sources is highly complementary, with only 993 species (12.1%) being represented in all 6, but 2,245 species (27.4%) being represented in only one of them (Fig. 4). Despite the lead of iNaturalist, the Reptile Database does display photos from other sources such as CalPhotos (when permitted) and thus its number of species with photos exceeds that of iNaturalist.

Taxonomic and geographic representation

Images were not distributed equally between reptile clades, with crocodiles, turtles and tuataras being the smallest and thus best-covered groups, with over 90% of species photographed (Fig. 5). The worst covered clade was Amphisbaenia, with less than half of the species photographed. Snakes and lizards, the largest extant (non-avian) reptile groups, currently contain 3,837 and 6,876 species, respectively. Both were represented almost equally with about 72% of their species covered by photos. This is surprising given the more secretive lifestyle of snakes, which typically makes them harder to find than lizards.

Size and lifestyle also appear to determine photographic representation within families; with larger species appearing better represented than smaller species. For instance, 90.2 and 96.3% of all pythons (Pythonidae) and monitors lizards (Varanidae) have photos, while small and nocturnal species have fewer species covered by photos, e.g., 64.7 % of geckos (Gekkonidae; Fig. 6). Predominantly fossorial families, like Typhlopidae (36.4%), similarly have lower percentages of species photographed.

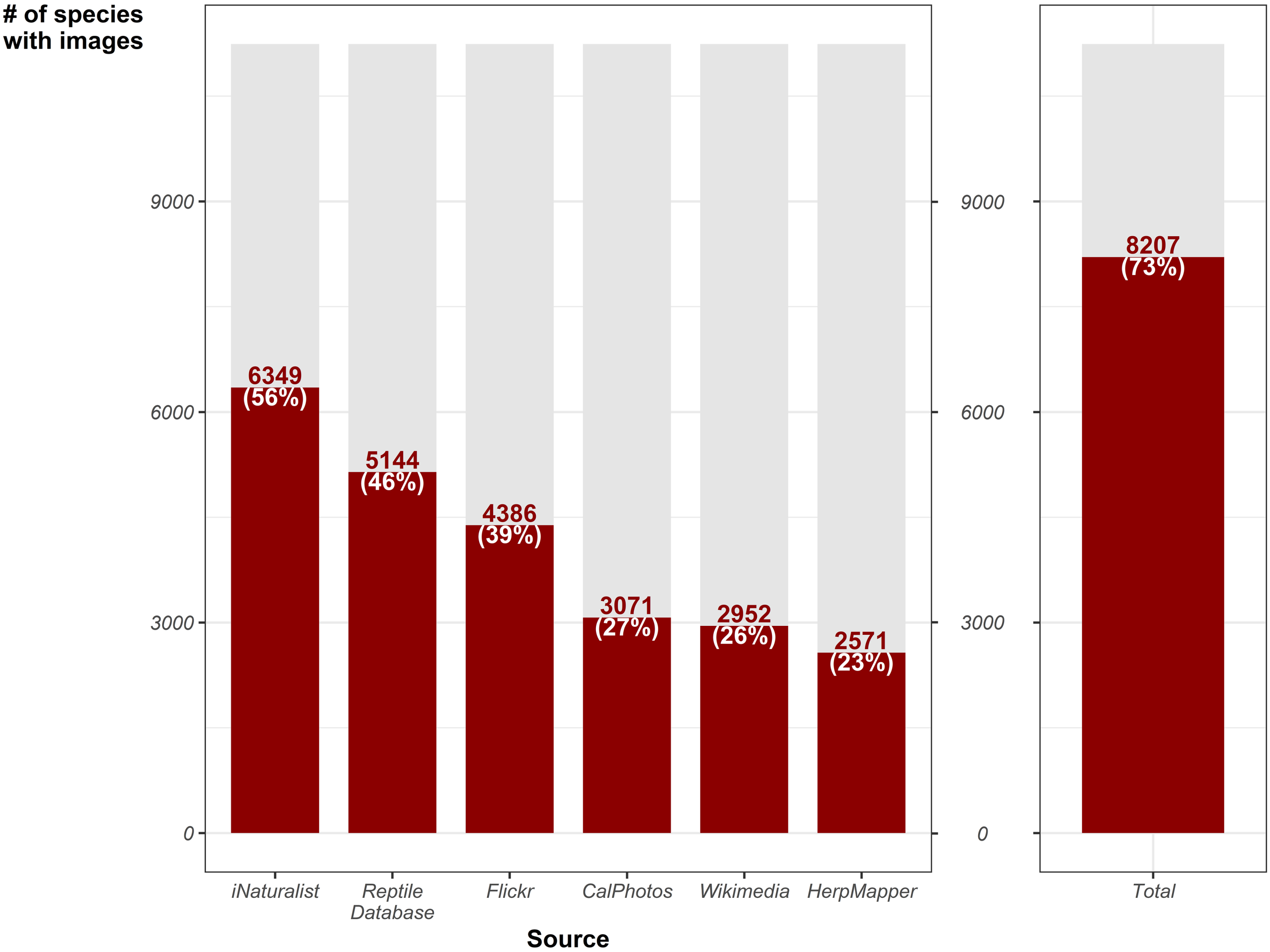

FIGURE 2. Number of species with images on each of the data sources, as well as the total number of species across all sources (right panel).

Species lacking images are concentrated in tropical areas (Fig. 7A), particularly in Central Africa, Pacific Islands, and parts of the Andes. Eurasian and North American species are almost completely covered by images, with very few exceptions (Fig. 7B). Australia stands out among the highly biodiverse countries as having the most species with photographic records –1119 species (Fig. 7C; Supplementary Table 3). The percentage of photographed species per country largely reflect cold spots highlighted by the GARD range maps, namely low percentages of species in Pacific Islands, Central Africa, but also Caribbean islands (Fig. 7D).

Discussion

By examining multiple online photo repositories, we have attempted to quantify the extent, coverage, and gaps in reptile photographic documentation. Our results provide only a snapshot; images are constantly being uploaded and community science initiatives continue to constitute a growing source for ecological and evolutionary data. We show that currently over 73% of species are covered by at least one image. Not surprisingly, considerably better coverage exists for testudines and crocodilians compared to lepidosaurians; and the largest geographic gaps exist in equatorial regions, which are also the most species-rich and often less accessible areas.

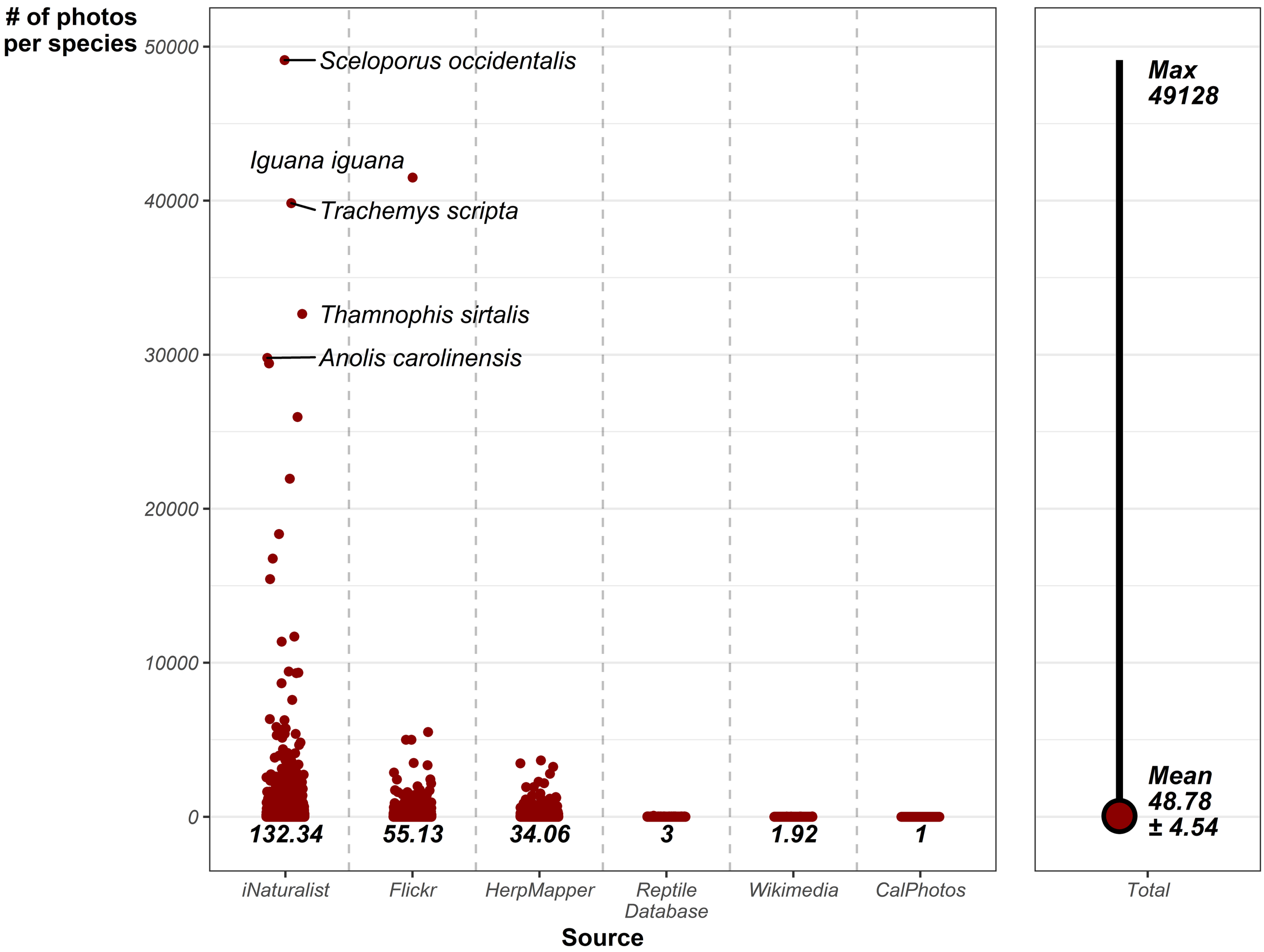

FIGURE 3. Number of images per species by source, excluding species without images. Top five species with images are highlighted with species labels.

Species images in books and journal articles

We have deliberately excluded the regularly published literature from our analysis, that is, journal articles and books such as field guides. While most books and articles are still not available online or remain behind paywalls, major technical problems remain to extract and annotate them even in open access journals. The classical literature is certainly a more comprehensive and often more reliable source of images than web sites, hence we will investigate these sources in a future study.

Species coverage and intraspecies variation

Few species in our dataset are covered by thousands of photos, while many are covered by only a few. Gaps in the photograph dataset present a problem, including both species with low number of images and 27% of species with no readily available online images. Low numbers of photos likely underrepresent the phenotypic variation within species and across ages or sexes. Many species undergo ontogenetic changes in colour or pattern, and many more display sexual dimorphism (but still poorly documented on a larger scale). Most field guides and taxonomic monographs have similar issues, illustrating a species with only one or a few photos because of space limitations or production costs. Expanding image collections can address these issues, potentially building towards dynamic "next-generation field guides" (Farnsworth *et al.* 2013). We encourage the herpetology community to continue to submit images to one of the repositories used here (preferably with minimal copyright restrictions and to those platforms with direct data access routes such as iNaturalist), to capture the full phenotypic diversity.

Image annotation and metadata

For a more detailed analysis of images, photographers must provide metadata, e.g., location, time and date, descriptions such as age class and size. Metadata allows users to extract and infer numerous biological insights. For example, Maritz & Maritz (2020) used Facebook images to catalog the diets of snakes illuminating the ecological roles of these reptiles. More data may be extracted from images, such as size estimates (Mahendiran *et al.* 2018), if photographers upload images with exif (exchangeable image file format) data still attached (e.g., focal length, white balance, lens). Drury *et al*. (2019) present two case studies with damselflies, where virtually-taken measurements combined with location data allowed description of continent-wide phenotypic variation and insight into evolutionary questions.

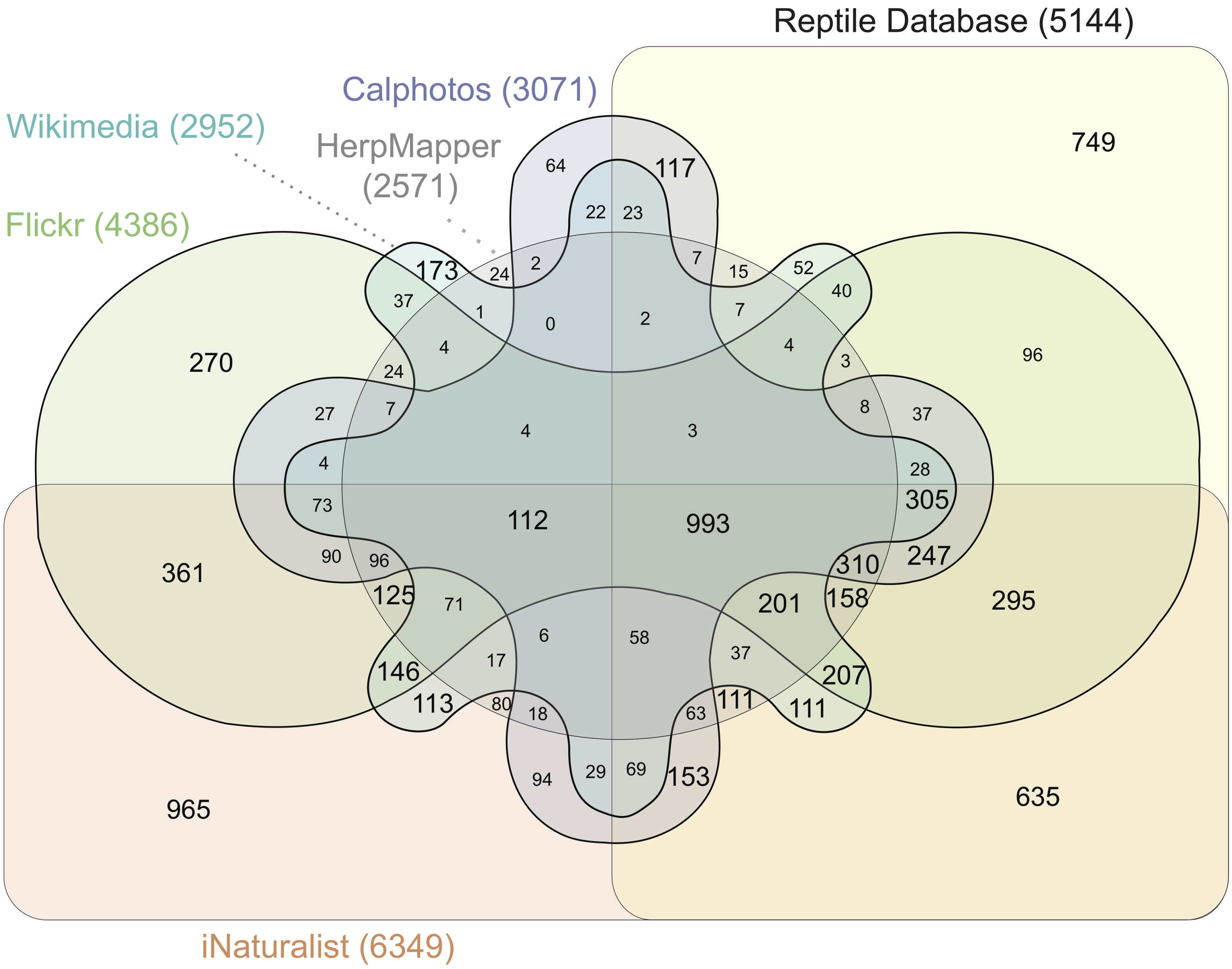

FIGURE 4. Overlap of species content in the 6 repositories used. For instance, 749 species (top right) have only photos in the Reptile Database while 635 species (bottom right) have photos both in the Reptile Database and iNaturalist, but not in any other repository. 295 species (middle right) have photos in the Reptile Database, iNaturalist, and Flickr, and so on. Numbers >99 are printed in larger font to emphasize their disproportional contribution.

Confidence in species identifications

A common problem with nature photographs is that species identities cannot always be ascertained from a single image, especially when similar species occur in the same locality (or if species are photographed in captivity). For instance, Ziegler *et al*. (2020) recently reported on a *Tupinambis* in the Cologne Zoo that was first thought to be *T. teguixin*. The animal also had some diagnostic characters of *T. cryptus*. DNA sequencing was required to finally confirm the specimen as the new species *T. cryptus*, which had been recently split from *T. teguixin* by Murphy *et al*. (2016).

Species identification can be difficult both for lay and expert contributors to community science projects (Austen *et al.* 2016). Verifying species identities are further complicated when images are of low resolution or only show parts of an animal (i.e., missing key characters; Austen *et al.* 2018; Gibbon *et al.* 2015). Our analysis involved tens of thousands of photos, hence there was no way for us to verify the reliability of species identifications in the 6 repositories used, forcing us to rely on IDs provided alongside images.

Identification issues will be particularly pronounced on platforms like Flickr, where there exists no formal system for correcting or crowd-sourcing species identities (as implemented by iNaturalist). While we suspect the species numbers provided by Flickr searches are likely inflated due to misidentifications, the overlap between Flickr and other sources (only 270/4386 species unique to Flickr) suggests that its overall impact on species coverage was limited. Any analysis built upon community data must address species identification issues (e.g., species distribution models' sensitivity to false positive records [Fernandes *et al.* 2018]) and treat community data on a case-by-case basis (Kosmala *et al.* 2016).

iNaturalist has also been working on image analysis for species identification, but this feature is still under development, although with certain taxonomic groups and image classes the success rate seems to be surprisingly high. While there is still substantial room for improvement this technology is advancing rapidly.

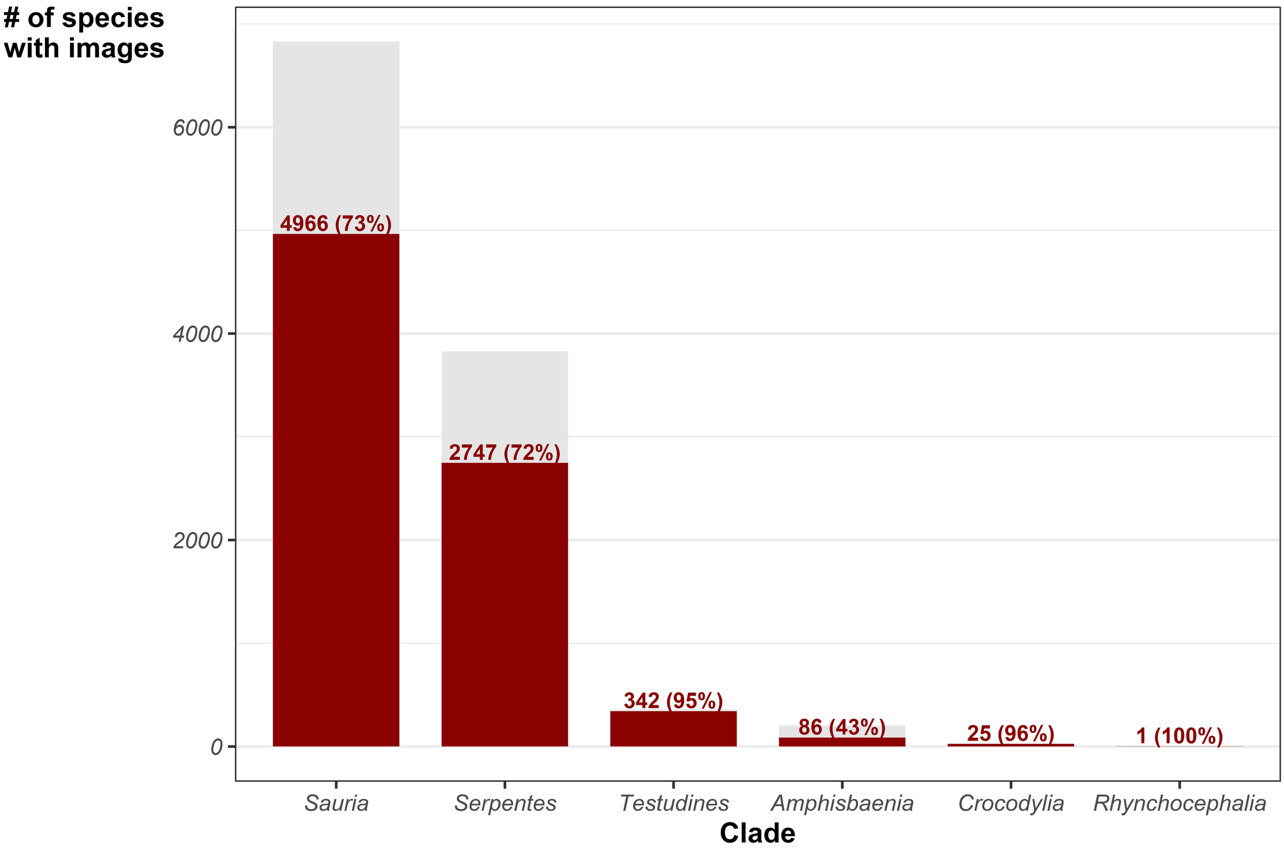

FIGURE 5. Number of species, per clade, with images from any of the data sources.

Changing taxonomy

As the aforementioned *Tupinambis* example indicates, another problem with species identifications is the ever-changing taxonomy. Species get split, synonymized, elevated from subspecies to species or simply change names. We have attempted to standardize names to the August 2020 version of the Reptile Database, but many names will continue to update. Constant changes can frustrate identification efforts and comparison of data sets. Originally correct identifications can become incorrect when species are split into new binomials. Rigorous tracking of locations (geo-tagging) and detailed photographs that allow users to reassess key characters can help mitigate the identification issues caused by species splits.

Automated image analysis and species recognition

Automated image classification has shown promise for streamlining species identification (Botella *et al.* 2018; Wäldchen *et al.* 2018), with iNaturalist's Seek App presenting a prominent example of image recognition implementation (https://www.inaturalist.org/pages/seek_app). One long-term goal of this study is to aid the creation of a dataset that could enable image analysis and automated species identification (Joly *et al.* 2020). While this will be possible for certain small groups, especially colourful snakes, it may be more difficult for large, homogenous

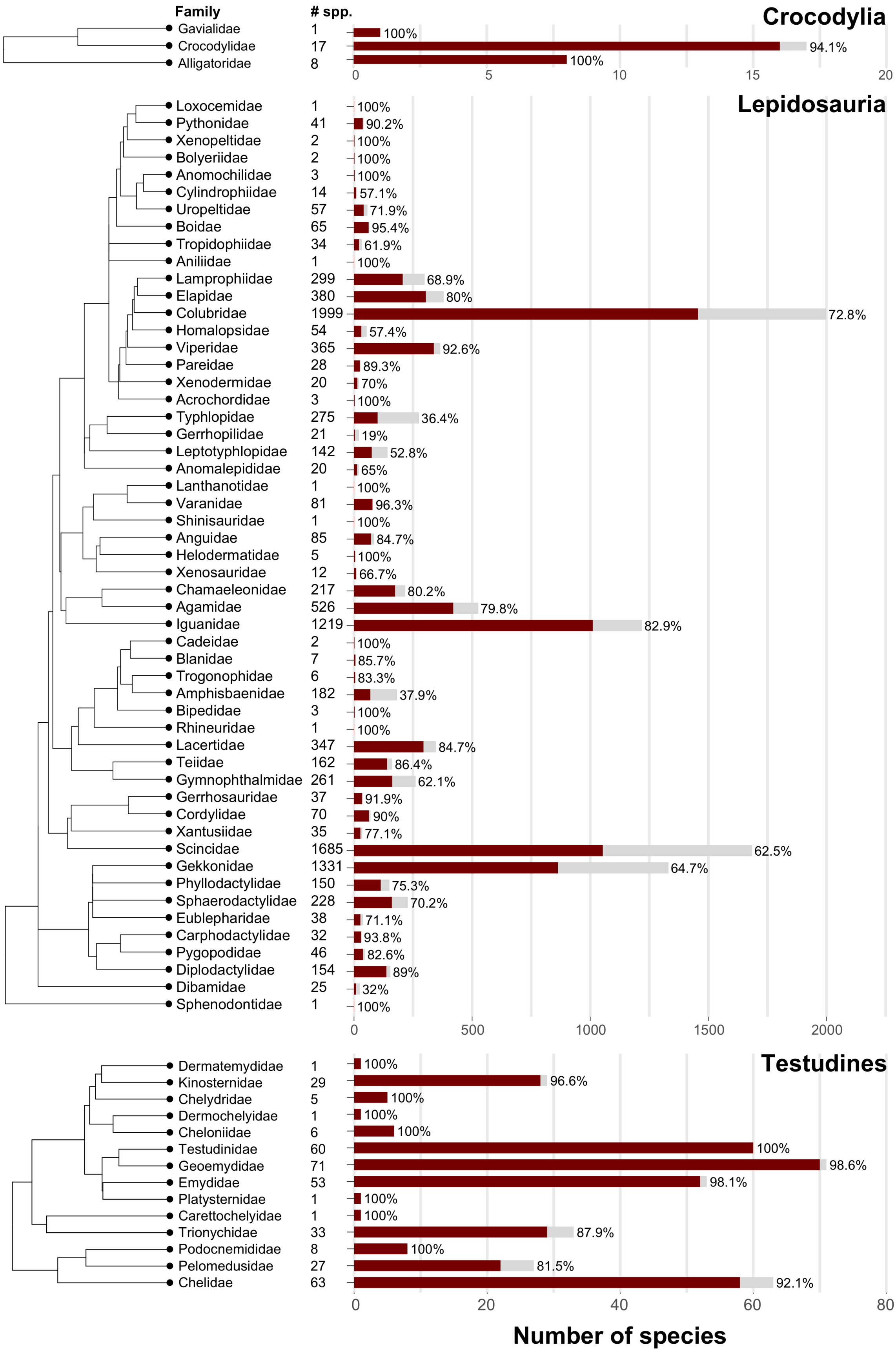

FIGURE 6. Total number of species (grey), number of species with photos (red), and percentage of species with photos per family from any of the data sources. Note different X-axis scales for each of the three clades shown (Crocodylia, Lepidosauria, Testudines).

groups such as blind snakes or many skinks that are frequently small, brown, and with esoteric field marks (Zhu *et al.* 2013). Metadata (e.g., location) in cases of look-alike species may be particularly important (Joly *et al.* 2020). Equally important, the training of such systems requires hundreds, if not thousands, of photos per species. Only the more easily identified species currently have photos in such numbers. Nevertheless, our data can serve as a foundation for future studies, and we hope that others will help us to expand this resource.

Use of images and copyright

Science is enriched by open and accessible datasets enabling reanalyses and replications, while setting foundations for larger studies (Poisot *et al.* 2013). However, images are by default protected by copyright. Some authors have argued that scientific (standardized) photographs are not creative products, thus do not qualify for copyright protection (Egloff *et al.* 2017). While the legal situation remains unresolved in many situations, photographers and scientists can help to mitigate this problem by defining a creative commons license (https://creativecommons.org/about/cclicenses/) when submitting images to a repository. For instance, iNaturalist and Flickr contributors can choose from a selection of creative commons copyright licenses to waive the "all rights reserved" default and ensure the availability of images. We strongly encourage everybody to use such open creative commons licenses in the interest of science.

Conclusion and outlook

We show that more than 70% of all reptile species are illustrated by photos in just 6 public image repositories. However, that leaves more than 3000 species without photos or with photos only available in less accessible sites. Many species are represented by only a few photos, indicating considerable progress required to complete this dataset. The Reptile Database has successfully worked with both iNaturalist and CalPhotos to link current names and photos; hopefully other repositories will join these efforts in order to make images systematically and transparently available to the scientific community and wider public.

Our inventory of photos highlights only one aspect of large biodiversity datasets. There are many other data types that need to be integrated into future biodiversity data projects (Miralles *et al*. 2020). Evidently, much more needs to be done to achieve a global biodiversity data infrastructure.

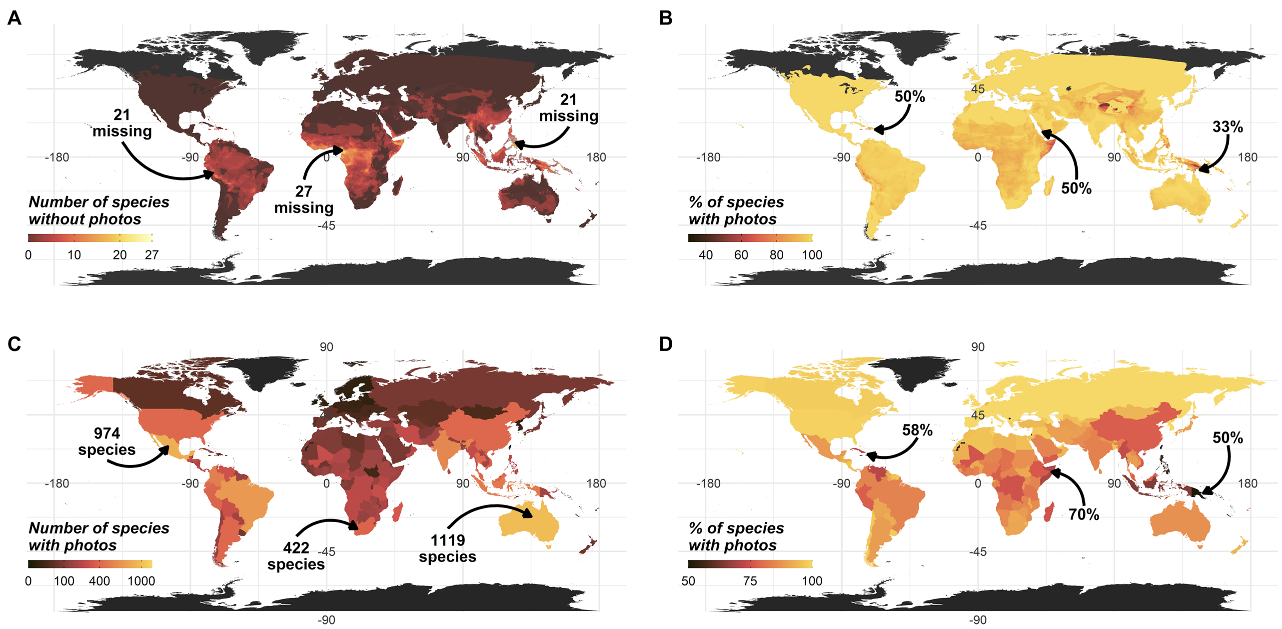

FIGURE 7. The distributions of species and images. (A) The number of species with zero images, with three hotspots for missing species highlighted. (B) The percentage of species with photos, with three cold spots highlighted. A and B are limited to the 10,064 species included in the GARD dataset (Roll *et al*. 2017). Spatial resolution is 0.1667 degrees. (C) The number of species with photos, on a per country basis, with hotspot highlighted (Mexico, South Africa, Australia). Note the colour scheme is square-rooted to aid differentiation of lower values. (D) The percentage of a country's species with images, with cold spots highlighted (from left to right: Haiti, Somalia, and Papua New Guinea).

Acknowledgements

We thank Suranaree University of Technology, Institute of Science and Institute of Research and Development for facilitating our research and providing logistical support. We thank the numerous photographers and community scientists tirelessly documenting and uploading records and photos of wildlife. Andrew Durso and Shai Meiri and one anonymous reviewer provided constructive comments that helped to improve the manuscript.

Author contributions

Conceptualization: PU. Data Curation: BMM, PU. Formal Analysis: BMM, PU. Investigation: BMM, PU. Resources: PF, LJV, PB, GV, SL, MF, JH, RS, BB, MRD, LJA, DJ, BK, BM, JH (*>4800 photos of >100 species each, 2,248 species total).* Methodology: BMM, PU. Visualization: BMM, PU. Writing—Original Draft Preparation: BMM, PU. Writing—Review & Editing: BMM, PU, LJV.

References

Allain, S. (2019) Mining Flickr: a method for expanding the known distribution of invasive species. *Herpetological Bulletin*, 148, 11–14.
https://doi.org/10.33256/hb148.1114

Auliya, M., Altherr, S., Ariano-Sanchez, D., Baard, E.H., Brown, C., Brown, R.M., Cantu, J., Gentile, G., Gildenhuys, P., Henningheim, E., Hintzmann, J., Kanari, K., Krvavac, M., Lettink, M., Lippert, J., Luiselli, L., Nilson, G., Nguyen, T.Q., Nijman, V., Parham, J.F., Pasachnik, S.A., Pedrono, M., Rauhaus, A., Rueda Córdova, D., Sanchez, M.E., Schepp, U., van Schingen, M., Schneeweiss, N., Segniagbeto, G.H., Somaweera, R., Sy, E.Y., Türkozan, O., Vinke, S., Vinke, T., Vyas, R., Williamson, S. & Ziegler, T. (2016) Trade in live reptiles, its impact on wild populations, and the role of the European market. *Biological Conservation*, 204, 103–119.
https://doi.org/10.1016/j.biocon.2016.05.017

Austen, G.E., Bindemann, M., Griffiths, R.A. & Roberts, D.L. (2016) Species identification by experts and non-experts: comparing images from field guides. *Scientific Reports*, 6, 33634. https://doi.org/10.1038/srep33634

Austen, G.E., Bindemann, M., Griffiths, R.A. & Roberts, D.L. (2018) Species identification by conservation practitioners using online images: accuracy and agreement between experts. *PeerJ*, 6, e4157.
https://doi.org/10.7717/peerj.4157

Botella, C., Joly, A., Bonnet, P., Monestiez, P. & Munoz, F. (2018) Species distribution modeling based on the automated identification of citizen observations. *Applications in Plant Sciences*, 6, e1029.
https://doi.org/10.1002/aps3.1029

Chang, W. (2015) *downloader: Download files over HTTP and HTTPS*. Available from: https://CRAN.R-project.org/package=downloader (accessed 31 October 2020)

Drury, J.P., Barnes, M., Finneran, A.E., Harris, M. & Grether, G.F. (2019) Continent-scale phenotype mapping using citizen scientists' photographs. *Ecography*, 42, 1436–1445.
https://doi.org/10.1111/ecog.04469

Dylewski, Ł., Mikula, P., Tryjanowski, P., Morelli, F. & Yosef, R. (2017) Social media and scientific research are complementary—YouTube and shrikes as a case study. *The Science of Nature*, 104, 48.
https://doi.org/10.1007/s00114-017-1470-8

Egloff, W., Agosti, D., Kishor, P., Patterson, D. & Miller, J. (2017) Copyright and the use of images as biodiversity data. *Research Ideas and Outcomes*, 3, e12502. https://doi.org/10.3897/rio.3.e12502

Farnsworth, E.J., Chu, M., Kress, W.J., Neill, A.K., Best, J.H., Pickering, J., Stevenson, R.D., Courtney, G.W., VanDyk, J.K. & Ellison, A.M. (2013) Next-generation field guides. *BioScience*, 63, 891–899.
https://doi.org/10.1525/bio.2013.63.11.8

Fernandes, R.F., Scherrer, D. & Guisan, A. (2018) Effects of simulated observation errors on the performance of species distribution models. *Diversity and Distributions*, 25.3, 1–14. https://doi.org/10.1111/ddi.12868

Flickr Development Team (2019) *Flickr API*. Flickr. Available from: https://www.flickr.com/services/api/ (accessed 31 October 2020)

Gibbon, G.E.M., Bindemann, M. & Roberts, D.L. (2015) Factors affecting the identification of individual mountain bongo antelope. *PeerJ*, 3, e1303.
https://doi.org/10.7717/peerj.1303

Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P. & Minghim, R. (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 16, 169. https://doi.org/10.1186/s12859-015-0611-3

Hedges, S.B., Dudley, J. & Kumar, S. (2006) TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics*, 22, 2971–2972. https://doi.org/10.1093/bioinformatics/btl505

Hijmans, R.J. (2020) *raster: Geographic data analysis and modeling*. Available from: https://CRAN.R-project.org/package=raster (accessed 31 October 2020)

Jarić, I., Correia, R.A., Brook, B.W., Buettel, J.C., Courchamp, F., Di Minin, E., Firth, J.A., Gaston, K.J., Jepson, P., Kalinkat, G., Ladle, R., Soriano-Redondo, A., Souza, A.T. & Roll, U. (2020) iEcology: Harnessing large online resources to generate ecological insights. *Trends in Ecology & Evolution*, S016953472030077X. https://doi.org/10.1016/j.tree.2020.03.003

Jiménez-Valverde, A., Peña-Aguilera, P., Barve, V. & Burguillo-Madrid, L. (2019) Photo-sharing platforms key for characterising niche and distribution in poorly studied taxa. *In*: Leather, S. & Stewart, A. (Eds.), *Insect Conservation and Diversity*, icad.12351. https://doi.org/10.1111/icad.12351

Joly, A., Goeau, H., Deneu, B., Cole, E., de Castaneda, R.R., Bolon, I., Durso, A., Lorieul, T., Glotin, H., Eggel, I., Vellinga, W.-P. & Muller, H. (2020) Overview of LifeCLEF 2020: a system-oriented evaluation of automated species identification and species distribution Prediction. *Lecture Notes in Computer Science*, 25, 342–363 https://doi.org/10.1007/978-3-030-58219-7_23

Kassambara, A. (2018) *ggpubr: "ggplot2" based publication ready plots*. Available from: https://CRAN.R-project.org/package=ggpubr (accessed 31 October 2020)

Keyes, O., Jacobs, J., Schmidt, D., Greenaway, M., Rudis, B., Pinto, A., Khezrzadeh, M., Meilstrup, P., Costello, A.M., Bezanson, J., Meilstrup, P. & Jiang, X. (2019) *urltools: Vectorised tools for URL handling and parsing*. Available from: https://cran.r-project.org/package=urltools (accessed 31 October 2020)

Kidd, L.R., Gregg, E.A., Bekessy, S.A., Robinson, J.A. & Garrard, G.E. (2018) Tweeting for their lives: Visibility of threatened species on twitter. *Journal for Nature Conservation*, 46, 106–109. https://doi.org/10.1016/j.jnc.2018.10.001

Kosmala, M., Wiggins, A., Swanson, A. & Simmons, B. (2016) Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14, 551–560. https://doi.org/10.1002/fee.1436

Kumar, S. & Hedges, S.B. (2011) Timetree2: Species divergence times on the iPhone. *Bioinformatics*, 27, 2023–2024. https://doi.org/10.1093/bioinformatics/btr315

Kumar, S., Stecher, G., Suleski, M. & Hedges, S.B. (2017) TimeTree: A resource for timelines, timetrees, and divergence times. *Molecular biology and evolution*, 34, 1812–1819. https://doi.org/10.1093/molbev/msx116

Lang, D.T. & The CRAN Team (2018) *RCurl: General Network (HTTP/FTP/...) Client interface for R*. Available from: https://CRAN.R-project.org/package=RCurl (accessed 31 October 2020)

Mahendiran, M., Parthiban, M., Azeez, P.A. & Nagarajan, R. (2018) In situ measurements of animal morphological features: A non-invasive method. *In*: Fisher, D. (Ed.), *Methods in Ecology and Evolution*, 9, pp. 613–623. https://doi.org/10.1111/2041-210X.12898

Maritz, R.A. & Maritz, B. (2020) Sharing for science: high-resolution trophic interactions revealed rapidly by social media. *PeerJ*, 8, e9485. https://doi.org/10.7717/peerj.9485

Marshall, B.M. & Strine, C.T. (2019a) Exploring snake occurrence records: Spatial biases and marginal gains from accessible social media. *PeerJ*, 7, e8059. https://doi.org/10.7717/peerj.8059

Marshall, B.M. & Strine, C.T. (2019b) Supplementary information for Marshall & Strine 2019: Code, data and packages. *PeerJ*, 7, e8059. https://doi.org/10.7717/peerj.8059

Marshall, B.M., Strine, C. & Hughes, A.C. (2020) Thousands of reptile species threatened by under-regulated global trade. *Nature Communications*, 11, 4738. https://doi.org/10.1038/s41467-020-18523-4

McClain, C.R. (2019) Likes, comments, and shares of marine organism imagery on Facebook. *PeerJ*, 7, e6795. https://doi.org/10.7717/peerj.6795

Measey, J., Basson, A., Rebelo, A.D., Nunes, A.L., Vimercati, G., Louw, M. & Mohanty, N.P. (2019) Why have a pet amphibian? Insights from YouTube. *Frontiers in Ecology and Evolution*, 7, 52. https://doi.org/10.3389/fevo.2019.00052

Miranda, E.B.P., Ribeiro, R.P. & Strüssmann, C. (2016) The ecology of human-anaconda conflict: a study using internet videos. *Tropical Conservation Science* 9, 43–77. https://doi.org/10.1177/194008291600900105

Miralles, A., Bruy, T., Wolcott, K., Scherz, M.D., Begerow, D., Beszteri, B., Bonkowski, M., Felden, J., Gemeinholzer, B., Glaw, F., Glöckner, F.O., Hawlitschek, O., Kostadinov, I., Nattkemper, T.W., Printzen, C., Renz, J., Rybalka, N., Stadler, M., Weibulat, T., Wilke, T., Renner, S.S. & Vences, M. (2020) Repositories for Taxonomic Data: Where We Are and What is Missing. *Systematic biology*, syaa026. https://doi.org/10.1093/sysbio/syaa026

Murphy, J.C., Jowers, M.J., Lehtinen, R.M., Charles, S.P., Colli, G.R. Jr., A.K.P., Hendry, C.R. & Pyron, R.A. (2016) Cryptic,

sympatric diversity in Tegu Lizards of the *Tupinambis teguixin* group (Squamata, Sauria, Teiidae) and the description of three new species. *PLOS ONE*, 11, e0158542.
https://doi.org/10.1371/journal.pone.0158542

Ooms, J. (2014) The jsonlite Package: A practical and consistent mapping between JSON data and R objects. *arXiv*, 1403.2805. [published online]

Otsuka, R. & Yamakoshi, G. (2020) Analyzing the popularity of YouTube videos that violate mountain gorilla tourism regulations. *In*: Lau, E.H. (Ed.), *PLOS ONE*, 15, e0232085. https://doi.org/10.1371/journal.pone.0232085

Paradis, E. & Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528.

Pebesma, E. (2018) Simple features for r: Standardized support for spatial vector data. *The R Journal*, 10, 439–446. https://doi.org/10.32614/RJ-2018-009

Pedersen, T.L. & Crameri, F. (2018) scico: Colour Palettes Based on the Scientific Colour-Maps.

Poisot, T., Mounce, R. & Gravel, D. (2013) Moving toward a sustainable ecological science: don't let data go to waste! *Ideas in Ecology and Evolution*, 6.
https://doi.org/10.4033/iee.2013.6b.14.f

R Core Team (2020) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna. Available from: https://www.r-project.org/ (accessed 31 October 2020)

R Studio Team (2020) *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA. Available from: http://www.rstudio.com/ (accewssed 10 November 2020)

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A.M., Bernard, R., Böhm, M., Castro-Herrera, F., Chirio, L., Collen, B., Colli, G.R., Dabool, L., Das, I., Doan, T.M., Grismer, L.L., Hoogmoed, M., Itescu, Y., Kraus, F., LeBreton, M., Lewin, A., Martins, M., Maza, E., Meirte, D., Nagy, Z.T., de C. Nogueira, C., Pauwels, O.S.G., Pincheira-Donoso, D., Powney, G.D., Sindaco, R., Tallowin, O.J.S., Torres-Carvajal, O., Trape, J.-F., Vidan, E., Uetz, P., Wagner, P., Wang, Y., Orme, C.D.L., Grenyer, R. & Meiri, S. (2017) The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology & Evolution*, 1, 1677–1682.
https://doi.org/10.1038/s41559-017-0332-2

Roll, U., Mittermeier, J.C., Diaz, G.I., Novosolov, M., Feldman, A., Itescu, Y., Meiri, S. & Grenyer, R. (2016) Using Wikipedia page views to explore the cultural importance of global reptiles. *Biological Conservation*, 204, 42–50.
https://doi.org/10.1016/j.biocon.2016.03.037

Ross, N. (2020) *fasterize: Fast polygon to raster conversion*. Available from: https://CRAN.R-project.org/package=fasterize (accessed 31 October 2020)

Slowikowski, K. (2018) *ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2."* Available from: https://CRAN.R-project.org/package=ggrepel (accessed 31 October 2020)

Wäldchen, J., Rzanny, M., Seeland, M. & Mäder, P. (2018) Automated plant species identification—Trends and future directions. *In*: Bucksch, A. (Ed.), *PLOS Computational Biology*, 14, e1005993.
https://doi.org/10.1371/journal.pcbi.1005993

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York. Available from: https://cran.r-project.org/web/packages/ggplot2/index.html (accessed 31 October 2020)

Wickham, H. (2019a) *rvest: Easily Harvest (Scrape) Web Pages*. Available from: https://CRAN.R-project.org/package=rvest (accessed 31 October 2020)

Wickham, H. (2019b) *stringr: Simple, consistent wrappers for common string operations*. Available from: https://CRAN.R-project.org/package=stringr (accessed 31 October 2020)

Wickham, H., François, R., Henry, L. & Müller, K. (2019) *dplyr: A Grammar of Data Manipulation*. Available from: https://CRAN.R-project.org/package=dplyr (accessed 31 October 2020)

Wickham, H. & Henry, L. (2019) *tidyr: Easily Tidy Data with "spread()" and "gather()" Functions*. Available from: https://CRAN.R-project.org/package=tidyr (accessed 31 October 2020)

Wickham, H., Hester, J. & Ooms, J. (2018) *xml2: Parse XML*. Available from: https://CRAN.R-project.org/package=xml2 (accessed 31 October 2020)

Wilke, C.O. (2019) *cowplot: Streamlined plot theme and plot annotations for "ggplot2."* Available from: https://CRAN.R-project.org/package=cowplot (accessed 31 October 2020)

Yu, G., Lam, T.T.-Y., Zhu, H. & Guan, Y. (2018) Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Molecular Biology and Evolution*, 35, 3041–3043.
https://doi.org/10.1093/molbev/msy194

Yu, G., Smith, D., Zhu, H., Guan, Y. & Lam, T.T.-Y. (2017) ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8, 28–36.
https://doi.org/10.1111/2041-210X.12628

Zaher H, Murphy RW, Arredondo JC, Graboski R, Machado-Filho PR, Mahlow K, *et al*. (2019) Large-scale molecular phylogeny, morphology, divergence-time estimation, and the fossil record of advanced caenophidian snakes (Squamata: Serpentes). *PLOS ONE*, 14 (5), e0216148.
https://doi.org/10.1371/journal.pone.0216148

Zhu, W., Drewes, J. & Gegenfurtner, K.R. (2013) Animal detection in natural images: Effects of color and image database.

PLOS ONE, 8, 1–14.
https://doi.org/10.1371/journal.pone.0075816

Ziegler, T., Rauhaus, A. & Vences, M. 2020 Welt-Zooerstzucht des Kryptischen Goldtejus (*Tupinambis cryptus*) im Terrarium des Kölner Zoos. *Elaphe*, 2020 (1), 22–31

SUPPLEMENTARY INFORMATION

Supplementary Table 1. List of all species and the number of photos in each of the 6 repositories.

Supplementary Table 2. List of species without photos in any of the 6 repositories.

Supplementary Table 3. Per country summary data of number of species present and number with images.

Supplementary information (available at Zenodo: https://doi.org/10.5281/zenodo.4010155):
SuppCode1_Flickr_search.R: Code used to retrieve Flickr image data
SuppCode2_Wikimedia_query.R: Code used to retrieve Wikimedia image data,
SuppCode3_HerpMapper_search.R: Code used to retrieve HerpMapper image data
SuppCode4_Figure Generation.R: Code used to generate figures 2, 3 and 5-7, also.
We have included additional data in the Zenodo repository required to reproduce figures: reptile_checklist_2020_04.csv, reptile names 2019 syno.csv, GARD non-repDB_fixed.csv, reptiles_family_crocodylia.nwk, reptiles_family_lepidosauria.nwk, reptiles_family_testudines.nwk.
GARD species distribution data are available on Dryad: https://doi.org/10.5061/dryad.83s7k.