Received: 27 March 2023 | Revised: 24 August 2023 | Accepted: 4 September 2023

DOI: 10.1111/ddi.13771

RESEARCH ARTICLE

Diversity and Distributions WILEY

Using automated content analysis to monitor global online trade in endemic reptile species

Jooel Rinne[1] | Ritwik Kulkarni[1] | Andrea Soriano-Redondo[1] | Ricardo Correia[1,2,3] | Enrico Di Minin[1,2,4]

[1]Helsinki Lab of Interdisciplinary Conservation Science, Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

[2]Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland

[3]Biodiversity Unit, University of Turku, Turku, Finland

[4]School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa

Correspondence
Enrico Di Minin, Helsinki Lab of Interdisciplinary Conservation Science, Department of Geosciences and Geography, University of Helsinki, Helsinki 00014, Finland.
Email: enrico.di.minin@helsinki.fi

Funding information
Academy of Finland, Grant/Award Number: 348352; H2020 European Research Council, Grant/Award Number: 802933; H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 101022521; KONE Foundation research grant, Grant/Award Number: 202101976 and 202103830

Editor: Matthew Luskin

Abstract

Aim: Online reptile trade poses new challenges to species conservation and requires automated monitoring. Range-restricted and endemic reptile species are especially vulnerable to wildlife trade and unsustainable exploitation. In this study, we investigated the magnitude and geographic distribution of online trade of 96 endemic and range-restricted reptile species from the Lesser Antilles.

Location: Global.

Methods: We developed methods for automated collection, filtering and processing of wildlife trade content for the targeted species from publicly accessible online platforms.

Results: We identified 599 relevant advertisements originating from 231 different advertisers and 41 websites focusing on 43 species. Species advertised included threatened species according to the International Union for the Conservation of Nature (IUCN) Red List and species listed in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) Appendices. Among threatened species, five are Critically Endangered, three are Endangered and two are Vulnerable according to the IUCN Red List. Moreover, three of the six most advertised species were classified as Near Threatened. Germany was the country with the highest number of advertisements ($N=124$), followed by the United States ($N=55$), the Netherlands ($N=15$) and United Kingdom ($N=15$). Based on data from sale advertisements that included price and currency data, prices ranged from one to over a thousand Euros.

Main Conclusions: We present a framework for automated analysis of online trade in reptiles that can be extended to other taxonomic groups. Our results highlight countries, such as Germany and the United States, where enhanced monitoring actions would be important to assess the origin (i.e. captive bred or wild sourced individuals) and the legality of the trade. Immediate conservation actions, such as population monitoring, are also needed to ensure wildlife trade is not threatening the persistence of endemic reptile populations in the wild.

KEYWORDS
biodiversity loss, digital media, internet, online wildlife trade, reptiles, threatened species

1 | INTRODUCTION

An unprecedented crisis triggered by humans is affecting biodiversity globally (Barnosky et al., 2011). Despite some local conservation successes and increasing protected area coverage, biodiversity continues to decline (Johnson et al., 2017). This is primarily the result of human pressures on biodiversity continuing to increase (Betts et al., 2017; IPBES, 2019; Mantyka-Pringle et al., 2012; Urban, 2015). According to the International Union for the Conservation of Nature (IUCN), habitat loss and unsustainable exploitation are the major human pressures affecting biodiversity worldwide (Maxwell et al., 2016). Regions under high-intensity threat from unsustainable exploitation are especially concentrated in North and South America and Asia (Di Minin, et al., 2019). Unsustainable wildlife trade, a dimension of unsustainable exploitation, affects at least 24% of all terrestrial vertebrate species on Earth (Scheffers et al., 2019).

The internet has made wildlife trade more accessible as species can be traded on different platforms, such as online marketplaces, forums and social media sites (Altherr & Lameter, 2020; Di Minin, Fink, et al., 2019; Fink et al., 2021; Lavorgna, 2014). Because of the extremely large volume of content on internet websites, digital platforms and trade advertisements, online trade can be difficult to monitor and assess (Demeau, Vargas, & Jeffrey, 2019). Advanced methods for data filtering and analysis are therefore needed to effectively monitor online wildlife trade. Computer vision models and natural language processing are being used increasingly to automatically identify image and/or text content pertaining to wildlife trade from online platforms (Cardoso et al., 2023; Di Minin, Brooks, et al., 2019; Kulkarni & Di Minin, 2021, 2023). However, research on these topics should follow the highest data privacy and protection standards as it involves using personal data to investigate potentially illegal activities (Di Minin et al., 2021). Still, automated mining and analysis of online data can provide new insights to help investigate the geography and intensity of wildlife trade at multiple scales.

Reptiles are among the most traded vertebrate species (Scheffers et al., 2019). International reptile trade of wild specimens can be legal when performed following regulations by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) or illegal if done in violation of these regulations. At the national level, regulations of trade are country and species-specific. Thus, while international trade might be banned for certain species, it might still be legal within the national borders of certain countries (Auliya et al., 2016). Up to 35% of all reptile species are traded online, with only 9% of these species included in CITES (Marshall et al., 2020). Worryingly, approximately 90% of traded reptile species and half of traded individuals are sourced from the wild (Marshall et al., 2020). Distinct appearance and rarity in the wild or in the pet market are some of the attributes desired by reptile owners, thus threatened and/or range-restricted species are especially popular (Altherr & Lameter, 2020; Auliya et al., 2016; Courchamp et al., 2006). More detailed assessments are needed to assess the online trade in endemic, range-restricted, reptile species threatened by the illegal wildlife trade.

In this study, we investigated the online trade of 96 reptile species originating from the Lesser Antilles (Table S1; Figure S1) that are restricted to small islands or island banks (Hedges, 2006; Wilson et al., 2011). Small population sizes and endemic status make these species popular as exotic pets. Moreover, the difficulty in captive breeding some of these species increases the risk of exploitation of wild populations (e.g. Langner et al., 2022; Shepherd et al., 2019). Here, we developed methods for automated monitoring (i.e. collection, filtering and processing) of wildlife trade content for the targeted species from publicly accessible online platforms. Specifically, we investigated (i) which species were involved in the trade; (ii) which countries played an important role in the trade; and (iii) what were the prices demanded for these species.

2 | METHODS

2.1 | Study area and species

We compiled a list of Lesser Antillean species of interest combining species information from Fauna and Flora international (2021) ($N=113$) and Caribherp (Hedges, 2020) ($N=86$). We excluded any species with geographic ranges expanding beyond the Lesser Antilles islands ($N=19$). We used the Reptile Database (Uetz et al., 2020), the Global Assessment of Reptile Distributions database (Roll et al., 2017) and the IUCN Red List of Threatened Species (IUCN, 2023) to assess the ranges of the species. Our list consisted of 93 species and 3 subspecies (*Iguana iguana melanoderma*, *Iguana iguana insularis* and *Iguana iguana sanctaluciae*). These three subspecies are endemic to the region and are considered as distinct species in some taxonomies (e.g. Breuil et al., 2020, 2022). Among these 96 species and subspecies, 69 are lizards and 27 are snakes, and 40 are considered threatened with extinction by the IUCN (i.e. assessed as either Vulnerable, Endangered or Critically Endangered) (IUCN, 2023). Both regional (e.g. the Specially Protected Areas and Wildlife Protocol) and national legislation in the Lesser Antilles require protecting threatened species, for example, by regulating or prohibiting the harvest of wild populations (Noseworthy, 2017; UN Environment, 1990). In addition, nine species or subspecies are listed in CITES appendices (Table S2). One species, the Union Island Gecko (*Gonatodes daudini*), is listed in Appendix I, meaning that the commercial international trade of wild specimens of this species is prohibited (CITES, 2023). The rest of the species are listed in

Appendix II meaning that international trade of the species can be authorized, but requires appropriate export permits or certificates (CITES, 2023).

2.2 | Data collection, processing and analysis

We identified a list of websites on which the species of interest were being traded by using (i) Google Search (see e.g. Marshall et al., 2020) and (ii) a list of websites identified as part of a previous study (Noseworthy, 2017; Figure 1). In the case of Google searches, we used two query approaches; one using a search string composed by the scientific name of each reptile species and 'for sale', and another using the search string 'reptiles for sale' in 12 languages (English, Dutch, Polish, French, Italian, Indonesian, German, Portuguese, Spanish, Japanese, Hindi and Filipino). To choose the languages for the search, we looked for evidence of search interest for the species using data from Google Trends (2021). We sampled internet search volume from Google Trends using topic searches and the species as topics, which considers language variations in the ways species are searched for in different contexts. Based on the data retrieved, we identified countries with relevant search volume and the dominant language of these locations from the list of languages spoken in each country provided by CIA World Factbook (Central Intelligence Agency, 2021). We then used this information to do online searches using the Google Search Engine in the relevant languages, aiming to replicate the behaviour of potential buyers. For each Google search, we collected websites likely to show relevant trade activity from the first 100 search results as these contained the most relevant search results. The websites were manually vetted to assess if any Lesser Antillean species were potentially being traded, thus making them a target for further data collection. These websites included online shops of pet stores, reptile forums, marketplaces and websites of individual reptile breeders and hobbyists. Using this method, we gathered a list of 90 websites of interest (Figure S2).

We scraped data from the websites of interest using ScraPy version 2.4.1 (Kouzis-Loukas, 2016), which is a Python extension that can be used to extract data from websites (Figure S3). Specifically, we developed a web scraper that accesses and collects information from the contents of a website and stores this information in a SQL database (Diouf et al., 2019). The scraper uses Cascading Style Sheets (CSS) selectors to indicate the location of relevant data elements found on the website's HTML document. For each website scraped, we created a configuration file which consisted of the website's address and the CSS selectors. Each entry in the database represented a reptile trade item or a forum post depending on the type of website. We collected data between 25 January 2021 and 6 September 2021. Each website was scraped once. We retrieved a total of 366,951 entries from 90 websites with the data fields shown in Table 1.

After the data collection process, we processed the raw data for further analysis. First, we extracted data entries mentioning Lesser Antillean reptiles. For this purpose, we created a list of synonyms of scientific names and common names, so any nicknames and variations of names used to refer to target species would be included. The list was created using the list of reptile names from Marshall et al. (2020) as a basis and adding species synonyms from the Reptile database (Uetz et al., 2020), Wikidata and Wikipedia by manually searching each species on each site. We then created a Python script that searched the raw data for any mentions of the reptile names on the synonyms list. Data entries with no mentions of Lesser Antillean reptiles were dropped. To further process the data, another Python script was used to determine the exact species mentioned on each data entry and to create individual species entries if more than one species of interest were mentioned on the original data entry.

The next data processing phase included extracting information on the numbers of individual reptiles being traded using a custom-made Python script. The script searched the 'quantity', 'reptile' and 'description' data elements for numerical values indicating the quantity of individual reptiles. To extract quantity information from

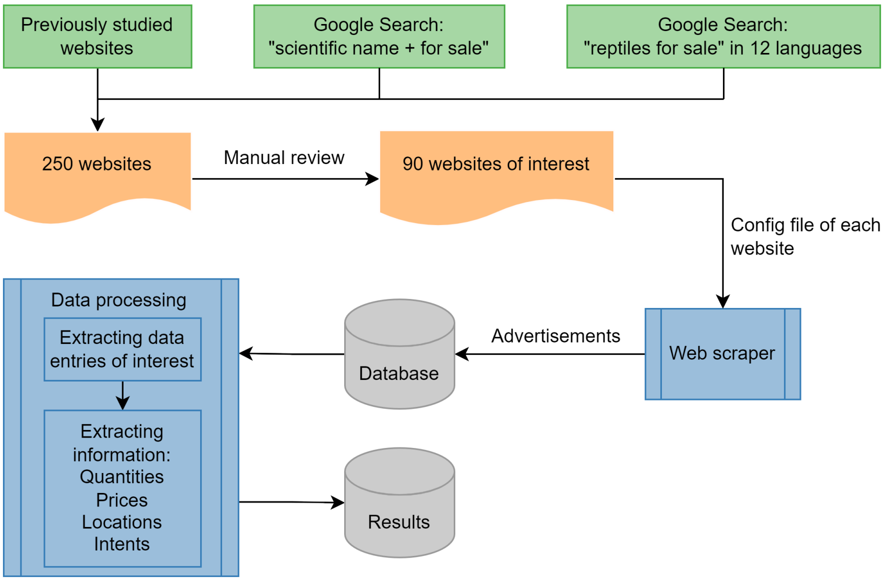

FIGURE 1 Flowchart of data collection and processing. The list of websites was collected from three different sources. After manual review, 90 websites were scraped for data. Data were then stored into a database and processed.

TABLE 1 Data elements collected from the websites.

Data element	Explanation
Timestamp	Information on when the sale item was posted
Reptile	The name of the reptile or the title of the item traded
Price	Price of the item traded
Quantity	Quantity of stock of the item traded
Description	Longer description of the species or item traded
Comments	Customer or other comments regarding the item traded
Location	Assumed location of the seller or buyer of the item traded
Image	URL to an image of the reptile or the item traded
Acquired	Date the data were scraped from the internet
Source	Web URL to the website where the sales item is listed
Intent	Is the item traded being sold or looked to be bought
Other	Other information
Website type	The type of the website (online shop/forum/marketplace/other)

text-based data elements, we used the Natural Language Processing Library SpaCy version 3.0.5 (Honnibal & Montani, 2017) for Python. The script also extracted price information from a data entry by checking if the 'price' and 'currency' element had any data stored. If price data were found, the script converted the price to Euros using forex-python library (Micro Pyramid Informatic Pvt. Ltd, 2016). The date in the 'timestamp' data element of a data entry was used to calculate the correct conversion rate between the currency in question and the Euro. Lastly, the script searched the data fields for the intent of the data entry, meaning whether the advertisement was posted for selling or buying reptiles, or for other reasons.

The script also pseudonymized the advertisement poster's name or nickname by converting it to a numerical ID, so the number of advertisements per trader could be counted while ensuring the anonymity of the original poster (Di Minin et al., 2021). We only scraped data from websites that were publicly available and did not prohibit web scraping. All data elements from which an individual user could be identified were removed or replaced by an ID number. Location data are presented only at a country level. We created an additional script that used SpaCy to label words in 'location' data element. The words identified as geopolitical entities were fed to Python's Geocoder version 1.38.1 library (Carriere, 2013) to find the matching coordinates for that location from OpenStreetMaps (OpenStreetMap contributors, 2015). If the location was not detected using this approach, the 'description' data element was searched for any geographic information. The accuracy of the location information varied depending on the accuracy of the location in the advertisements. Therefore, we considered such uncertainty over precise location, by upscaling to country-level information. This was also done to address data privacy issues, as more accurate location data could enable the identification of the advertisers. Finally, to avoid duplicate entries if the same advertisement was posted multiple times or on several websites, we compared the species names, quantity, price, currency, intent and location information. If all the information matched between two entries and the trader ID was the same, the entries were considered as duplicates and only one was kept for analysis. No names, nicknames, usernames, content of the trade advertisements or specific locations were kept after data processing. Advertisement content (e.g. title of the advertisement and the description of the species) was removed and only species, quantity, price, currency, intent, trader ID and location on a country level were kept.

An accuracy assessment of the scripts was conducted to estimate the performance of the scripts in extracting relevant information. The assessment was done in two stages by estimating the accuracy of extracting (i) a post that contained a target species from raw data and (ii) an entity of interest such as quantity, location and price from a relevant advertisement. Specifically, we selected a test data set of 380 data entries from the raw data and used it to measure how well the data processing algorithm extracted data entries of interest (advertisements mentioning Lesser Antillean Reptiles) from the raw data. We selected 300 of those entries of which three were of interest and 297 not of interest based on manual review. The remaining 80 data entries were manually reviewed and determined to be of interest before adding them to the test data. We included these entries because of the small portion of data entries of interest in the raw data. To sum up, the test data consisted of 83 data entries of interest and 297 data entries not of interest.

We then let the data processing algorithm classify the data entries in the test data into two categories: of interest and not of interest. The algorithm classified 366 of the 380 data entries correctly meaning that the overall accuracy of the algorithm was 96.3% (Table S3). The algorithm classified the data entries of interest with accuracy of 83.1% and irrelevant data entries with an accuracy of 100%. Accuracy scores for each assessment are reported in Table S4.

We also assessed the algorithm's accuracy in extracting data elements from the data entries of interest. The data elements were species, quantity, price, currency, location and intent. The test data consisted of 100 randomly selected data entries of interest. We manually reviewed the test data and then compared the performance of the algorithm to the reviewed data. We then calculated the accuracy for each data element. If an element was not found in either reference data or classified data, we did not take it into consideration when assessing accuracy.

We used the number of advertisements instead of the actual number of reptiles to quantify trade instances because information on the quantity of individuals being traded was not consistently available. We then explored the relationship between the intent of the advertiser, the species conservation status, the CITES Appendix status and the number of trade advertisements. Data entries with location information were used to map the number of advertisements per country using QGIS version 3.10.2 (QGIS Development Team, 2019). We also plotted the number of advertisements selling CITES listed Lesser Antillean species. Price information was used to calculate mean, median, standard deviation and range of reptile selling prices.

3 | RESULTS

We found 599 advertisements of 43 Lesser Antillean reptile species originating from 231 distinct advertisers and 41 websites. The advertisements were posted online between 21 April 2003 and 6 September 2021. We identified 428 sales advertisements, 106 buying advertisements and 65 that were categorized as other advertisements (i.e. listings of pets that were not on sale) or advertisements for which the intent was not clear (Table 2).

The species most commonly listed for sale were the Least Concern Martinique Anole (*Anolis roquet*, *N*=70), the Least Concern Guadeloupe Anole (*Anolis marmoratus*, *N*=35), the Near Threatened Saint Vincent Treeboa (*Corallus cookii*, *N*=33), the Least Concern Leeward Banded Geckolet (*Sphaerodactylus sputator*, *N*=32), the Near Threatened Dominica Anole (*Anolis oculatus*, *N*=29) and the Near Threatened Saba Anole (*Anolis sabanus*, *N*=29) (Figure 2). We also found instances of online trade in threatened species, including five Critically Endangered, three Endangered and two Vulnerable species. We also found instances of online trade for nine species listed in the CITES appendices. Among these species, the Appendix II listed Saint Vincent Treeboa was the species with most trade records, followed by the Appendix II listed Critically Endangered Lesser Antillean Iguana (*Iguana delicatissima*, *N*=15), the Appendix I listed Union Island Gecko (*Gonatodes daudini*, *N*=6) and the Appendix II listed Saba Iguana (*Iguana melanoderma*, *N*=5). It should be noted that three of the six advertisements selling the Union Island Gecko were dated before the species was listed in Appendix I, meaning that, at the time, international trade of the species was not regulated.

Germany was the country with the highest number of advertisements of Lesser Antillean reptiles (*N*=124), followed by the United States (*N*=55), the Netherlands (*N*=15) and United Kingdom (*N*=15) (Figure 3a). Other countries involved in trade were Czech Republic (*N*=9), Japan (*N*=9), Canada (*N*=6), Poland (*N*=5), Spain (*N*=2), Austria (*N*=1), Belgium (*N*=1) and Greece (*N*=1). Germany and the United States were also the countries with the highest number of species on sale or requested to buy in the identified advertisements (Figure 3b). Based on advertisements with location information, CITES listed species were found for sale in United States (*N*=13), Germany (*N*=6), United Kingdom (*N*=6), Japan (*N*=3), Czech Republic (*N*=2), the Netherlands (*N*=1) and Austria (*N*=1). Germany was confirmed to be the country with the highest number of sales across all species (Figure 2). However, trade in Critically Endangered species appeared to be especially important in the United Sates.

Based on data from 42 selling advertisements that had price and currency data, the sale price of the Lesser Antillean reptile species ranged from 1 to 1381.2 Euros. The advertisements were posted online over approximately 6 years, and the total sum of the asking prices was 9735.3 Euros. The median listing price in a selling advertisement was approximately 92 Euros and the interquartile range 140.2 Euros. *Iguana* species were the most expensive and listed for sale for over 1000 Euros, followed by *Thecadactylus* species that were listed for sale between 200 and 800 Euros.

TABLE 2 Conservation status and number of advertisements per intent of the advertiser.

Conservation status	Seller	Buyer	Other	Total
Least Concern (LC)	255	76	42	373
Near Threatened (NT)	103	24	12	139
Critically Endangered (CR)	30	3	6	39
Data Deficient (DD)	21	1	5	27
Not Evaluated (NE)	11			11
Endangered (EN)	4	2		6
Vulnerable (VU)	4			4
Total	428	106	65	599

4 | DISCUSSION

Our results show that almost half of the endemic Lesser Antillean reptile species are found on sale or on buying advertisements online. The fact that trade involves species at high risk of extinction or that is affected by the trade is noteworthy. Similar results have been found for other taxonomic groups and regions (e.g. Fink et al., 2021). Under CITES Appendix I, species bred in captivity can be traded according to the appropriate resolutions. Still, considering the varying difficulty in breeding reptile species in captivity, it is likely that some of these species are being sourced from the wild, calling for action to understand whether trade is threatening populations in the wild. For example, establishing legal captive breeding facilities might be done illegally by harvesting wild individuals to be used for breeding (see e.g. Wyatt et al., 2018). As information on the origin of the individuals for sale was not readily available in many advertisements, future studies should investigate more in depth whether individuals were sourced from the wild or not, and the legality of the trade and compliance with CITES regulation. Evidence, for example, suggests that wild-caught Critically Endangered turtles are found for sale online at higher market price (Sung & Fong, 2018). In the case of individuals sourced from the wild, the impacts of pet trade may act synergistically to other threats (e.g. habitat loss) already affecting the persistence of these species in the wild. The most traded *Iguana* species, the Lesser Antillean Iguana (*Iguana delicatissima*), for example, is already threatened by habitat loss and fragmentation as well as hybridization with the Green Iguana (*Iguana iguana*) (Pounder et al., 2020). The fact that Near Threatened and Data Deficient species are also involved in the trade calls for further assessments of the impact of trade on the conservation of these species.

Our results suggest that Europe, especially Germany, and the United States are the largest hubs of Lesser Antillean reptile online trade. The only other country where Lesser Antilles reptiles were found to be traded online outside North America and Europe is Japan. These results are in line with previous studies on reptile pet trade, including of Lesser Antillean reptile species (e.g. Altherr & Lameter, 2020; Bush et al., 2014; Noseworthy, 2017;

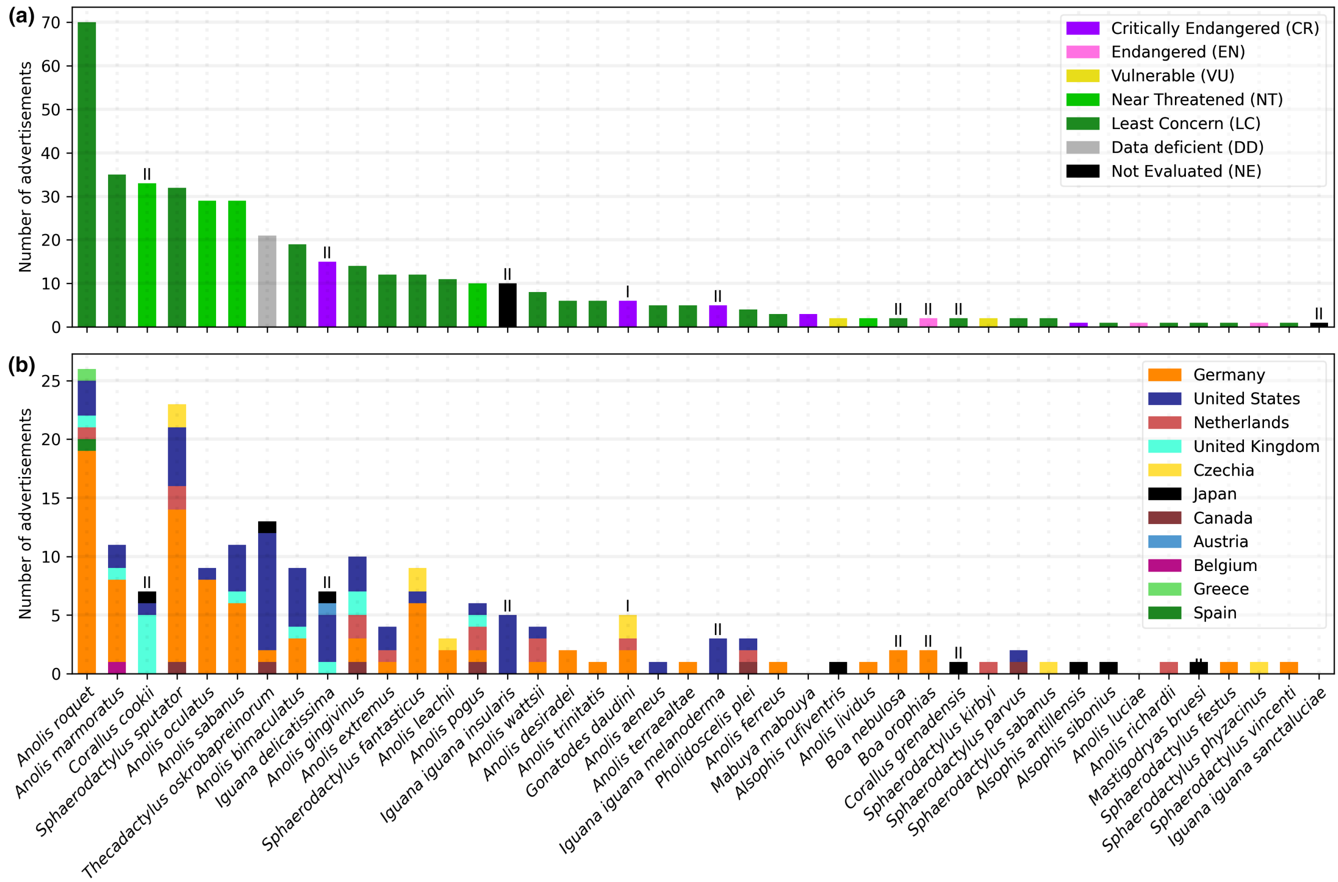

FIGURE 2 Number of sales advertisements found per Lesser Antillean species. (a) All sales advertisements. (b) Sales advertisements with location information categorized by the country of the location. The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) appendix number is reported on top of the bar of the CITES-listed species. *Iguana iguana insularis*, *Iguana iguana melanoderma* and *Iguana iguana sanctaluciae* are considered subspecies of *Iguana iguana*.

Robinson et al., 2015). The central role of Germany in the trade of these species was also noted in previous studies (e.g. Altherr & Lameter, 2020; Auliya et al., 2016; Jensen et al., 2019). Small lizard species, such as *Anolis* and *Sphaerodactylus* species, are particularly popular in Germany. One potential justification for this is the fact that these species are easier to breed in captivity, and it is possible that there is already an advanced infrastructure for breeding these reptiles in the country (Langner et al., 2022). Under this scenario, it is plausible that most individuals of these species featuring in trade originate from captive breeding populations. From reading the descriptions of sales advertisements, a recurring location for trading Lesser Antillean reptile species, both in German and English, was the Terraristika reptile exhibition in Hamm, Germany (see Altherr & Lameter, 2020). Other exhibitions were also mentioned, for example, in the United States and the United Kingdom. Looking into the role of these exhibitions in enhancing the popularity of reptile species in the wildlife trade would be an important future topic of study. While the Lacey Act prohibits the import of species that are protected in the exporting country and might hinder the trade of reptile species in the United States, the United States stands out in trade of CITES listed species as well as threatened species. For example, most of the instances of trade of *Iguana* species we detected originate from the United States. The United States are already known for being an important market for wild-collected Indonesian pet reptiles (Lyons & Natusch, 2011).

Research conducted in English is likely to exclude non-English-speaking regions from the assessment. In this study, we aimed to include non-English-speaking regions by searching for reptile trading websites in 12 different languages and including the names of the species in as many languages as possible. We acknowledge that the terms used in some languages might not have been the most used to advertise the species or instances of trade. Nevertheless, the abundance of German websites and inclusion of, for example, Japanese, Dutch and Czech websites indicate that to some extent, we were able to identify content from non-English-speaking countries. To avoid some limitations associated with language in efforts to monitor wildlife trade, future studies may consider taking advantage of recent developments in alternative approaches such as image analysis using machine learning methods (e.g. Cardoso et al., 2023; Kulkarni & Di Minin, 2023).

The total monetary value of the online trade of Lesser Antillean reptiles could not be fully estimated, as most advertisements did not report price information. Common species such as *Anolis* lizards are relatively cheap to buy and only a few rare *Iguana* species trade for more than a thousand Euros per individual. This

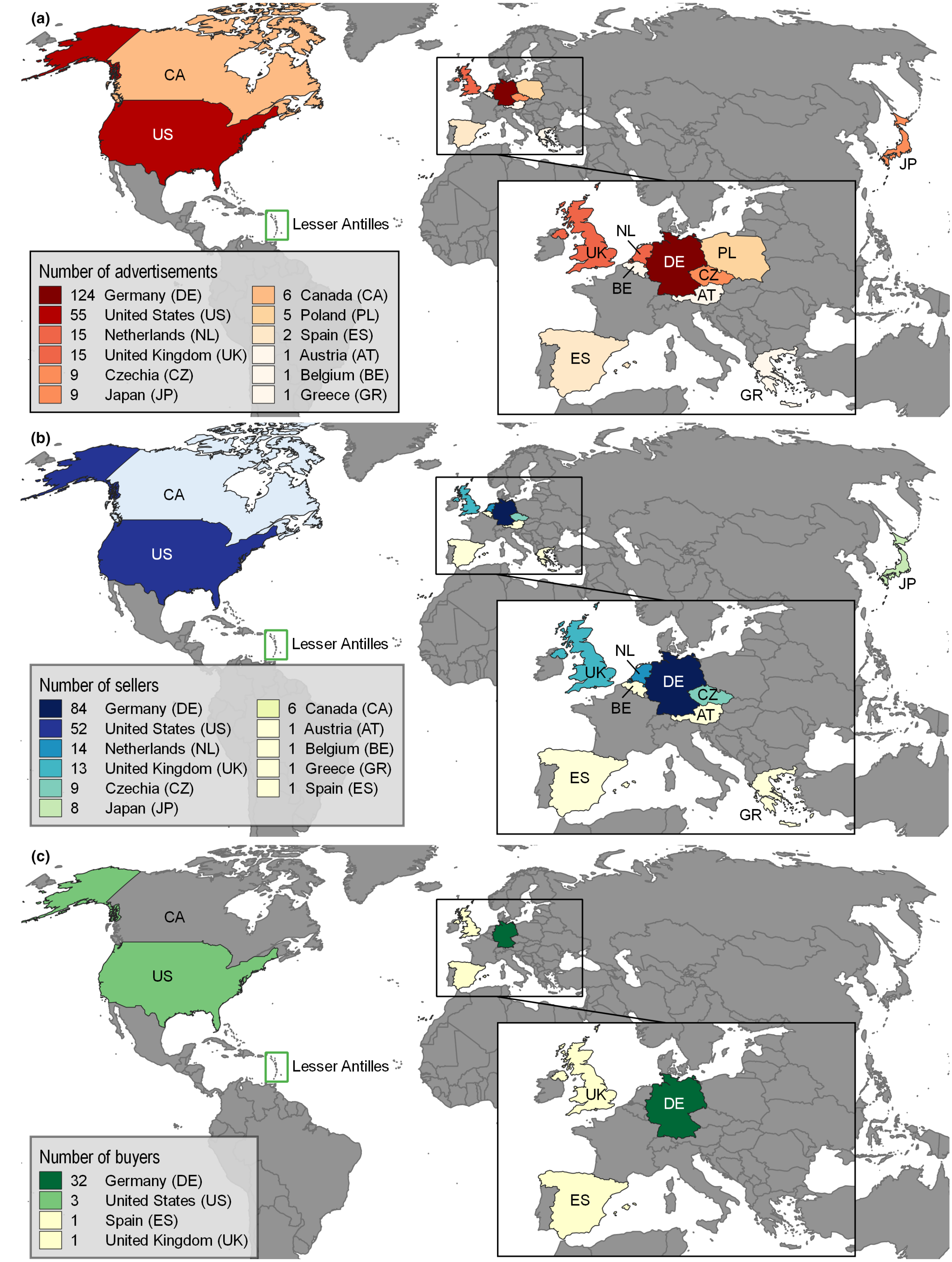

FIGURE 3 Number of Lesser Antillean reptile advertisements per country. a, all advertisements including selling, buying and other advertisements; b, sales advertisements; c, buying advertisements.

is expected as rarity drives the price up and could also be linked to the fact that these individuals may be taken from the wild as found in other studies (see e.g. Sung & Fong, 2018). Furthermore, it is also possible that the trade of rare species for which trade is forbidden or highly regulated may happen elsewhere and involve different prices. For example, Fink et al. (2021) observed that Indonesian bird species traded for significantly higher prices online when compared to physical markets.

The methods used in this study can be adapted to assess the trade of any list of species. However, the need for automation to assess the vast amount of data online limits the accuracy of the data. More manual surveying or more sophisticated methods, such as trained natural language processing algorithms using named entity recognition (e.g. Kulkarni & Di Minin, 2021), are needed for the data processing to yield more accurate results. Either way, more manual effort would be needed because a large amount of reviewed data are required to train machine learning algorithms. Especially, recognition of different elements from text would be required to assess the quantities and prices of the species studied, as most of the trade of these species is happening in online forums between pet owners. It is easier to classify data collected from, for example, websites where each data element has a distinct location on the website. One element of interest is of course whether individuals for sale are caught from the wild or bred in captivity. However, the origin of the reptiles was rarely stated in online advertisements and, therefore, it was difficult to assess with an automated script. Manual verification of this information may still be important for this purpose (see, e.g. Marshall et al., 2020). Nevertheless, even if the origin is stated in a sales advertisement, it may not correspond to the real one. Expanding our data pipeline to other digital platforms would also be important. Accessing social media data, for example, would be crucial for assessing the full scope of online reptile trade as commerce is also common on these sites (Demeau et al., 2019; Sung et al., 2021).

In conclusion, the proposed methods were used to investigate the magnitude and geographic distribution of online trade of 96 endemic and range-restricted reptile species from the Lesser Antilles. The same methods can be adapted to investigate the global extent of online trade in other taxonomic groups. Being able to study the geography of the online trade allows identification of countries where targeted monitoring efforts can help reveal the nature of the wildlife trade (legal and or illegal) and the source of individuals in the trade (wild caught and/or captive bred), especially in countries such as Germany and United States that appear to play a bigger role in the online trade of these species.

AUTHOR CONTRIBUTIONS

J.R. and E.D.M. developed the research ideas. J.R. developed the framework, collected the data, and performed the analysis with assistance from R.K. J.R. and E.D.M. led the writing with contributions from R.K., A.S.R, and R.C. E.D.M. obtained funding for the research.

ACKNOWLEDGEMENTS

J.R., R.K. and E.D.M. thank the European Research Council (ERC) for funding under the European Union's Horizon 2020 research and innovation programme (grant agreement 802933). R.K was supported by the KONE Foundation research grant (grant 202103830). A.S.R. was supported by a Marie Skłodowska-Curie Actions Postdoctoral Fellowship (grant agreement 101022521). R.A.C. acknowledges personal funding from the Academy of Finland (grant 348352) and from the KONE Foundation (grant 202101976). All authors would like to thank Isabel Vique at Fauna and Flora International.

CONFLICT OF INTEREST STATEMENT

None.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ddi.13771.

DATA AVAILABILITY STATEMENT

Data and code are available from https://etsin.fairdata.fi/dataset/fabe5192-e31e-494d-8eef-3da6f4faba00.

ORCID

Ricardo Correia https://orcid.org/0000-0001-7359-9091
Enrico Di Minin https://orcid.org/0000-0002-5562-318X

REFERENCES

Altherr, S., & Lameter, K. (2020). The rush for the rare: Reptiles and amphibians in the European pet trade. *Animals*, *10*(11), 2085.

Auliya, M., Altherr, S., Ariano-Sanchez, D., Baard, E. H., Brown, C., Brown, R. M., Cantu, J. C., Gentile, G., Gildenhuys, P., Henningheim, E., Hintzmann, J., & Ziegler, T. (2016). Trade in live reptiles, its impact on wild populations, and the role of the European market. *Biological Conservation*, *204*, 103–119.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, *471*(7336), 51–57.

Betts, M. G., Wolf, C., Ripple, W. J., Phalan, B., Millers, K. A., Duarte, A., Butchart, S. H. M., & Levi, T. (2017). Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature*, *547*(7664), 441–444.

Breuil, M., Schikorski, D., Vuillaume, B., Krauss, U., Daltry, J. C., Gaymes, G., Gaymes, J., Lepais, O., Bech, N., Jelić, M., Becking, T., & Grandjean, F. (2022). *Iguana insularis* (Iguanidae) from the southern Lesser Antilles: An endemic lineage endangered by hybridization. *ZooKeys*, *1086*, 137–161. https://doi.org/10.3897/zookeys.1086.76079

Breuil, M., Schikorski, D., Vuillaume, B., Krauss, U., Morton, M. N., Corry, E., Bech, N., Jelić, M., & Grandjean, F. (2020). Painted black: *Iguana melanoderma* (Reptilia, Squamata, Iguanidae) a new melanistic endemic species from Saba and Montserrat islands (Lesser Antilles). *ZooKeys*, *926*, 95–131. https://doi.org/10.3897/zookeys.926.48679

Bush, E. R., Baker, S. E., & Macdonald, D. W. (2014). Global trade in exotic pets 2006-2012. *Conservation Biology*, *28*(3), 663–676.

Cardoso, A. S., Bryukhova, S., Renna, F., Reino, L., Xu, C., Xiao, Z., Correia, R., Di Minin, E., Ribeiro, J., & Vaz, A. S. (2023). Detecting wildlife trafficking in images from online platforms: A test case using deep learning with pangolin images. *Biological Conservation*, *279*, 109905.

Carriere, D. (2013). *Geocoder*. https://github.com/DenisCarriere/geocoder

Central Intelligence Agency. (2021). Field listing – Languages. *The world factbook*. Retrieved February 19, 2021, from https://www.cia.gov/the-world-factbook/field/languages/

CITES. (2023). *The CITES Appendices*. Retrieved February 24, 2023, from https://cites.org/eng/app/index.php

Courchamp, F., Angulo, E., Rivalan, P., Hall, R. J., Signoret, L., Bull, L., & Meinard, Y. (2006). Rarity value and species extinction: The anthropogenic Allee effect. *PLoS Biology*, *4*(12), 2405–2410.

Demeau, E., Vargas, M., & Jeffrey, K. (2019). Wildlife trafficking on the internet: a virtual market similar to drug trafficking? *Revista Criminalidad*, *61*(2), 101–112. https://www.policia.gov.co/file/205799/download?token=4fZC7sH_

Di Minin, E., Brooks, T. M., Toivonen, T., Butchart, S. H. M., Heikinheimo, V., Watson, J. E. M., Burgess, N. D., Challender, D. W. S., Goettsch, B., Jenk-ins, R., & Moilanen, A. (2019). Identifying global centers of unsustainable commercial harvesting of species. *Science Advances*, *5*(4), eaau2879.

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., & Kulkarni, R. (2021). How to address data privacy concerns when using social media data in conservation science. *Conservation Biology*, *35*(2), 437–446.

Di Minin, E., Fink, C., Hiippala, T., & Tenkanen, H. (2019). A framework for investigating illegal wildlife trade on social media with machine learning. *Conservation Biology*, *33*(1), 210–213.

Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., & Mbaye, S. N. (2019). *Web scraping: State-of-the-art and areas of application*. 2019 IEEE international conference on big data (big data), 6040–6042.

Fauna & Flora international. (2021). *Fauna & flora international*. Retrieved November 18, 2021, from https://www.fauna-flora.org/

Fink, C., Toivonen, T., Correia, R. A., & Di Minin, E. (2021). Mapping the online songbird trade in Indonesia. *Applied Geography*, *134*, 102505.

Google Trends. (2021). *Google trends*. Retrieved February 19, 2021, from https://trends.google.com/trends/

Hedges, S. B. (2006). Paleogeography of the Antilles and origin of west Indian terrestrial vertebrates. *Annals of the Missouri Botanical Garden*, *93*(2), 231–244.

Hedges, S. B. (2020). *Caribherp*. Retrieved October 12, 2020, from http://www.caribherp.org/

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing*. https://spacy.io/

IPBES. (2019). Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services. *Population and Development Review*, *45*(3), 680–681. https://doi.org/10.1111/padr.12283

IUCN. (2023). *The IUCN Red List of Threatened Species*. Retrieved January 27, 2023, from https://www.iucnredlist.org

Jensen, T. J., Auliya, M., Burgess, N. D., Aust, P. W., Pertoldi, C., & Strand, J. (2019). Exploring the international trade in African snakes not listed on CITES: Highlighting the role of the internet and social media. *Biodiversity and Conservation*, *28*(1), 1–19.

Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Lei, G. C., & Wilmshurst, J. M. (2017). Biodiversity losses and conservation responses in the Anthropocene. *Science*, *356*(6335), 270–274.

Kouzis-Loukas, D. (2016). *Learning Scrapy: Learn the art of efficient web scraping and crawling with python*. Packt Publishing.

Kulkarni, R., & Di Minin, E. (2021). Automated retrieval of information on threatened species from online sources using machine learning. *Methods in Ecology and Evolution*, *12*(7), 1226–1239.

Kulkarni, R., & Di Minin, E. (2023). Towards automatic detection of wildlife trade using machine vision models. *Biological Conservation*, *279*, 109924.

Langner, C., Pfau, B., Bernardes, M., Gerlach, U., Felix, H., van Schingen-Khan, M., Schepp, U., Arranz, C., Riedling, M., & Kwet, A. (2022). *Evaluation of the captive breeding potential of selected amphibian and reptile taxa included in appendices I and II at CITES CoP18 evaluation of the captive breeding potential of selected amphibian and reptile taxa included in appendices I and II at CITES CoP18 results of the identical F+E-project (FKZ: 3520 53 2054)*.

Lavorgna, A. (2014). Wildlife trafficking in the internet age. *Crime Science*, *3*(1), 5.

Lyons, J. A., & Natusch, D. J. D. (2011). Wildlife laundering through breeding farms: Illegal harvest, population declines and a means of regulating the trade of green pythons (*Morelia viridis*) from Indonesia. *Biological Conservation*, *144*(12), 3073–3081. https://doi.org/10.1016/j.biocon.2011.10.002

Mantyka-Pringle, C. S., Martin, T. G., & Rhodes, J. R. (2012). Interactions between climate and habitat loss effects on biodiversity: A systematic review and metaanalysis. *Global Change Biology*, *18*(4), 1239–1252.

Marshall, B. M., Strine, C., & Hughes, A. C. (2020). Thousands of reptile species threatened by under-regulated global trade. *Nature Communications*, *11*(1), 4738.

Maxwell, S., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. (2016). The ravages of guns, nets and bulldozers. *Nature*, *536*(7615), 143–145.

Micro Pyramid Informatic Pvt. Ltd. (2016). *Forex Python*. https://pypi.org/project/forexpython/

Noseworthy, J. (2017). Cold-blooded conflict: Tackling the illegal trade in endemic Caribbean Island reptiles. In *Unpublished master's thesis, masters of philosophy in conservation leadership*. University of Cambridge.

OpenStreetMap contributors. (2015). *Planet dump*. https://planet.openstreetmap.org

Pounder, K. C., Mukhida, F., Brown, R. P., Carter, D., Daltry, J. C., Fleming, T., & Soanes, L. M. (2020). Testing for hybridisation of the critically endangered *Iguana delicatissima* on Anguilla to inform conservation efforts. *Conservation Genetics*, *21*(3), 405–420.

QGIS Development Team. (2019). QGIS geographic information system. Open Source Geospatial Foundation. http://qgis.org

Robinson, J. E., Griffiths, R. A., St John, F. A., & Roberts, D. L. (2015). Dynamics of the global trade in live reptiles: Shifting trends in production and consequences for sustainability. *Biological Conservation*, *184*, 42–50.

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., Meiri, S., Bohm, 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., … Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology & Evolution*, *1*(11), 1785.

Scheffers, B. R., Oliveira, B. F., Lamb, I., & Edwards, D. P. (2019). Global wildlife trade across the tree of life. *Science*, *366*(6461), 71–76.

Shepherd, C. R., Janssen, J., & Noseworthy, J. (2019). A case for listing the Union Island gecko Gonatodes daudini in the appendices of CITES. *Global Ecology and Conservation*, *17*, e00549.

Sung, Y.-H., & Fong, J. J. (2018). Assessing consumer trends and illegal activity by monitoring the online wildlife trade. *Biological Conservation*, *227*, 219–225. https://doi.org/10.1016/j.biocon.2018.09.025

Sung, Y. H., Lee, W. H., Leung, F. K. W., & Fong, J. J. (2021). Prevalence of illegal turtle trade on social media and implications for wildlife trade monitoring. *Biological Conservation*, *261*, 109245. https://doi.org/10.1016/j.biocon.2021.109245

Uetz, P., Freed, P., & Hošek, J. (2020). *The Reptile Database*. Retrieved December 17, 2020, from http://www.reptile-database.org/

UN Environment Programme. (1990). *Specially protected areas and wildlife protocol*. https://www.unep.org/cep/resources/policy-and-strategy/spaw-protocol-text

Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, *348*(6234), 571–573.

Wilson, B. S., Horrocks, J. A., & Hailey, A. (2011). Introduction. Conservation of insular Herpetofaunas in the West Indies. *Conservation of Caribbean Island Herpetofaunas*, *1*, 3–17.

Wyatt, T., Johnson, K., Hunter, L., George, R., & Gunter, R. (2018). Corruption and wildlife trafficking: Three case studies involving Asia. *Asian Journal of Criminology*, *13*(1), 35–55. https://doi.org/10.1007/s11417-017-9255-8

BIOSKETCH

The interdisciplinary team of scientists, including geographers, conservation scientists and machine learning scientists, from the Helsinki Lab of Interdisciplinary Conservation Science at the University of Helsinki and the Biodiversity Unit at the University of Turku has expertise on using digital data and methods to investigate people–nature interactions. Particularly, their research focuses on applying natural language processing and computer vision methods for automated digital content analysis.

SUPPORTING INFORMATION

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

How to cite this article: Rinne, J., Kulkarni, R., Soriano-Redondo, A., Correia, R., & Di Minin, E. (2023). Using automated content analysis to monitor global online trade in endemic reptile species. *Diversity and Distributions*, *00*, 1–10. https://doi.org/10.1111/ddi.13771