Received: 4 March 2021 | Revised: 6 August 2021 | Accepted: 13 August 2021

DOI: 10.1111/jbi.14256

Journal of Biogeography WILEY

METHODS

Bio-Dem, a tool to explore the relationship between biodiversity data availability and socio-political conditions in time and space

Alexander Zizka[1,2] | Oskar Rydén[3,4,5,6] | Daniel Edler[4,7,8] | Johannes Klein[4] | Allison Perrigo[4,7] | Daniele Silvestro[4,7,9,10] | Sverker C. Jagers[5,6] | Staffan I. Lindberg[3,6] | Alexandre Antonelli[4,7,11,12]

[1]German Center for Integrative Biodiversity Research Halle-Jena-Leipzig (iDiv), sDiv, University of Leipzig, Leipzig, Germany

[2]Naturalis Biodiversity Center, Leiden University, Leiden, The Netherlands

[3]Varieties of Democracy Institute, University of Gothenburg, Gothenburg, Sweden

[4]Gothenburg Global Biodiversity Centre, University of Gothenburg, Gothenburg, Sweden

[5]Centre for Collective Action Research, University of Gothenburg, Gothenburg, Sweden

[6]Department of Political Science, University of Gothenburg, Gothenburg, Sweden

[7]Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden

[8]Integrated Science Lab, Department of Physics, Umeå University, Umeå, Sweden

[9]Department of Biology, University of Fribourg, Fribourg, Switzerland

[10]Swiss Institute of Bioinformatics, Fribourg, Switzerland

[11]Department of Plant Sciences, University of Oxford, Oxford, UK

[12]Royal Botanic Gardens Kew, Richmond, UK

Correspondence
Alexander Zizka, German Center for Integrative Biodiversity Research Halle-Jena-Leipzig (iDiv), sDiv, University of Leipzig, Leipzig, Germany.
Email: alexander.zizka@idiv.de

Funding information
German Research Foundation, Grant/Award Number: DFG FZT 118; Centre for Collective Action Research; Riksbankens Jubileumsfond, Grant/Award Number: M13- 0559:1; Knut and Alice Wallenberg Foundation, Grant/Award Number: 2013.0166; European Research Council, Grant/Award Number: 724191; University of Gothenburg; Swedish Research Council, Grant/Award Number: 2019-05191, 2019-04739 and 2016-00796; Swedish Foundation for Strategic Research, Grant/Award Number: FFL15-0196; Royal Botanic Gardens; Swiss National Science

Abstract

Aim: Geo-referenced species occurrence records are a prime example of biodiversity data and a cornerstone of biodiversity research. Yet, their availability depends on non-biological factors, including the political framework in the region of collection. Here we present Bio-Dem (www.bio-dem.surge.sh), an open-source software to explore the relationship between the availability of species occurrence records and the political framework in countries worldwide and through time.

Location: Global.

Taxon: Any.

Methods: Bio-Dem accesses the number of occurrence records available from countries worldwide from the Global Biodiversity Information Facility (www.gbif.org) and socio-political information from these countries from the Varieties of Democracy database (www.v-dem.net) as well as information on colonial history from the Issue Correlates of War Project.

Staffan I Lindberg and Alexandre Antonelli shared last authors.

Foundation, Grant/Award Number: PCEFP3_187012 and FN-1749

Handling Editor: Ricardo Correia

Results: Bio-Dem is available as a free graphical user interface web application at www.bio-dem.surge.sh. Through three interactive graphs, Bio-Dem enables users to explore the relationship between species occurrence records and countries' regime type, freedom of movement, freedom of association, gross domestic product, education length, political corruption, armed conflict and colonial history. We describe possible links of these specific political indicators with the collection and mobilization of biodiversity data. Illustrative examples and video tutorials explaining the use of Bio-Dem are available at www.bio-dem.surge.sh.

Main conclusions: Bio-Dem provides a user-friendly way to explore the link between socio-political variables and the availability of species occurrence records in user-selected taxa and geographic regions, and to generate hypotheses on their connection. We envision Bio-Dem as a useful tool for teachers communicating the caveats of available biodiversity data, for biodiversity scientists exploring potential spatial and temporal bias in their data, and for social scientists exploring the impact of political systems on scientific data collection and biodiversity conservation.

KEYWORDS

autocracy, colonial history, colonialism, democracy, GBIF, herbarium specimens, museum specimens, occurrence record availability, regime type, sampling bias

1 | INTRODUCTION

The geographic distribution of biodiversity has fascinated humans for centuries. On a global scale, the scientific knowledge of species' geographic distributions is mainly based on natural history collections; traditionally as physical specimens in museums or herbaria, and more recently complemented by observations (e.g. vegetation plots or bird surveys), camera trap observations and citizen science data linked to photographs (e.g. www.inaturalist.org, www.floraincognita.com). The public availability of such biodiversity data has increased dramatically in the last decade, in good part due to large-scale data aggregators, such as the Global Biodiversity Information Facility (GBIF, www.gbif.org).

Geo-referenced species occurrence records provided by GBIF (and other aggregators) have become central for ecological and biodiversity research, yet they have important shortcomings and caveats. A major issue is that the availability of species occurrence records differs systematically among regions and taxa. These differences have multiple reasons, which, beyond biological explanations, also include biases associated with geographic accessibility (Engemann et al., 2015; Lin et al., 2015; Yang et al., 2014), political and socio-economic differences (Daru et al., 2018; Meyer et al., 2015; Rydén et al., 2020) and colonial history (Eichhorn et al., 2020). For instance, disproportionally more information is available from accessible regions close to roads and cities (Zizka et al., 2021), from countries with a high investment in research (Daru et al., 2018; Meyer et al., 2015), as well as open societies and regions with a stable political situation (Rydén et al., 2020). Further, in some regions it has been shown that scientists tend to repeatedly visit the same localities preferentially (e.g. field stations) instead of documenting novel areas, leading to a disparity in sampling frequency among regions (Farooq et al., 2020).

Geographic sampling bias is well recognized: methods exist to quantify it (e.g. Monsarrat et al., 2019; Ruete, 2015; Zizka et al., 2021) and account for it in biogeographic research (e.g. Aiello-Lammens et al., 2015; Phillips et al., 2009; Varela et al., 2014). In contrast, relatively little is known about socio-political biases in relation to sampling and the associated impact that countries' society, political regime and history may have on the availability of species occurrence records (Meyer et al., 2015). This is surprising because a considerable body of literature exists on the links between political regimes and their environmental performance—that is, a country's ambition to address environmental problems with adequate policy and implementation efficiency to generate positive outcomes for the environment (Neumayer, 2002; Povitkina, 2018; Rydén et al., 2020; Scruggs, 2003; Sjöstedt & Jagers, 2014). Environmental performance can be considered in two ways. It can be aggregated across several environmental issues, for example, climate change, deforestation and ocean pollution. Or it can focus on a particular issue, such as the collection of species occurrence records, which may represent the commitment to better understand and thereby protect flora and fauna.

Socio-political factors may impact the availability of species occurrence records either by affecting the capacity for record collection in a country, or by affecting governments' willingness to mobilize and publicly share data, especially in the context of intergovernmental organizations such as the GBIF (Lee & Zhang, 2016; Neumayer, 2002). Biases may exist between different countries but also between different time intervals within individual countries. A

general assumption is that certain types of regimes (e.g. open, safe, democratic) would favour the collection of species occurrence records whereas others (e.g. repressive, unsafe, authoritarian) would not (Amano & Sutherland, 2013). However, the empirical relationship seems to be more complex (Beeson, 2010; Rydén et al., 2020), since some authoritarian countries have a strong tradition in systematic biology (for example, former members of the Soviet Union). Furthermore, countries may differ substantially in distinct axes of political freedom, for instance, having a high level of freedom of movement, but very little freedom of expression, which may affect the collection and mobilization of biodiversity data differently (e.g. Brunei; Povitkina, 2018).

Specific political dimensions that may impact the collection and mobilization of species occurrence records include: regime type (democracies and autocracies), electoral democracy (the extent to which a regime elects their leaders through competitive multiparty elections), freedom of movement, the strength of civil society, levels of corruption, economic development, education, and the extent of physical violence and major armed conflicts (Amano & Sutherland, 2013; Coppedge, 2020; Meyer et al., 2015). The exploration of socio-political bias on the availability of species occurrence records is therefore an essential step in building and interpreting biodiversity knowledge.

Here we present Bio-Dem, a graphical user interface web application to visualize the availability of species occurrence records from GBIF (as proxy for scientific biodiversity knowledge) in relation to political systems and colonial history in countries around the world.

2 | MATERIALS AND METHODS

2.1 | Biodiversity records

Bio-Dem obtains the number of species occurrence records from GBIF in real time from the GBIF API. The data accessible via GBIF are contributed from institutions and citizen science projects around the world and include, among others, museum and herbarium collections and scientific and citizen science observation records.

2.2 | Socio-political variables

Bio-Dem obtains data on political indicators from the Varieties of Democracy (V-Dem) database (www.v-dem.net), currently in Version 10 (Coppedge, 2020). V-Dem is the world's largest database dedicated to the conceptualization and collection of data on democracy, comprising over 28 million data points and covering 201 countries and territories. V-Dem enables researchers to work with both aggregated higher-level concepts such as 'electoral democracy' and disaggregated concepts capturing, for example, freedom of academic expression. The database contains both factual data and expert-based measures using Bayesian IRT-modelling that accounts for coder bias (i.e. the subjectivity of individual researchers reporting the data) and uses this to adjust for heterogeneity across countries and years. The point estimates for each variable are based on the posterior medians from the measurement mode (Pemstein et al., 2020).

We included 13 socio-political variables in Bio-Dem, and only included the variables from V-Dem for which we postulated a conceptual impact on species occurrence record availability (see Table 1). Since socio-political variables are often correlated, and visualization of correlations as done in Bio-Dem may highlight spurious correlations, we emphasize that tangible mechanistic links need to be clearly described and that data exploration using Bio-Dem must be followed by rigorous statistical analyses (e.g. Keele, 2015). Below, we provided two empirical examples of putative mechanistic connections between selected socio-political variables and the availability of species occurrence records. See Table 1 for a summary of putative mechanistic links for all variables included in Bio-Dem and Data S1 for more detailed variable descriptions.

Regime type is a core variable summarizing the political system of countries (Lührmann et al., 2018). This variable is an ordinal classification of regimes as either a (1) closed autocracy, (2) electoral autocracy, (3) electoral democracy, or (4) liberal democracy. The first two categories are separated from the latter by evaluating how competitive the process of access to power is. A competitive process is characterized by free and fair multiparty elections with *de facto* consequences for the political leadership that are supported by political and civil liberties such as freedom of expression, right to alternative information and associational freedom (Munck, 2016, p. 11). In a second step, the autocratic categories are distinguished based on the existence of elections, where electoral autocracies hold *de jure* (legally recognized) elections while closed autocracies do not (e.g. Schedler, 2002). The democratic regimes are differentiated based on a liberal principle, characterized by access to justice regardless of gender, the so-called rule of law, and the presence of both legislative and judicial constraints on the executive branch, as well as an independent and competent bureaucracy. This distinction is based on a negative view of political power, where even if power is reached through democratic means, it needs checks and balances to prevent the tyranny of the majority. It also indicates that democracies need a certain level of capacity in order for their citizens to enjoy its benefits (e.g. Coppedge et al., 2011; Munck, 2016).

The political regime type is relevant for the generation and mobilization of biodiversity data since closed or electoral autocracies are expected to have fewer species occurrence records in relation to electoral or liberal democracies, mostly through accountability and preference aggregation (Povitkina, 2018; see also electoral democracy index below for more details). Furthermore, the rule of law and impartial bureaucracy make leaders more likely to honour international agreements and implement environmental policies more efficiently (Li & Reuveny, 2006; Sjöstedt, 2013). Lastly, liberal democratic regimes also allow for civil liberties, such as freedom of movement and physical integrity, which likely facilitate the collection of species occurrence records and biodiversity data in general.

TABLE 1 Socio-political variables included in Bio-Dem (Coppedge, 2020) with a short explanation and their postulated relevance for the collection of biodiversity data. See www.bio-dem.surge.sh and Data S1 for a more detailed description of each variable

Variable	Explanation	Relevance to biodiversity data
Civil society	This variable captures the extent to which a state allows civil society organizations (CSO) to freely organize and operate without being regulated or repressed.	Environmental performance is generally higher in countries with robust civil societies. CSOs may push governments to better manage biodiversity, provide infrastructure to collect biodiversity data or by collecting it themselves.
Conflict	Did a country experience either an internal armed conflict (at least 25 battle deaths) or a civil war (at least 1000 battle deaths per year) during a given year?	Decreased record collection by restricting accessibility, reducing capacity of domestic researchers and naturalists and disincentivize collection by international researchers. Furthermore, reduces resources for biodiversity research and data and degrade the environment itself, which indirectly can impede record collection.
Education 15+	The average level of education among citizens that are older than fifteen years old.	Higher levels of environmental awareness, more resources being allocated to educational and scientific purposes, and more sound environmental policy. In turn, higher information collection and larger allocation of resources to research.
Freedom of association	Can civil society organizations operate freely and without fear of repression from the government, and can oppositional political parties take form and participate in elections.	Higher likelihood of the existence of naturalist societies in countries where people are free to organize themselves, which can either collect data or facilitate data collection by campaigning and lobbying efforts, aimed at both the public and decision-makers. Furthermore, international environmental organizations can more easily operate in those countries.
Freedom of expression	This includes to what extent a country has freedom of expression and alternative sources of information do exist (e.g. press, media, academic and cultural freedom).	Higher level of interest in environmental issues and hence engagement for biodiversity data collection due to free information and discussion, including in cases where information collection is potentially in conflict with interests of the government.
Freedom of movement	To what extent enjoy citizens formal freedom of movement and residence? The variable gauges aspects such as the right for men and women to move at all times of the day, across regions using public thoroughfares	Restrictions on travel and residence may limit accessibility for collection, and disincentivize international researchers.
GDP per capita	The gross domestic product (GDP) per capita in a country.	Higher allocation of resources to research and education and hence Thus, higher capacity and infrastructure for biodiversity data collection and mobilization. Furthermore, more development can lead to a higher interest in environmental issues and hence data collection.
Physical violence index	To what extent a regime respects citizen's physical integrity? This variable captures to what degree torture and killings for political purposes are carried out by the government.	Low respect for citizens' physical integrity may decrease the capacity and motivation of domestic practitioners and scientists and disincentivize international researchers.
Political corruption	The level of political corruption in a country. This variable includes the extent of corruption within the public sector, the judiciary, the executive and the legislative bodies. As such, this captures different forms of corruption, operating at both grand and petty levels, and aimed at affecting both political decision-making and the implementation of policy.	High corruption draws resources from conservation and research and may hence hamper the collection of biodiversity data, especially since data collection depends on sufficient economic resources and support from public policy. Further, corruption may cause safety reasons, for instance by facilitating high levels of poaching.
Polyarchy (electoral democracy)	To what degree does a regime selects is executive and legislative through popular elections, whether these elections are free and fair, and how widespread the right to vote is.	Countries that are more democratic and have a higher accountability by election can be expected to have higher levels of environmental commitments, including the provision of infrastructure for the collection of information on biodiversity data

(Continues)

TABLE 1 (Continued)

Variable	Explanation	Relevance to biodiversity data
Protected areas	The share of total land area that is put under conservational status, as defined by the IUCN. It specifically captures terrestrial protected areas.	Protected areas of comprise charismatic habitats and species which may trigger collection, and protected areas often provide necessary infrastructure for data collection. Furthermore, the amount of protected area may be a surrogate for a countries commitment to environment and biodiversity protection in general, which is related to higher investment in data collection.
Regime type	How democratic is a regime? This variable evaluates regimes according to two principles: how competitive is the process of access to power (free and fair multiparty elections supported by political and civil liberties such as freedom of expression, right to alternative information and associational freedom) and how liberal democratic a regime is (rule of law and legislative and judicial constraints on the executive).	More democratic and liberal regimes may be more suitable for data collection (due to personal and institutional freedom, as well as resources available) and more likely to share data in international frameworks.
Time since independence	The time since independence from the last colonizing country. Only for colonial ties outside Europe and for connection of countries on different continents.	The longer a country is independent, the more time there may be for establishment of domestic scientific institutions and the lower the on-going ties of scientific dependency to the former colonizing country may be.

Categorization of political regimes through regime type is convenient because it summarizes many aspects of democracy (e.g. Schmitter & Karl, 1991). However, more specific indices may be more informative, for instance, the electoral democracy index, which is theoretically derived from Dahl's (1977) conception of polyarchy—rule by the many. The electoral democracy index quantifies to which degree a regime elects its executive and legislative bodies through popular multiparty elections, whether the elections are free and fair, and how widespread the right to vote is. It also measures the presence of freedom of expression, the right to obtain information from diverse sources, and the freedom of people to organize themselves freely in political parties and civil society organizations. This index ranges from 0 to 1, where higher scores indicate more electoral democracy (Coppedge, 2020). The electoral democracy index is relevant for the availability of species occurrence records since countries that are more democratic are expected to have higher levels of environmental commitments (e.g. Neumayer, 2002). These commitments can include, for example, financing and supplying infrastructure for the collection of information on environmental issues, such as biodiversity measurements.

Democracies can also be expected to have higher levels of commitment to biodiversity-related issues due to freedom of expression, freedom of association and electoral accountability. One reason to expect this is that as democracies put fewer restrictions on information flows, citizens' knowledge of environmental issues should increase. As such, if environmental issues, such as biodiversity decline, are more salient and discussed among influential segments of society, then political leaders can have stronger incentives for investing in, for example, biodiversity conservation (Morrow et al., 2008). However, if political leaders disregard such pressures, then voters and civil society organizations can, in return, put pressure on them for disregarding their interests, either through electoral accountability or informational campaigns (Li & Reuveny, 2006). Through this process, conventional or 'green' parties can find support for prioritizing the environment. As democracies allow for civil society organizations to exist to a larger extent, this can have positive effects on collecting species occurrence records, because domestic or international organizations can carry out data collection on their own initiative. Lastly, many of the hypothesized links are procedural and several of the evoked arguments have a 'negative' counterpart, meaning that aspects of democracy can contribute negatively to environmental performance (Rydén et al., 2020), giving rise to conditional relationships (Western, 1998).

While regime type and the electoral democracy index may correlate with the availability and the collection of species occurrence records on multiple levels, these effects can be indirect. In contrast, the physical violence index and the occurrence of conflict in a country quantify more direct obstacles to the availability of species occurrence records (e.g. Midlarsky, 1998). If countries have relatively low respect for citizens' physical integrity, one can expect this to have a negative effect on practitioners' and scientists' motivation and ability to travel within or into these countries. Hence, in countries that are experiencing conflicts, one can expect the collection of species occurrence records to be lower. In Bio-Dem, we included information on conflict from the UCDP/PRIO armed conflict dataset (Gleditsch et al., 2002; Pettersson & Öberg, 2020). In the dataset, armed conflict is defined as 'a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in a calendar year' (Pettersson, 2020). There are two intensity levels: (1) war (at least 1000 battle related deaths per year), and (2) minor conflict (25 – 999 battle related deaths per year).

2.3 | Colonial history

In addition to socio-political data, we included information on colonial history into BIO-DEM. There is a pervasive impact of historic European colonialism on scientific research across disciplines (Boshoff, 2009; Dahdouh-Guebas et al., 2003; Keller, 2006). Recently the discussion on systemic inequalities in science in general, and biodiversity research in particular (Erondu et al., 2021; Trisos et al., 2021; Vorontsova et al., 2021), has become a wider dialog. Particularly, the origins of modern biogeographic research (Eichhorn et al., 2020) as well as tropical systematics (Figueiredo & Smith, 2010) and ecology (Baker et al., 2019) are tightly connected to colonialism. Illustrative examples include the direct connection of the historic European 'exploration' of tropical ecosystems with military expeditions and explorative missions (for example, Darwin's travels on the Beagle), as well as the knowledge bias stemming from the dominance of Western scientific practice and language, for instance, the origins of modern taxonomy (including scientific species names based on Latin grammar and letters). This relevance is strengthened by research suggesting that colonial heritage affects contemporary deforestation rates (Ellis et al., 2021; Marchand, 2016). In more general terms, colonial history is argued to have long-lasting effects on political institutions (e.g. Herbst, 2000).

We obtained information on former colonial ties and dates of independence from the Issue Correlates of War Project colonial history dataset (Hensel & Mitchell, 2007), and included this data into BIO-DEM to enable researchers to explore the geographic patterns of European colonial history in relation to the availability of species occurrence records worldwide. We excluded colonial relationships that either ended before 1800, occurred within Europe or existed between neighbouring countries. Additionally, for practical reasons, we only included the most recent colonial relationship when the colonizing country has changed over time. In total, we included information for 114 formerly colonized countries.

2.4 | BIO-DEM code and description

BIO-DEM is accessible at www.bio-dem.surge.sh and the code is available under a MIT license at https://github.com/AntonelliLab/Bio-Dem. BIO-DEM is implemented in JavaScript and runs within the browser. The user interface is generated with the help of d3 for the interactive charts and React with Material-UI for the other components.

BIO-DEM provides a user-friendly interface, with three interactive graphs to visualize available biodiversity data in relation to countries' socio-political frameworks. First, the 'Biodiversity knowledge & political regimes' graph is an interactive scatterplot, where each data point represents a country and the x- and y-axes are user-defined. The axes can be chosen to represent the number of occurrence records, the number of occurrence records per area, or one of ten socio-political variables with a potential impact on record collection (see Figure 1a for an example). The colour of the data points can represent the political regime, the geographic region, GBIF membership status, or former colonial connections; the point size can represent either the total

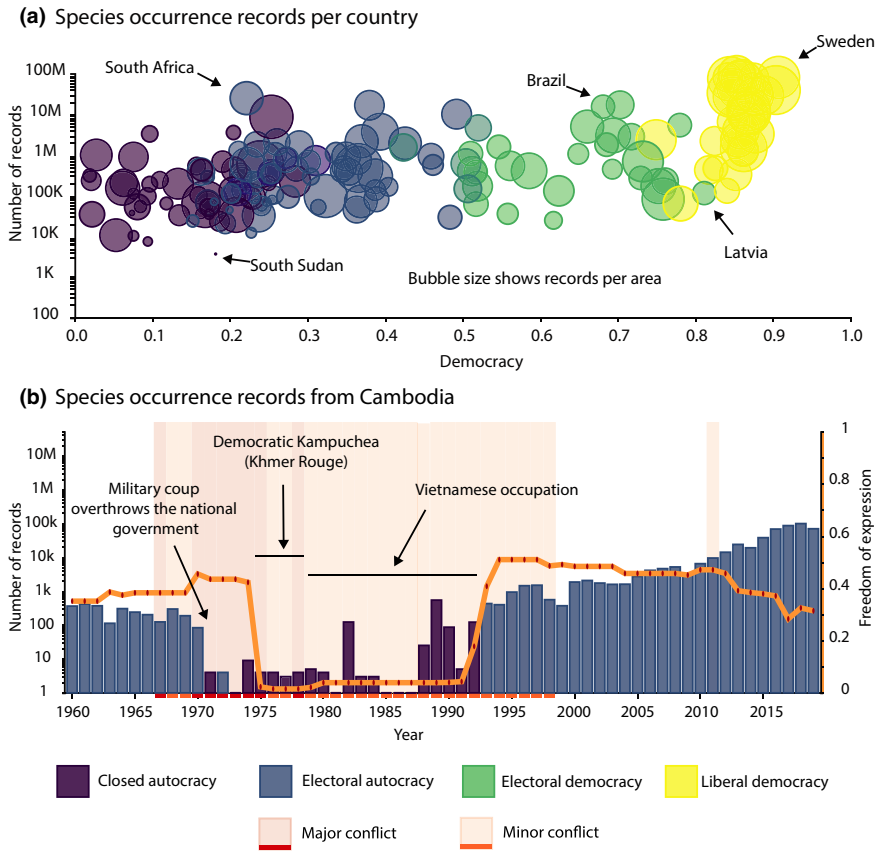

FIGURE 1 Number of species occurrence records reflects the state of political systems through space and time. Colours indicate regime type. (a) Occurrence records per country. There is no clear correlation between democracy type and number of records, but liberal democracies generally have more total records and more records available per area. (b) Occurrence records from Cambodia. In Cambodia, a period of political change and armed conflict between 1970 and 1992 is related to a decrease in biodiversity data available. Note the logarithmic scale for the y-axis related to the record number. Subfigure (a) shows interactive graph 1

number of occurrence records, or the number of occurrence records standardized by country area. Hovering over a data point displays additional information on the country. Illustrative examples with pre-set variables are available via a navigation bar. The graph can be exported with publication-level quality in svg format and the underlying data can be exported as a table in csv format for further analysis.

Second, the 'Biodiversity knowledge through time' graph combines a bar- and a line chart (see Figures 1b and 2 for examples). The bar chart shows the number of available occurrence records by year, from 1960 to 2019, for a user-defined country. The bars can be coloured by political regime of the respective year, the type of record (preserved specimen vs others) or the origin of the data contributor (domestic/former colonizing country/other; see Figure 2 for examples). The line chart shows the temporal development of a user-defined socio-political variable, which is potentially linked to record availability (Table 1). The y-axes of both charts are independent. Furthermore, years with local armed conflicts and the date of independence (for former colonies) are indicated, and users can choose to exclusively show records associated with a photograph or records based on preserved specimens. Users can also choose to filter records to a specific taxon on any taxonomic level, using text filters supported by auto-complete with the GBIF taxonomic backbone. The two interactive graphs are linked: marking a data point in the scatterplot triggers an automatic selection of the respective country in the second chart, showing the availability of records through time.

Third, the 'World map' graph is an interactive map visualizing the distribution of the included socio-political variables and the publishing countries of the species occurrence records (see Figure 3 for an example). By clicking on any country on the map, users can explore the publishing countries of species occurrence records from the selected country. In this case the colour scheme shows percent contribution, hovering over a country will display additional information, and the time bar can be used to limit the display to species occurrence records from specific time intervals between the years 1960 to 2019.

3 | EMPIRICAL EXAMPLES

We present two example cases illustrating the use of Bio-Dem for data exploration, and for hypothesis generation and testing, respectively. First, we used Bio-Dem to explore links between the political system (democratic v. autocratic), political turmoil and the availability of biodiversity data in space and time. Second, we used data exported from Bio-Dem to test predictions about the role of colonial ties in the availability of biodiversity data.

3.1 | Example 1: Exploring links between regime type and political turmoil and biodiversity data availability

The first example visualized the number of records per country in relation to the electoral democracy index, the number of records per area, and the regime type. This example suggested a relationship between regime type and record availability, with the highest number of records available from liberal democracies and a general increase in records with increases in the electoral democracy index (Figure 1a). South Africa and Brazil (high number of records at a relatively low electoral democracy index) as well as South Sudan and Latvia (unexpectedly low number of records) emerged as countries of particular interest.

In addition to visualizing net trends, Bio-Dem also enables the exploration of how regime change, as well as armed conflict, affect the availability of primary biodiversity data through time. For instance, the number of species occurrence records available from Cambodia from 1960 to 2019 suggests a relationship with the electoral democracy index, regime type, and political instability (Figure 1b). We found a decrease in record availability by two orders of magnitude in the 1970s, coinciding with the beginning of a period of conflict and autocratic consolidation (Lührmann & Lindberg, 2019). The end of this period and the corresponding increase in the level of democracy led to a sudden increase in record availability. Despite historical turmoil and a minor recent decline in the electoral democracy index, Cambodia mirrors most other countries in exhibiting a general increase in available biodiversity data. Although we are not suggesting any causal links, the recent democratic backslides in large and biodiverse countries (e.g. Brazil) emphasize the importance of potential connections between democracy and biodiversity exploration from a conservation perspective as well (e.g. Alizada et al., 2021). This is because the allocation of resources for conservation focuses on presumably biodiverse countries and regions, as inferred from species distribution records, and often without explicitly accounting for the socio-political bias we report here.

3.2 | Example 2: Hypotheses relating to the impact of colonial history on the availability of biodiversity data

While the colonial origins of biogeography and ecology and the remaining legacy of this history have been pointed out (Eichhorn et al., 2020), it is not clear to what extent former colonial ties continue to impact the availability of biodiversity data today. Here, to illustrate the use of Bio-Dem for hypothesis testing, we test two hypotheses against a null expectation:

- Hypothesis 1 (H_1): The proportion of species occurrence records contributed to GBIF from domestic sources in formerly colonized African countries *increased* since independence.
- Hypothesis 2 (H_2): The proportion of herbarium and museum specimens collected in formerly colonized African countries and contributed to GBIF from the last colonizing country *decreased* since independence.
- Null hypothesis (H_0): The proportion of herbarium and museum specimens available in GBIF did not change since independence.

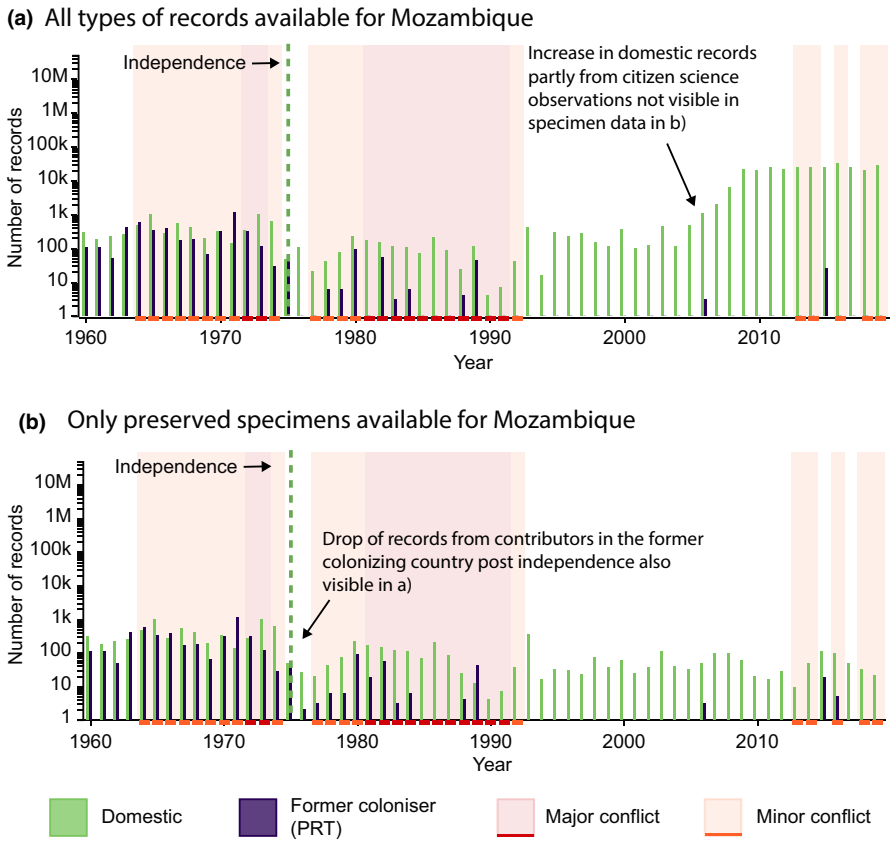

FIGURE 2 Number of species occurrence records from Mozambique through time, available from www.gbif.org. Records are split according to the country of the data contributor. (a) Records based on any type of species occurrence data. (b) Only records based on preserved specimens. Approximately 50% of the records available from 2006 and onward are from citizen science data (eBird & iNaturalist), the other major contributor is from fisheries. Only records contributed from Mozambique and Portugal are shown for better visibility. This figure shows a simplified version of interactive graph 2 of Bio-Dem

To test these hypotheses, we used Bio-Dem to obtain two data sets for 40 formerly colonized African countries that achieved independence after 1960 (when our data series starts). The first data set included the fraction of all species occurrence records per year after independence contributed from domestic sources for each country. The second data set included the fraction (per-country, per-year) of preserved herbarium and museum specimens contributed from sources in the respective former colonizing country (Figure S1). We then fitted two generalized linear mixed effects models using time since independence as fixed effect and country as random effect.

Specifically, we defined

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 x_{ij} + b_{0j} + b_{1j} x_{ij} + e_{ij} \quad (1)$$

where $\log\left(\frac{p_{ij}}{1-p_{ij}}\right)$ is the log odds, with p_{ij} the probability that a species occurrence record from the j th country contributed to GBIF at time index i comes from a domestic source (model 1) or that a preserved specimen from the j th country contributed to GBIF at time index i comes from an institution in the former colonizing country (model 2). x_{ij} denotes the years since independence of the j th country at years since independence index i; b_{0j} and b_{1j} are the random intercept and random slope effects for the j th country; β_0 and β_1 are the overall fixed intercept and fixed slope; and e_{ij} is the random error (Tanaka & Hui, 2019). We assumed different scales for slope and intercept and correlated random intercept and slope within countries, but this remained independent across countries. Furthermore, we assumed mutual independence of random effects and random error, and normally and identically distributed (NID) effects, so that the distribution assumptions for the random effects are as follows:

$$\begin{bmatrix} b_{0j} \\ b_{1j} \end{bmatrix} \sim NID\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right) and\, e_{ij} \sim NID(0, \sigma^2) \quad (2)$$

We ran the analysis in R (R Core Team, 2021) using the glmer function of the lme4 package (Bates et al., 2015) and used additional functions from the countrycode (Arel-Bundock et al., 2018), readxl (Wickham & Bryan, 2019), sjPlot (Lüdecke, 2021), states (Beger, 2021), tidyverse (Wickham et al., 2019) and ggeffects (Lüdecke, 2018) packages.

We found a significant positive effect of time since independence on the proportion of domestic records ($p < 0.001$), with an estimated overall effect size of 0.077 (2.5% 0.058, 97.5% = 0.095) on the logit scale, which was equivalent to the odds ratio of a record being contributed from a domestic institution increasing 8% per year since independence (1.0795; 2.5% 1.0595, 97.5% = 1.1; Figure S2). We did not find a significant effect of years since independence on the proportion of specimens contributed to GBIF from institutions from the former colonizing countries, with a positive per group effect for 14 countries and a negative per group effect for 26 countries (Figure S3 in, e.g. Mozambique; Figure 2).

These results reject the null hypothesis (H_0) and the alternative hypothesis (H_2), and corroborate H_1. Hence, our results are

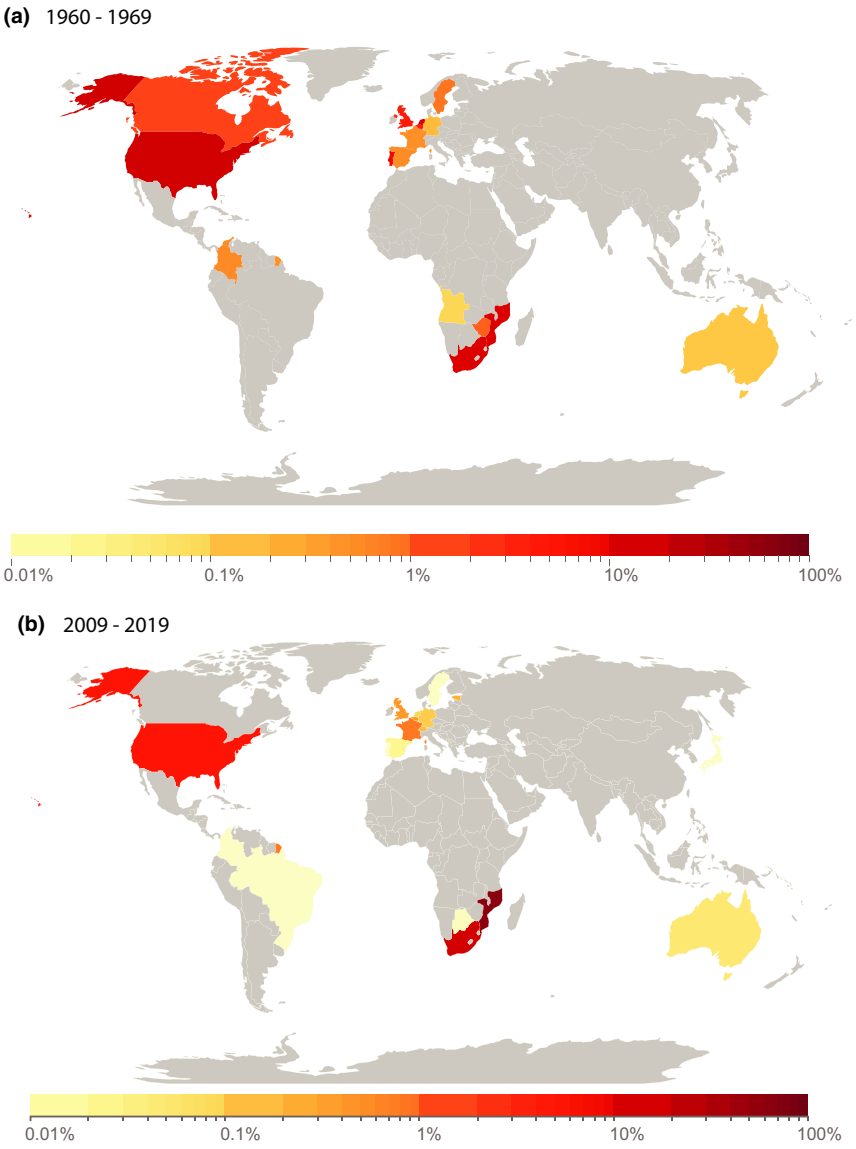

FIGURE 3 Number of Mozambican species occurrence records published from contributors in all countries worldwide, via GBIF. (a) Species occurrence records from 1960–1969; (b) Species occurrence records from 2009–2019. During these two periods, the proportion of domestically published records increased and the proportion of records published from European countries decreased. This figure shows interactive graph 3 of BIO-DEM

ambiguous on the lasting effect of colonial history on species occurrence record availability. The clear increase in the fraction of domestic records across countries suggests an increasing strength of domestic biodiversity research and a decreasing impact of former colonial ties. In contrast, the lack of a significant change of the fraction of preserved specimen data contributed from institutions in the former colonizing country since independence, and hence rejection of H_2, suggests that the effect of colonial history persists more strongly for preserved specimens. This was expected, because of the costly facilities needed to host specimens and the lengthy process of training scientific personal to curate scientific collections—neither of which are resources commonly available in newly-independent countries. Furthermore, the known inertia of taxonomic institutions likely contributed to this effect: since taxonomic collections and expertise of individual institutions and researchers are necessarily geographically and taxonomically limited, researchers and institutions in former colonizing countries often continue to collect material from former colonies for decades after independence, if possible (Figueiredo & Smith, 2010). Furthermore, while our results suggest no overall effect across countries, the large variance in per country slopes indicates that the effect differed among countries, and in multiple cases the proportion of preserved specimens available from the former colonizing country did decrease to almost zero (e.g. Mozambique, Figures 2 and 3).

We limited our analyses here to tests for correlation and simple time-dependence for illustrative purposes. However, analyses targeted specifically to understand the empirical effects of colonial history on species occurrence record collection would benefit from more complex and realistic models, including, for instance, a broader geographic scope, global temporal trends, and additional country-specific effects.

This example also illustrates how BIO-DEM may be used to generate novel research hypotheses. For instance, for multiple countries in our example data set, the number of species occurrence records has increased since the early 2000s (e.g. Mozambique; Figure 2). Further investigation on the origin of the data at www.gbif.org reveals that many of these records were contributed from citizen science project such as eBird and iNaturalist (data not shown). Since,

at least in theory, any citizen with a smartphone (amounting to over 3 billion globally in 2021) can contribute to these databases, a new hypothesis could be that citizen science data may override colonial legacies in data available from GBIF in comparison with countries never colonized. This hypothesis could be tested with data exported from Bio-Dem using the following predictions: (1) the number of domestic records per year has increased since smartphones became widespread; (2) the majority of records contributed are from citizen science data contributors.

4 | DISCUSSION

The availability of biodiversity data from public repositories varies due to the socio-political setting in countries worldwide and due to pervasive effects of colonial history, which may bias analysis in biodiversity research. Yet, this bias is often neglected. Bio-Dem is an easily accessible web application to visually explore potential temporal connections between the socio-political situation in countries worldwide and the availability of biodiversity data, namely species occurrence records. We demonstrated the use of Bio-Dem to identify patterns worth further exploration, and to generate and test research hypotheses.

Correlation does not imply causation and we acknowledge that observed patterns might result from chance or confounding variables. To minimize the impact of spurious correlations, we have only included socio-political variables in Bio-Dem for which we postulate a direct or indirect causal connection with the collection of species occurrence records (Table 1, Data S1). Yet, a causal interpretation of the observed patterns is difficult, due to indirect or unclear mechanisms, and the large number of potential factors involved (Rydén et al., 2020). The relationship between socio-political variables and biodiversity data availability is multi-faceted, and likely also multi-directional (Rydén et al., 2020). In addition, political variables are often correlated and disentangling their effects may be challenging or impossible given the limited number of data points (ultimately the number of countries). Furthermore, countries' overall biodiversity and endemism may impact record availability since high diversity and endemism may motivate collections, and sophisticated methodological choices may be necessary to disentangle these factors (Rydén et al., 2020). For these reasons, we refrain from including correlation analyses, such as regression statistics, into Bio-Dem since we think that they, in most cases, may be overly simplistic and easily over-interpreted.

We consider Bio-Dem primarily to be a tool for data exploration and communication. Researchers may use Bio-Dem to explore the potential impact of political variables on the availability of species occurrence records and to identify interesting patterns or outliers that may spark further research. Additionally, we think that Bio-Dem is well-suited to communicate and teach potential caveats of species occurrence data, which is now used ubiquitously in biogeographic and (macro-)ecological research and teaching. Future developments could include the exploration of further links to biodiversity data collection, such as the impact of the COVID-19 pandemic or extreme weather events (e.g. years of unusual drought, or the 2019/2020 widespread fires in Australia and Brazil). We provide the option to export the data used for visualization as tables for user-designed statistical analyses, and illustrated how researchers may use this data to test specific hypotheses in one of our empirical examples. For instance, the analysis of country data and time series could be exported from Bio-Dem and used to determine causal links in a suitable framework, for example, based on causal models (Pearl, 2018). When used with these caveats in mind, we expect that Bio-Dem will be a thought-provoking visual tool to explore spatial biodiversity data availability.

ACKNOWLEDGEMENTS

The authors thank all data collectors and contributors to GBIF and V-Dem. The authors also thank Sonia Dhanda for helpful feedback on the text about colonial history. No permits were obtained for this research, since no data collection was performed. The research presented in this paper is a contribution to the strategic research area Biodiversity and Ecosystem services in a Changing Climate (BECC), and funded in part by the BECC Action Group 'V-Bio: Assessing the link between biodiversity and democracy'. Additional funding was provided by iDiv/sDiv via the German Research Foundation (DFG FZT 118) to AZ; the Centre for Collective Action Research to OR; Riksbankens Jubileumsfond (Grant M13- 0559:1), Knut and Alice Wallenberg Foundation (Grant 2013.0166) and the European Research Council (Grant 724191) to SIL; internal grants from the Vice-Chancellor's office, the Dean of the College of Social Sciences, and the Department of Political Science at the University of Gothenburg to SIL and SJ; the Swedish Research Council (2019-05191), the Swedish Foundation for Strategic Research (FFL15-0196) and the Royal Botanic Gardens, Kew to AA. DS received funding from the Swiss National Science Foundation (PCEFP3_187012, FN-1749) and from the Swedish Research Council (VR: 2019-04739). DE was supported by the Swedish Research Council (2016-00796).

CONFLICT OF INTEREST

We declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data used in Bio-Dem are publicly available from www.gbif.org (species occurrence records) and www.v-dem.net (political variables). The code for Bio-Dem is accessible under a MIT license at github.com/AntonelliLab/Bio-Dem and video tutorials on the use of Bio-Dem are available at www.bio-dem.surge.sh. The data and analysis code for the empirical example are available at a zenodo repository (https://doi.org/10.5281/zenodo.5166809).

ORCID

Alexander Zizka https://orcid.org/0000-0002-1680-9192
Oskar Rydén https://orcid.org/0000-0002-8717-3574
Allison Perrigo https://orcid.org/0000-0002-6565-6305
Daniele Silvestro https://orcid.org/0000-0003-0100-0961

Sverker C. Jagers https://orcid.org/0000-0001-5491-8819
Staffan I. Lindberg https://orcid.org/0000-0003-0386-7390
Alexandre Antonelli https://orcid.org/0000-0003-1842-9297

REFERENCES

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5), 541–545. https://doi.org/10.1111/ecog.01132

Alizada, N., Cole, R., Gastaldi, L., Grahn, S., Hellmeier, S., Kolvani, P., Lachapelle, J., Lührmann, A., Maerz, S. F., Pillai, S., & Lindberg, S. (2021). *Autocratization turns viral. Democracy report 2021*. University of Gothenburg: V-Dem Institute. https://www.v-dem.net/files/25/DR%202021.pdf

Amano, T., & Sutherland, W. J. (2013). Four barriers to the global understanding of biodiversity conservation: Wealth, language, geographical location and security. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1756), 20122649. https://doi.org/10.1098/rspb.2012.2649

Arel-Bundock, V., Enevoldse, N., & Yetman, C. J. (2018). countrycode: An R package to convert country names and country codes. *Journal of Open Source Software*, *3*(28), 848–8489. https://doi.org/10.21105/joss.00848

Baker, K., Eichhorn, M. P., & Griffiths, M. (2019). Decolonizing field ecology. *Biotropica*, *51*(3), 288–292. https://doi.org/10.1111/btp.12663

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beeson, M. (2010). The coming of environmental authoritarianism. *Environmental Politics*, *19*(2), 276–294. https://doi.org/10.1080/09644010903576918

Beger, A. (2021). states: Create panels of independent states. R package version 0.3.1. https://CRAN.R-project.org/package=states

Boshoff, N. (2009). Neo-colonialism and research collaboration in Central Africa. *Scientometrics*, *81*(2), 413. https://doi.org/10.1007/s11192-008-2211-8

Coppedge, M. (2020). *V-Dem Dataset 2020*. Varieties of Democracy (V-Dem) Project. https://doi.org/10.23696/VDEMDS20

Coppedge, M., Gerring, J., Altman, D., Bernhard, M., Fish, S., Hicken, A., Kroenig, M., Lindberg, S. I., McMann, K., Paxton, P., Semetko, H. A., Skaaning, S. E., Staton, J., & Teorell, J. (2011). Conceptualizing and measuring democracy: A new approach. *Perspectives on Politics*, *9*(2), 247–267. https://doi.org/10.1017/S1537592711000880

Dahdouh-Guebas, F., Ahimbisibwe, J., Van Moll, R., & Koedam, N. (2003). Neo-colonial science by the most industrialised upon the least developed countries in peer-reviewed publishing. *Scientometrics*, *56*(3), 329–343. https://doi.org/10.1023/A:1022374703178

Dahl, R. A. (1977). *Polyarchy: Participation and Opposition* (Illustrated edition). Yale University Press.

Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfeld, T. J. S. S., Seidler, T. G., Sweeney, P. W., Foster, D. R., Ellison, A. M., & Davis, C. C. (2018). Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytologist*, *217*(2), 939–955. https://doi.org/10.1111/nph.14855

Eichhorn, M. P., Baker, K., & Griffiths, M. (2020). Steps towards decolonising biogeography. *Frontiers of Biogeography*, *12*(1). https://doi.org/10.21425/F5FBG44795

Ellis, E. C., Gauthier, N., Goldewijk, K. K., Bird, R. B., Boivin, N., Díaz, S., Fuller, D. Q., Gill, J. L., Kaplan, J. O., Kingston, N., Locke, H., McMichael, C. N. H., Ranco, D., Rick, T. C., Shaw, M. R., Stephens, L., Svenning, J.-C., & Watson, J. E. M. (2021). People have shaped most of terrestrial nature for at least 12,000 years. *Proceedings of the National Academy of Sciences*, *118*(17). https://doi.org/10.1073/pnas.2023483118

Engemann, K., Enquist, B. J., Sandel, B., Boyle, B., Jørgensen, P. M., Morueta-Holme, N., Peet, R. K., Violle, C., & Svenning, J.-C. (2015). Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution*, *5*(3), 807–820. https://doi.org/10.1002/ece3.1405

Erondu, N. A., Aniebo, I., Kyobutungi, C., Midega, J., Okiro, E., & Okumu, F. (2021). Open letter to international funders of science and development in Africa. *Nature Medicine*, *27*(5), 742–744. https://doi.org/10.1038/s41591-021-01307-8

Farooq, H., Azevedo, J. A. R., Soares, A., Antonelli, A., & Faurby, S. (2020). Mapping Africa's biodiversity: More of the same is just not good enough. *Systematic Biology*, syaa090. https://doi.org/10.1093/sysbio/syaa090

Figueiredo, E., & Smith, G. F. (2010). The colonial legacy in African plant taxonomy. *South African Journal of Science*, *106*(3–4), 1–4. https://doi.org/10.4102/sajs.v106i3/4.161

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed conflict 1946–2001: A new dataset. *Journal of Peace Research*, *39*(5), 615–637. https://doi.org/10.1177/0022343302039005007

Hensel, P. R., & Mitchell, S. M. (2007). The issue correlates of war (ICOW) project supplementary data set: Colonial history data set. *Harvard Dataverse*. https://doi.org/10.7910/DVN/5EMETG

Herbst, J. (2000). *States and power in Africa: Comparative lessons in authority and control* (STU-Student edition, 2). Princeton University Press. https://www.jstor.org/stable/j.ctt9qh05m

Keele, L. (2015). The statistics of causal inference: A view from political methodology. *Political Analysis*, *23*(3), 313–335. https://doi.org/10.1093/pan/mpv007

Keller, R. C. (2006). Geographies of power, legacies of mistrust: Colonial medicine in the global present. *Historical Geography*, *34*, 26–48.

Lee, M. M., & Zhang, N. (2016). Legibility and the informational foundations of state capacity. *The Journal of Politics*, *79*(1), 118–132. https://doi.org/10.1086/688053

Li, Q., & Reuveny, R. (2006). Democracy and environmental degradation. *International Studies Quarterly*, *50*(4), 935–956. https://doi.org/10.1111/j.1468-2478.2006.00432.x

Lin, Y.-P., Deng, D., Lin, W.-C., Lemmens, R., Crossman, N. D., Henle, K., & Schmeller, D. S. (2015). Uncertainty analysis of crowd-sourced and professionally collected field data used in species distribution models of Taiwanese moths. *Biological Conservation*, *181*, 102–110. https://doi.org/10.1016/j.biocon.2014.11.012

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Lüdecke, D. (2021). sjPlot: Data visualization for statistics in social science. R package version 2.8.9. https://CRAN.R-project.org/package=sjPlot

Lührmann, A., & Lindberg, S. I. (2019). A third wave of autocratization is here: What is new about it? *Democratization*, *26*(7), 1095–1113. https://doi.org/10.1080/13510347.2019.1582029

Lührmann, A., Tannenberg, M., & Lindberg, S. I. (2018). Regimes of the world (RoW): opening new avenues for the comparative study of political regimes. *Politics and Governance*, *6*(1), 60. https://doi.org/10.17645/pag.v6i1.1214

Marchand, S. (2016). The colonial origins of deforestation: An institutional analysis. *Environment and Development Economics*, *21*(3), 318–349. https://doi.org/10.1017/S1355770X1500025X

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, *6*(1), 8221. https://doi.org/10.1038/ncomms9221

Midlarsky, M. I. (1998). Democracy and the environment: An empirical assessment. *Journal of Peace Research*, *35*(3), 341–361. https://doi.org/10.1177/0022343398035003005

Monsarrat, S., Boshoff, A. F., & Kerley, G. I. H. (2019). Accessibility maps as a tool to predict sampling bias in historical biodiversity

occurrence records. *Ecography*, *42*(1), 125–136. https://doi.org/10.1111/ecog.03944

Morrow, J. D., De Mesquita, B. B., Siverson, R. M., & Smith, A. (2008). Retesting selectorate theory: Separating the effects of W from other elements of democracy. *American Political Science Review*, *102*(3), 393–400. https://doi.org/10.1017/S0003055408080295

Munck, G. L. (2016). What is democracy? A reconceptualization of the quality of democracy. *Democratization*, *23*(1), 1–26. https://doi.org/10.1080/13510347.2014.918104

Neumayer, E. (2002). Do democracies exhibit stronger international environmental commitment? A cross-country analysis. *Journal of Peace Research*, *39*(2), 139–164. https://doi.org/10.1177/0022343302039002001

Pearl, J. (2018). *Causality*. Cambridge University Press.

Pemstein, D., Marquardt, K. M., Tzelgov, E., Wang, Y., Medzihorsky, J., Krusell, J., Miri, F., & von Römer, J. (2020). *The V-Dem measurement model: Latent variable analysis for cross-national and cross-temporal expert-coded data* (SSRN Scholarly Paper ID 3595962). Social Science Research Network, https://doi.org/10.2139/ssrn.3595962

Pettersson, T. (2020). *UCDP/PRIO Armed Conflict Dataset Codebook v 20.1*. Published online. https://ucdp.uu.se/downloads/

Pettersson, T., & Öberg, M. (2020). Organized violence, 1989–2019. *Journal of Peace Research*, *57*(4), 597–613. https://doi.org/10.1177/0022343320934986

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, *19*(1), 181–197. https://doi.org/10.1890/07-2153.1

Povitkina, M. (2018). The limits of democracy in tackling climate change. *Environmental Politics*, *27*(3), 411–432. https://doi.org/10.1080/09644016.2018.1444723

R Core Team (2021). *R: A Language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Ruete, A. (2015). Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance maps. *Biodiversity Data Journal*, *3*, e5361. https://doi.org/10.3897/BDJ.3.e5361

Rydén, O., Zizka, A., Jagers, S. C., Lindberg, S. I., & Antonelli, A. (2020). Linking democracy and biodiversity conservation: Empirical evidence and research gaps. *Ambio*, *49*(2), 419–433. https://doi.org/10.1007/s13280-019-01210-0

Schedler, A. (2002). Elections without democracy: The menu of manipulation. *Journal of Democracy*, *13*(2), 36–50. https://doi.org/10.1353/jod.2002.0031

Schmitter, P. C., & Karl, T. L. (1991). What democracy is… and is not. *Journal of Democracy*, *2*(3), 75–88. https://doi.org/10.1353/jod.1991.0033

Scruggs, L. (2003). *Sustaining abundance: Environmental performance in industrial democracies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511615689

Sjöstedt, M. (2013). Horizontal and vertical resource dilemmas in natural resource management: The case of African fisheries. *Fish and Fisheries*, *14*(4), 616–624. https://doi.org/10.1111/j.1467-2979.2012.00481.x

Sjöstedt, M., & Jagers, S. C. (2014). Democracy and the environment revisited: The case of African fisheries. *Marine Policy*, *43*, 143–148. https://doi.org/10.1016/j.marpol.2013.05.007

Tanaka, E., & Hui, F. K. C. (2019). Symbolic formulae for linear mixed models. *ArXiv:1911.08628 [stat]*, *1150*, 3–21. https://doi.org/10.1007/978-981-15-1960-4_1

Trisos, C. H., Auerbach, J., & Katti, M. (2021). Decoloniality and anti-oppressive practices for a more ethical ecology. *Nature Ecology & Evolution*, https://doi.org/10.1038/s41559-021-01460-w

Varela, S., Anderson, R. P., García-Valdés, R., & Fernández-González, F. (2014). Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography*, *37*(11), 1084–1091. https://doi.org/10.1111/j.1600-0587.2013.00441.x

Vorontsova, M. S., Lowry, P. P., Andriambololonera, S. R., Wilmé, L., Rasolohery, A., Govaerts, R., Ficinski, S. Z., & Humphreys, A. M. (2021). Inequality in plant diversity knowledge and unrecorded plant extinctions: An example from the grasses of Madagascar. *PLANTS, PEOPLE, PLANET*, *3*(1), 45–60. https://doi.org/10.1002/ppp3.10123

Western, B. (1998). Causal heterogeneity in comparative research: A bayesian hierarchical modelling approach. *American Journal of Political Science*, *42*(4), 1233–1259. https://doi.org/10.2307/2991856

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2019). readxl: Read Excel Files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl

Yang, W., Ma, K., & Kreft, H. (2014). Environmental and socio-economic factors shaping the geography of floristic collections in China. *Global Ecology and Biogeography*, *23*(11), 1284–1292. https://doi.org/10.1111/geb.12225

Zizka, A., Antonelli, A., & Silvestro, D. (2021). Sampbias, a method for quantifying geographic sampling biases in species distribution data. *Ecography*, *44*(1), 25–32. https://doi.org/10.1111/ecog.05102

BIOSKETCH

AZ is a German biogeographer and botanist, developing methods to harness big datasets for biogeographic research. The development of Bio-Dem was triggered by the realization that the impact of socio-political factors and colonial history on biodiversity data availability is often underappreciated and needs to be clearly communicated. This realization resulted from various collaborative projects with researchers from institutions worldwide, and the differences in biodiversity data available for these projects depending on socio-political circumstances. AA is a Brazilian and Swedish biologist with a consolidated track record of collaborations with researchers from institutions world-wide. We acknowledge that our positions as researches at European institutions and of mostly European origin allow us access to funding, biodiversity data and analysis options not all researchers may have.

Author contribution: All authors designed this study. DE, JK, OR and AZ developed Bio-Dem. AZ and DS analysed the data. AZ wrote the manuscript with contributions from all authors.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zizka, A., Rydén, O., Edler, D., Klein, J., Perrigo, A., Silvestro, D., Jagers, S. C., Lindberg, S. I., & Antonelli, A. (2021). Bio-Dem, a tool to explore the relationship between biodiversity data availability and socio-political conditions in time and space. *Journal of Biogeography*, 48, 2715–2726. https://doi.org/10.1111/jbi.14256