Motivation: We generated a novel database of Neotropical snakes (one of the world’s richest herpetofauna) combining the most comprehensive, manually compiled distribution dataset with publicly available data. We assess, for the first time, the diversity patterns for all Neotropical snakes as well as sampling density and sampling biases.

Main types of variables contained: We compiled three databases of species occurrences: a dataset downloaded from the Global Biodiversity Information Facility (GBIF), a verified dataset built through taxonomic work and specialized literature, and a combined dataset comprising a cleaned version of the GBIF dataset merged with the verified dataset.

Spatial location and grain: Neotropics, Behrmann projection equivalent to 1° × 1°.

Time period: Specimens housed in museums during the last 150 years.
1 | INTRODUCTION

Reptiles are a highly diverse group of terrestrial vertebrates with 10,450 known species, with this number increasing at c. 100 per year (Tonini, Beard, Ferreira, Jetz, & Pyron, 2016; Uetz & Hošek, 2016). It is probably the most neglected group in conservation prioritizations, as only 52% of the described species have been assessed in the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (IUCN, 2017). Most of the assessed species have been categorized based on range size, of which 20% (Bland & Ohm, 2016; Böhmer, Collen, & Baillie, 2013) are considered data deficient owing to the lack of appropriate data on taxonomy, ecology, distribution, population trends and threats (Bland & Böhmer, 2016; Böhmer et al., 2013). This contrasts with that for other vertebrates, as for instance only 0.6% of birds and 15% of mammals are data deficient (Butchart & Bird, 2010; Schipper et al., 2008).

Among reptiles, there are c. 3,500 snake species globally, inhabiting temperate to tropical environments, in terrestrial and aquatic habitats (Uetz & Hošek, 2016; Wallach, Williams, & Boundy, 2014). As for most reptiles, distribution data for snake species remain scarce, and consequently, they are excluded from most large-scale studies of biodiversity and distribution patterns (e.g., Jenkins, Alves, Uezu, & Vale, 2015; Moura, Villalobos, Costa, & García, 2016). Although reliable estimates of snake diversity would contribute to global and regional strategies for biological conservation, no detailed data have yet been compiled for the Neotropics, despite it comprising one of the world’s richest herpetofaunas (Böhmer et al., 2013; Meiri & Chapple, 2016).

Here, we present a new database of snake occurrences covering the entire Neotropics and assess, for the first time, the diversity patterns for all Neotropical snakes as well as sampling artefacts. We hypothesize that snake diversity follows a similar pattern to those already described for other vertebrates in the Neotropics (Jenkins et al., 2015; Moura et al., 2016). We generated our novel database by combining the most comprehensive, manually compiled distribution dataset with publicly available data, from which we calculate species richness (SR) and phylogenetic diversity (PD) as well as sampling density and sampling biases. Finally, we discuss prospects for more informed conservation strategies and design research agendas.

2 | MATERIAL AND METHODS

2.1 | Data sources

We compiled three datasets for snakes recorded in the Neotropical region (sensu Olson et al., 2001), from central Mexico to southern South America, including all Caribbean islands. We included only records identified at the species level.

The raw dataset (RD) comprised georeferenced records for snakes downloaded from the Global Biodiversity Information Facility (GBIF; http://doi.org/10.15468/dlt.pbqv). We filtered our search for records linked to specimens, literature occurrences and material samples, leaving out records lacking associated vouchers.

The verified dataset (VD) comprised geographical occurrences from vouchered specimens examined in natural history museums (Supporting Information Appendix S1) and required a large collaborative effort among herpetologists. The initial focus of the VD was to gather data on Brazilian snakes but also including their distribution outside the country. This was then expanded also to include species and records from other Latin American countries outside Brazil, through point occurrence data from vouchers and scientific literature.

The combined dataset (CD) was constructed by merging a cleaned version of the RD with the VD. To produce the RD cleaned dataset, we taxonomically validated and updated the species names. Geographical coordinates were cleaned by verification of geographical data and map compilation using the speciesgeocodeR (Töpel et al., 2017) and maptools (Lewin-Koh et al., 2011) packages in R (R Core Team, 2017). A commented list containing all taxonomic and geographical changes
applied to the RD in this process (without voucher verification) is provided in Supporting Information Appendix S2. Then, we merged the GBIF cleaned dataset with the VD to form the CD. We also removed from CD all redundant coordinates for each species (i.e., records with identical latitude and longitude values).

2.2 | Analyses

2.2.1 | Species richness and phylogenetic diversity

We used the CD for species richness (SR) and phylogenetic diversity (PD) analyses. Both analyses were performed at two spatial resolutions: the grid cell scale, which was on an equal area Behrmann projection with 360 columns (corresponding to 1° × 1° at 30° N or 30° S, 1° × c. 0.75° at the equator and 1° × c. 9.5° at the poles), and the ecoregion scale corresponding to polygons (sensu Olson et al., 2001; Figure 1). We ran all analyses using the software Biodiverse, version 1.1 (Laffan, Lubarsky, & Rosauer, 2010).

Phylogenetic diversity and species richness are usually correlated (Morlon et al., 2011). However, SR takes into account only distribution data for each species, whereas PD is calculated by using distribution data plus branch lengths of the phylogeny. Thus, PD incorporates evolutionary history that is not expressed by SR (Faith, 1992, 2008; Tucker et al., 2017).

The PD analysis was based on distribution data and a sample of 100 trees, from which we calculated mean values for each grid cell and ecoregion.
TABLE 1  Number of species and amount of occurrence data in the three datasets of snakes recorded in the Neotropical region

| Dataset          | Number of occurrences | Number of species |
|------------------|-----------------------|-------------------|
| Raw dataset      | 7,299                 | 659               |
| Verified dataset | 140,368               | 488               |
| Combined dataset | 147,515               | 886               |

ecoregion polygons to synthesize the result of PD in a single map for each scale adopted. We used the phylogeny provided by Tonini et al. (2016). The variance in PD metrics across the sample of trees reported in their study was low; thus, we considered this a sufficient approximation of PD. Phylogenetic diversity analyses require a precise match between distribution data and the terminals of the phylogeny. Of the 886 species in the CD, 847 (96%) were present on the tree and were used for both SR and PD analyses to allow a more direct comparison between the two analyses.

2.2.2 | Sampling gaps
We calculated the number of occurrences across grid cells superimposed onto the Neotropical region to identify the intensity of sampling in each dataset.

3  | RESULTS

3.1 | Data availability
The RD includes 7,299 records of 659 species from 12 families (Table 1). The records are distributed over 25 countries (Figure 2a). A large number of records were derived from Central America and the West Indies, whereas the data are especially poor and sparse in South America (Figure 2a).

The VD contains almost 20 times more records than the RD. It includes 140,368 georeferenced records for 488 species from 10 families (Table 1). The records are distributed over 18 countries (Figure 2a), especially in South America, with 436 species recorded only in this region (Figure 2a).

We excluded 152 inconsistences in taxonomic and geographical cleaning of RD (Supporting Information Appendix S2). Thus, the CD has a total of 147,515 georeferenced records, representing 886 species in 12 families across 27 countries (Table 1 and Figure 2a). To our knowledge, this constitutes the most extensive and complete dataset of snake distributions for the Neotropical region, both in number of occurrences and number of species (Table 1 and Figure 2a). Maps for each species and a list of all Neotropical snake species included in this study, with their status of their known geographical distribution, are provided in Supporting Information Appendices S3 and S4.

3.2 | Spatial patterns of species richness and phylogenetic diversity

3.2.1 | Grid cell diversity
SR and PD are spatially very similar (Figure 2b,c). Species richness (60–120 species) and PD (1,000–2,000) are highest in the Atlantic Forest of southeast Brazil, closely followed by the Amazonian region along some large cities or close to important rivers, the coastal forests in northernmost South America, the Andean forests of Ecuador, the moist and montane forests of Central America, the Cerrado savannas in Central Brazilian Plateau and nearby Tocantins drainage, and the Pantanal wetlands. Intermediate values of SR (30–60 species) and PD (500–1,000) are found in the semi-arid Caatinga in northeast Brazil, in a large continuous (SR) or scattered (PD) area in the Cerrado savannas, the Pampas and Chaco regions in southern South America, in the Andean region over Colombia, Ecuador and Peru, and over a large portion in Central America.

For SR, values < 30 species are found in small portion of the Neotropical region (e.g., West Indies, north-western Amazonia, northern and southern Bolivia, and north of the Chaco). For PD, a large number of cells show low values (between zero and 500), with the largest patches in Amazonia.

3.2.2 | Ecoregion diversity
We recorded species in 187 ecoregions (Figure 2d,e; Supporting Information Appendix S5). The Cerrado is the richest ecoregion, with occurrences of 222 species of snakes, and also has the highest PD value (2,700). However, the ecoregions inside the Atlantic Forest domain also presented high values of SR and PD. Ecoregions in the Caatinga, the extra-Andean region from Colombia to Peru, Costa Rica, Chaco and Pampas have intermediate values of both SR and PD (Figure 2d,e; Supporting Information Appendix S5).

The ecoregions with lowest SR and PD are located in the southern part of the Andes and the West Indies islands. Thirty-one ecoregions include records for just one to three species each (Figure 2d,e; Supporting Information Appendix S5).

3.3 | Sampling gaps
Based on CD, the most poorly sampled Neotropical region is the Amazon, where all grid cells harbour < 500 records and 1,600,000 km² have no records at all (Figure 2a). The Andean region is also poorly sampled, with 900,000 km² empty and all others having < 500 records. The Lesser Antilles and Central America are also poorly sampled. The best-sampled region is the Atlantic Forest (400,000 km², containing 1,000–3,000 occurrences; Figure 2a). Some cells are well sampled, even though surrounding cells have very few records.

4  | DISCUSSION

4.1 | Data availability
We found errors associated with non-updated nomenclature and erroneous georeferences in the RD. This reinforces previous suggestions (e.g., Ficetola et al., 2013; Maldonado et al., 2015; Meyer, Weigelt, & Kreft, 2016) that GBIF data should not be used without proper verification and cleaning. The verified dataset, albeit smaller in the absolute number of species and records outside Brazil, can be considered well curated. As these two datasets are so different in geographical and taxonomic representation, merging them proved to be a suitable approach. Combining the RD cleaned dataset with the verified dataset almost
FIGURE 2  Species occurrence data and spatial patterns of Neotropical snake diversity. (a) Geographical coverage of sampling of snakes measured in $1\degree \times 1\degree$ grid cells. RD = raw dataset, obtained from www.gbif.org; VD = verified dataset, presented here; CD = combined dataset, produced by merging RD and VD. (b) Species richness at grid cells. (c) Phylogenetic diversity at grid cells. (d) Species richness at the ecoregion scale. (e) Phylogenetic diversity at the ecoregion scale.
doubled the number of species in the CD and substantially increased the geographical coverage to cover the Neotropical region more adequately.

This study provides the most comprehensive and novel database on snakes in the Neotropical region to date. Our CD increased the knowledge about Neotropical snakes, providing data for 886 species, an improvement of 670 species compared with previous studies (216 species by Böh m et al., 2013). We believe that CD also led to a considerable reduction in the number of ‘poorly known’ snake species regarding geographical distribution, especially for the tropical regions known to contain the most species classified as data deficient or threatened (see Supporting Information Appendix S4; Bland & Böh m, 2016; Böh m et al., 2013; Tingley, Meiri, & Chapple, 2016).

4.2 | Spatial patterns of species richness and phylogenetic diversity

The SR patterns found for Neotropical snakes broadly correspond to the patterns previously reported for other vertebrates (e.g., Fenker, Tedeschi, Pyron, & Nogueria, 2014; Jenkins et al., 2015; Moura et al., 2016). In contrast, our results contradict previous suggestions that Amazonia (here considered as poorly sampled) is the richest area for Neotropical reptiles (Böh m et al., 2013). This discrepancy may be explained by differences between our dataset and that of Böh m et al. (2013), which also included lizards, used species ranges instead of grid cells, adopted a different spatial scale, and was based on a random sampling, which in theory is meant to provide an adequate representation of species globally, but in practice may be problematical.

A different view of areas harbouring high SR and PD emerges on the scale of ecoregions (Figure 2d,e). For both indices, the Cerrado is the most diverse region. Accordingly, these results indicate that snake diversity in seasonally dry tropical forests may be more diverse than in rain forests, a pattern not previously inferred. The Cerrado is a global biodiversity hotspot (Mittermeier, Turner, Larsen, Brooks, & Gascon, 2011; Myers et al., 2000), harbouring ≥ 153 species of snakes, of which 49 are endemic (Guedes, Nogueira, & Marques, 2014). It is also the world’s most species rich savanna in number of woody plant species and has higher diversity than any other dry forests in the Neotropics (DRYFLOR et al., 2016). However, our results could be biased by the ecoregion boundaries used here, which separated the Atlantic Forest into distinct subregions, but did not do so to the Cerrado. As a whole, the Atlantic Forest harbours the richest snake fauna, including 236 species, of which 83 are endemic (Guedes et al., 2014). This situation reinforces the importance of refined data on species distributions for assessing the influence of spatial scale on patterns of biodiversity.

Despite the close relationship between SR and PD, the most species-rich areas are not fully coincident with areas of highest PD, as already reported by Fenker et al. (2014) for a clade of snakes. At the grid cell scale, we find highest SR and PD in forested areas, a similar pattern previously reported for amphibians, mammals and birds (Jenkins et al., 2015; Moura et al., 2016). Such areas also appear to contain high PD for particular groups of snakes, such as the relatively diverse genus Bothrops (Fenker et al., 2014). However, high-PD areas can probably be explained by sympathy of widely divergent lineages. This should occur in grid cells where species of open-habitat clades are found together with forest-adapted clades (Fenker et al., 2014).

4.3 | Sampling gaps

Our sampling gap map reflects a situation similar to that documented for other vertebrates (Meyer, Kref t, Guralnick, & Jetz, 2015). Amazonia has the smallest number of records of snakes in relationship to its area, which was predictable in face of the scattered data already reported for other groups (Peres, 2005). The region’s high inaccessibility, low investments in local research and the relative shortage of experts to explore this huge area are likely to explain this result. In contrast, well-sampled areas were coincident with the location of the most active universities and scientific collections of reptiles.

5 | CONCLUSIONS

Our study demonstrates that Neotropical snake diversity is unevenly distributed, with some ecoregions, such as the Cerrado, containing a disproportionately high diversity. We also showed that merging public and manually compiled data sources is likely to provide the largest taxonomic and geographical coverage for any system under study. However, a proper taxonomic verification, examination and assessment of biases of the public dataset proved crucial. As a result, we can now provide a solid and reliable foundation for any kind of meta-analysis, including the assessment of climate change effects, conservation strategies or design of future research agendas. Conservation priorities should focus on areas of high diversity values as well as high threat by landscape changes. Finally, we found highest diversity values in forested areas, reinforcing the need for general habitat protection compared with actions that are targeting specific species.

In order to increase our knowledge about Neotropical snakes, a geographically and taxonomically focused sampling is required, targeting Amazonia and those species whose distributions are so far largely unknown.

ACKNOWLEDGMENTS

We are grateful to Alejandro Giraudo and Vanesa Arzamendia (Universidad Nacional del Litoral, Argentina); António Argólo (State University of Santa Cruz, Brazil); Francisco Franco and Otavio Marques (Butantan Institute, Brazil); Christine Strüssmann (Federal University of Mato Grosso, Brazil); Marinus S. Hoogmoed (Museu Paraense Emílio Goeldi); Fausto Barbo, Giovanna Montingelli, Juan Arredondo, Hussam Zaher and Márcio Martins (University of São Paulo, Brazil); Márcio Borges-Martins (Federal University of Rio Grande do Sul, Brazil); and Josué Azevedo (University of Gothenburg, Sweden), who kindly provided refined data on species distributions that will be soon made available through the Atlas of Brazilian Snakes coordinated by C.C.N. We also thank all curators and technical managers.
of all institutions visited by the authors and our collaborators over many years of collecting data. [Correction added on 1 December 2017, after first online publication: Marinus S. Hoogmoed (Museu Paraense Emílio Goeldi) was added in this version].

DATA ACCESSIBILITY
Maps containing the distribution of all species used in this study are available as rasterfile in Supporting Information Appendix S3.

AUTHOR CONTRIBUTIONS
T.B.G. and A.A. conceived the research; T.B.G., R.J.S., R.S.B., M.J., P.P., A.L.C.P., D.F.C.-H., H.B.B. and C.C.N. collected the data by visiting natural history museums and consulting literature; A.P. provided the phylogenies; T.B.G. analysed the biogeographical data and drew the maps, assisted by A.Z., S.L. and S.F.; T.B.G. and A.A. led the writing, with contributions from all authors, interpretation and discussion of the results.

ORCID
Thais B. Guedes http://orcid.org/0000-0003-3318-7193
Søren Faurby http://orcid.org/0000-0002-2974-2628

REFERENCES
Bland, L. M., & Böhm, M. (2016). Overcoming data deficiency in reptiles. Biological Conservation, 204, 16–22.

Bohm, M., Collen, B., Baillie, J. E. M., Bowles, P., Chanson, J., Cox, N., ... Zug, G. (2013). The conservation status of the world’s reptiles. Biological Conservation, 157, 372–385.

Butchart, S. H. M., & Bird, J. P. (2010). Data deficient birds on the IUCN Red List: What don’t we know and why does it matter? Biological Conservation, 143, 239–247.

DryFlor, Banda-R, K., Delgado-Salinas, Dexter, K. G., Linares-Palomino, R., Oliveira-Filho, A., ... Pennington, R. T. (2016). Plant diversity patterns in neotropical dry forests and their conservation implications. Science, 353, 1383–1387.

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. Biological Conservation, 61, 1–10.

Faith, D. P. (2008). Threatened species and potential loss of phylogenetic diversity: Conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. Conservation Biology, 22, 1461–1470.

Fenker, J., Tedeschi, L. G., Pyron, R. A., & Nogueira, C. C. (2014). Phylogenetic diversity, habitat loss and conservation in South American pitvipers (Crotalinae: Bothrops and Bothrocophias). Diversity and Distributions, 10, 1108–1119.

Ficetola, G. F., Rondinini, C., Bonardi, A., Katařiá, V., Padoa-Schioppa, E., & Angulo, A. (2013). An evaluation of the robustness of global amphibian range maps. Journal of Biogeography, 41, 211–221.

Guedes, T. B., Nogueira, C., & Marques, O. A. V. (2014). Diversity, natural history, and geographical distribution of snakes in the Caatinga, Northeastern Brazil. Zootaxa, 3863, 1–93.

IUCN. (2017). The IUCN red list of threatened species (Version 2017-1). Retrieved from http://cmsdocs.s3.amazonaws.com/summarystats/2017-1_summary_stats_page_documents/2017-1_RL_Stats_Table_1.pdf

Jenkins, C. N., Alves, M. A. S., Uezu, A., & Vale, M. M. (2015). Patterns of vertebrate diversity and protection in Brazil. PloS One, 10, e0145064.

Laffan, S. W., Lubarsky, E., & Rosauer, D. F. (2010). Biodiverse, a tool for the spatial analysis of biological and related diversity. Ecography, 33, 643–647.

Lewin-Koh, N. J., Bivand, R., Pesbema, E. J., Archer, E., Baddeley, A., Bibiko, H. J., ... Turner, R. (2011). MapTools: Tools for reading and handling spatial objects (R package version 0.8–10). Retrieved from http://CRAN.R-project.org/package=maptools

Maldonado, C. M., Molina, C. I., Ziska, A., Persson, C., Taylor, C. M., Albán, J., ... Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? Global Ecology and Biogeography, 24, 973–984.

Meiri, S., & Chapple, D. G. (2016). Biases in the current knowledge of threat status in lizards, and bridging the ‘assessment gap’. Biological Conservation, 204, 6–15.

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. Nature Communications, 6, 8221.

Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters, 8, 992–1006.

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., & Gascon, C. (2011). In F. E. Zachos and J. C. Habel (Eds.), Global biodiversity conservation: The critical role of hotspots (pp. 3–22). London, U.K.: Springer.

Morlon, H., Schwilk, D. W., Bryant, J. A., Marquet, P. A., Rebelo, A. G., Tauss, C., ... Green, J. L. (2011). Spatial patterns of phylogenetic diversity. Ecology Letters, 14, 141–149.

Moura, M. R., Villalobos, F., Costa, G. C., & Garcia, P. C. A. (2016). Disentangling the role of climate, topography and vegetation in species richness gradients. PloS One, 11, e0152468.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403, 853–858.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., ... Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on earth a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience, 51, 933–938.

Peres, C. (2005). Porque precisamos de megareservas na Amazonia. Megadiversidade, 1, 174–180.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3–900051-07-0. Retrieved from http://www.R-project.org

Schipper, J., Chanson, J. S., Chiozza, F., Cox, N. A., Hoffmann, M., Katařiá, V., ... Young, B. E. (2008). The status of the world’s land and marine mammals: Diversity, threat, and knowledge. Science, 322, 225–230.

Tingley, R., Meiri, S., & Chapple, D. G. (2016). Addressing knowledge gaps in reptile conservation. Biological Conservation, 204, 1–5.

Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W., & Pyron, R. A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biological Conservation, 204, 23–31.

Topel, M., Ziska, A., Calió, M. F., Scham, R., Silvestro, D., & Antonelli, A. (2017). SpeciesGeoCoder: Fast categorisation of species occurrences for analyses of biodiversity, biogeography, ecology and evolution. Systematic Biology, 66, 145–151.
Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A.,… Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. Biological Reviews, 92, 698–715.

Uetz, P., & Ho/C20sek, J. (2016). The reptile database. Retrieved from http://www.reptile-database.org

Wallach, V., Williams, K. L., & Boundy, J. (2014). Snakes of the world. A catalogue of living and extinct species. New York, NY: CRC Press.

BIOSKETCH

THAÍS BARRETO GUEDES (http://www.tbguedes.com) is an evolutionary biologist interested in biogeography, macroecology and conservation of the herpetofauna. She is investigating these subjects on a broad scale especially in snakes, but also in amphibians and other reptiles.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Guedes TB, Sawaya RJ, Zizka A, et al. Patterns, biases and prospects in the distribution and diversity of Neotropical snakes. Global Ecol Biogeogr. 2018;27:14–21. https://doi.org/10.1111/geb.12679