Received: 25 January 2019 | Revised: 10 January 2020 | Accepted: 16 January 2020

DOI: 10.1111/jbi.13847

RESEARCH PAPER

Journal of Biogeography | WILEY

Large-scale DNA-based survey of frogs in Amazonia suggests a vast underestimation of species richness and endemism

Jean-Pierre Vacher[1] | Jérôme Chave[1] | Francesco Gentile Ficetola[2,3] | Guilhem Sommeria-Klein[1,4] | Shengli Tao[1] | Christophe Thébaud[1] | Michel Blanc[5] | Agustín Camacho[6] | José Cassimiro[7] | Timothy J. Colston[8] | Maël Dewynter[9] | Raffael Ernst[10] | Philippe Gaucher[11] | Jerriane Oliveira Gomes[12] | Rawien Jairam[13] | Philippe J. R. Kok[14] | Jucivaldo Dias Lima[15] | Quentin Martinez[11] | Christian Marty[16] | Brice P. Noonan[17] | Pedro M. Sales Nunes[18] | Paul Ouboter[19] | Renato Recoder[7] | Miguel Trefaut Rodrigues[7] | Andrew Snyder[20] | Sérgio Marques-Souza[7] | Antoine Fouquet[1,11]

[1]Laboratoire Évolution et Diversité Biologique, UMR5174, CNRS-UPS-IRD, Toulouse Cedex 9, France

[2]Laboratoire d'Écologie Alpine (LECA), Université Grenoble Alpes, CNRS, Grenoble, France

[3]Department of Environmental Science and Policy, Università degli Studi di Milano, Milan, Italy

[4]Institut de Biologie de l'Ecole normale supérieure (IBENS), Ecole Normale Supérieure, CNRS, INSERM, PSL Université Paris, Paris, France

[5]Pointe Maripa, RN2/PK35, Roura, French Guiana

[6]Departamento de Fisiologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

[7]Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

[8]Department of Biological Sciences, Florida State University, Tallahassee, FL, USA

[9]Biotope, Agence Amazonie-Caraïbes, Cayenne, French Guiana

[10]Museum of Zoology, Senckenberg Natural History Collections Dresden, Dresden, Germany

[11]Laboratoire Écologie, Évolution, Interactions des Systèmes Amazoniens (USR 3456 LEEISA), Université de Guyane, CNRS Guyane, Cayenne, French Guiana

[12]Laboratório de Herpetologia/CZO, Museu Paraense Emílio Goeldi, Pará, Brazil

[13]National Zoological Collection of Suriname

Abstract

Aim: Mapping Amazonian biodiversity accurately is a major challenge for integrated conservation strategies and to study its origins. However, species boundaries and their respective distribution are notoriously inaccurate in this region. Here, we generated a georeferenced database of short mtDNA sequences from Amazonian frogs, revised the species richness and the delimitation of bioregions of the Eastern Guiana Shield and estimated endemism within these bioregions.

Location: Amazonia, with a focus on the Eastern Guiana Shield.

Taxon studied: Amphibia: Anura.

Methods: We used an extensive DNA-based sampling of anuran amphibians of Amazonia using next-generation sequencing to delimit Operational Taxonomic Units (OTU) and their distribution. We analysed this database to infer bioregions using Latent Dirichlet Allocation modelling. We then compared endemism within these bioregions based on our results and the current IUCN database, and inferred environmental variables that contributed the most to the biogeographic pattern.

Results: The recognized anuran species richness within the focal area increased from 440 species currently listed by the IUCN Red List to as much as 876 OTUs with our dataset. We recovered eight bioregions, among which three lie within the Eastern Guiana Shield. We estimated that up to 82% of the OTUs found in this area are endemic, a figure three times higher than the previous estimate (28%). Environmental features related to seasonal precipitations are identified as playing an important role in shaping Amazonian amphibian bioregions.

Main conclusions: Our results have major implications for defining future conservation priorities of this vast area given that endemism in most Amazonian bioregions is vastly underestimated, and might therefore hide a large portion of threatened

(NZCS), Anton de Kom University of Suriname, Paramaribo, Suriname

[14]Biology Department, Amphibian Evolution Lab, Vrije Universiteit Brussel, Brussels, Belgium

[15]Instituto de Pesquisas Cientificas e Tecnológicas do Amapá, Macapá, AP, Brazil

[16]Impasse Jean Galot, Montjoly, French Guiana

[17]Department of Biology, University of Mississippi, University, MS, USA

[18]Departamento de Zoologia, Centro de Biocências, Universidade Federal de Pernambuco, Recife, Brazil

[19]Institute for Neotropical Wildlife and Environmental Studies (NeoWild), Paramaribo, Suriname

[20]Global Wildlife Conservation, Austin, TX, USA

Correspondence
Antoine Fouquet, Laboratoire Évolution et Diversité Biologique, UMR5174, CNRS-UPS-IRD, Bâtiment 4R1, 118 route de Narbonne, 31062 Toulouse Cedex 9, France
Email: fouquet.antoine@gmail.com

Funding information
Agence Nationale de la Recherche; German Research Foundation, Grant/Award Number: 2, 1, FWO12A7614N, FWO12A7617N and KDQF01000000; University of Sao Paulo; University of Mississippi, Grant/Award Number: 1, KDQF00000000 and 1

Handling Editor: Carina Hoorn

species. Moreover, these findings raise concern about meta-analyses based on public databases within Amazonia dealing with poorly known groups.

KEYWORDS
16S, amphibians, biodiversity, Eastern Guiana Shield, IUCN, richness, species assemblage

1 | INTRODUCTION

Amazonia harbours the largest lowland rainforest in the world, and it shelters an outstanding biological diversity in all terrestrial life forms (Gibson et al., 2011). However, it is exposed to multiple threats, notably climatic (Brienen et al., 2015; Malhi et al., 2008; Saatchi et al., 2013) and land use changes (Davidson et al., 2012; Lambin & Meyfroidt, 2011). In the global context of catastrophic biodiversity decline, the conservation of this region is thus of major concern.

To mitigate these threats, systematic species-mapping programs based on expert knowledge (Jenkins, Pimm, & Joppa, 2013; Kreft & Jetz, 2010) have helped to define priority areas for biological conservation (Lomolino, Riddle, & Whittaker, 2016; Olson et al., 2001) notably via the delimitation of bioregions, which are spatial areas defined by the presence of co-occurring species. The limits of these regions could correspond to dispersal barriers, environmental gradients and shared ecological and evolutionary history among species (Antonelli, 2017; Bloomfield, Knerr, & Encinas-Viso, 2018). However, these bioregions were mostly delimited according to avian distribution patterns (Cracraft, 1985; Naka, Bechtoldt, Magalli, & Pinto, & Brumfield, 2012), which undoubtedly represent the best known animal group in terms of species distribution and taxonomy in Amazonia. Still, they may not match those of other taxonomic groups as birds have high dispersal abilities and are generally less sensitive to abiotic conditions than many other organisms, especially terrestrial ectotherms (Calderón-Patrón, Moreno, Pineda-López, Sánchez-Rojas, & Zuria, 2013; Qian, 2009; Rueda, Rodríguez, & Hawkins, 2010; Silva, Almeida-Neto, Prado, Haddad, & Rossa-Feres, 2012). Using assemblages of small terrestrial vertebrates with more limited dispersal abilities and probably greater sensitivity to environmental variation such as tailless amphibians (anurans) could yield different patterns than those previously described (Zeisset & Beebee, 2008), both at the continental (Vilhena & Antonelli, 2015) and regional scales (Vasconcelos, Prado, da Silva, & Haddad, 2014). Delimiting these bioregions is important for conservation but also to understand the evolutionary processes at the origin of Amazonian biodiversity (Antonelli et al., 2010; Hazzi, Moreno, Ortiz-Movliav, & Palacio, 2018).

A recent biogeographic analysis based on the IUCN distribution of Amazonian anurans found less bioregions than for birds (Godinho & da Silva, 2018). This result was surprising because more bioregions would have been expected for anurans due to their limited dispersal abilities (Godinho & da Silva, 2018). Here, we contend that such a

pattern might result from the inaccuracy of currently available distribution data, which is mostly linked to notoriously prevalent taxonomic uncertainties in Amazonian anurans (Fouquet et al., 2007, 2015, 2016; Funk, Caminer, & Ron, 2012; Motta, Menin, Almeida, Hrbek, & Farias, 2018; Vacher et al., 2017).

To test our hypothesis and improve the estimate of anuran richness in Amazonia, we assembled the largest database of geotagged DNA sequences of anurans from Amazonia to date, with a focus on the Eastern Guiana Shield (Figure 1), using high-throughput sequencing and combined these data with previously published information. First, we assessed the current taxonomic framework for Amazonian amphibians using sequence similarity based on a DNA-only approach, and then we used this dataset to define a new set of bioregions within Amazonia, and re-evaluated the regional endemism for anurans.

2 | MATERIALS AND METHODS

2.1 | Database construction

Our focal area (limits: W 72° W 47° S 11° N 09°) includes the central, eastern and northern parts of Amazonia, including Pantepui, but excluding western and southern periphery of Amazonia where our sampling was scarcer (Figure 1). This focal area was delimited to encompass the Guiana Shield (sensu Lujan & Armbruster, 2011) and its surrounding areas: the central and eastern parts of the Rio Amazonas drainage, and the northern parts of the Rio Purus, Madeira, Tapajós, Xingú and Tocantins watersheds (Figure 1), as well as peripheral non-Amazonian areas.

Our database was a geolocalized collection of amphibian 16S ribosomal DNA sequences of ca. 400 bp in length, a region of the mitochondrial genome widely used in barcoding studies of Amazonian anurans (Mayer, Fonte, & Lötters, 2019). This DNA region has extensively been used in phylogenetic studies in the region and has been shown to provide high taxonomic resolution in anuran amphibians and successful PCR amplification is possible for all members of this group with classic primer pairs (Fouquet et al., 2007; Vences, Thomas, Meijden, Chiari, & Vieites, 2005). For this study, we included 4,492 unpublished sequence data for individuals collected in Amazonia (Appendix S1 for details). We combined it with a carefully curated sequence database of 6,672 accessions from 50 Amazonian anuran genera (including all congeneric species) extracted from the NCBI online repository. Hence, the data set contained 11,164 accessions, 10,268 of which were geotagged, with 8,181 records within Amazonia, of which 4,634 were from the Eastern Guiana Shield.

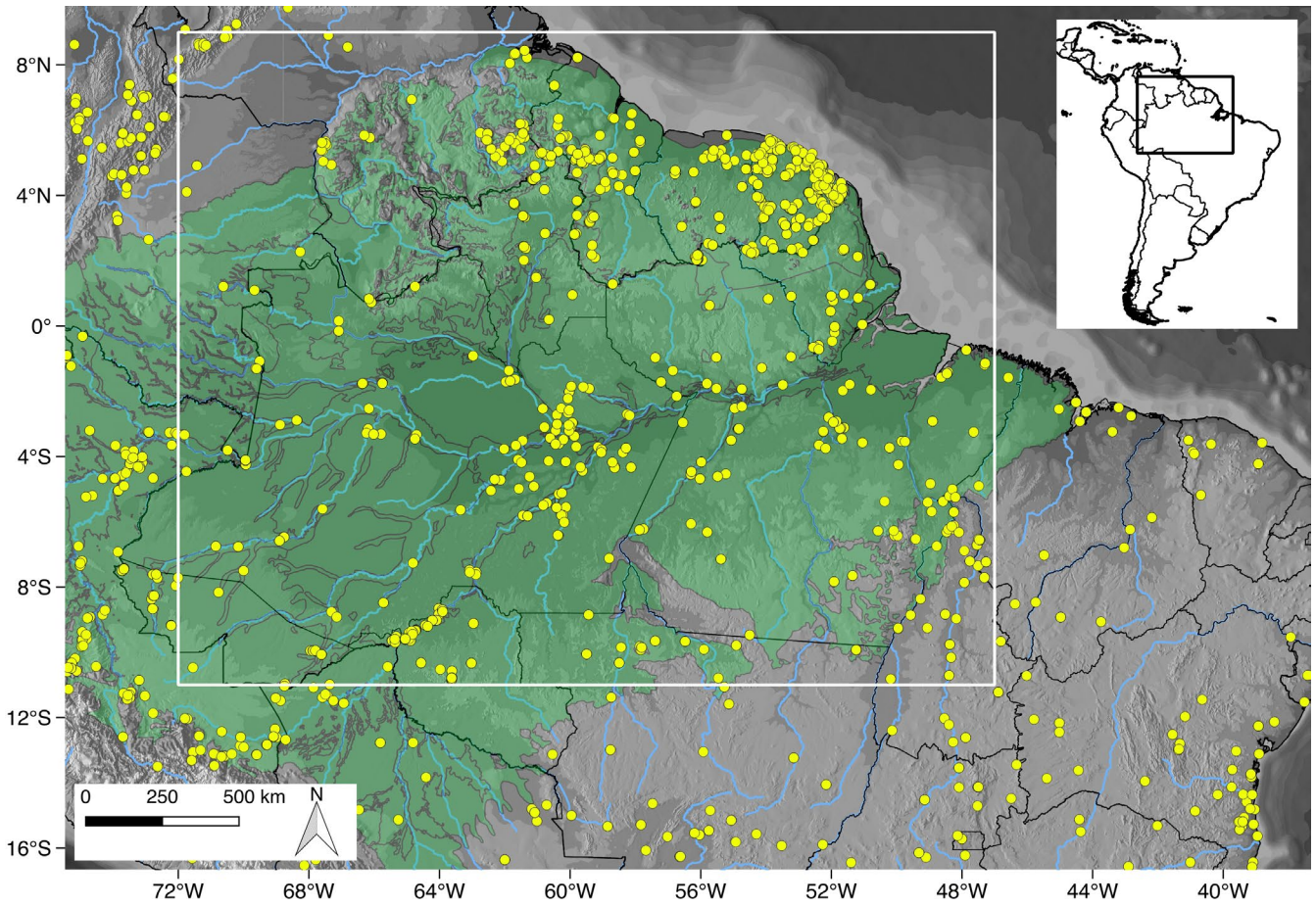

FIGURE 1 Map of all the occurrences (yellow points) of the barcoding dataset in Amazonia (green surface), which include the samples that we collected as well as data retrieved from GenBank. The white rectangle delimits the focal area of our study [Colour figure can be viewed at wileyonlinelibrary.com]

New samples were collected in the field in Guyana, Suriname, French Guiana and the Brazilian states of Amapá, Pará and Roraima, through nocturnal and diurnal active searches (visual and acoustic). Each specimen was identified to the species level, photographed in life and then euthanized using an injection of Xylocaine® (lidocaine chlorhydrate). Tissue samples (liver or muscle tissue from thigh or toe-clip) were taken and stored in 95% ethanol. Specimens were tagged and fixed (using formalin 10%) before being transferred to 70% ethanol for permanent storage. We also obtained biological material for sequencing maintained by several research institutions (see Acknowledgements), and covering the upper portions of Madeira, lower Xingú, Abacaxis and Purus Rivers.

We extracted DNA from liver or muscle tissue (thigh or toe-clip) of the 4,681 samples using the Wizard Genomic extraction protocol (Promega). We targeted a ca. 400bp fragment of the 16S rDNA. We used primers N16R and N16F (Salducci, Marty, Fouquet, & Gilles, 2005), to which we added NNN + 8-nucleotide labels (hereafter, designated as "tags"), for sample identification as all resulting PCR products were mixed into libraries: 32 tags for forward primer (N16R) and 36 tags for reverse primer (N16F). PCRs were carried out in a final volume of 20 µL, and contained 2 µL of 50 ng/µL DNA extract, 10 µL of AmpliTaq Gold® 360 Master Mix (Life Technologies), 5.84 µL of Nuclease-Free Ambion Water (Thermo Fisher Scientific), 0.25 µM of each primer and 3.2 µg of bovine serum albumin (BSA, Roche Diagnostic). The PCR conditions were as follow: 95°C for 10 min, then 40 cycles of 95°C for 30 s, 46°C for 30 s, 72°C for 30 s and followed by a final step of 72°C for 7 min. We prepared three complete libraries, each containing 1,152 samples, including 72 blanks (6 blanks per plate). Libraries of mixed PCR products were sequenced using 2 × 250 paired-end reads sequencing technology through MiSeq high-throughput sequencing (Illumina) at the Génopole. The resulting outputs were analysed with the OBITOOLS software suite (Boyer et al., 2016). Paired-end reads were assembled and merged, and we used the tag attached to the primer to assign each reads to its label. Then, we removed low-quality reads (alignment scores < 50, containing Ns or shorter than 50 bp). The resulting batch of reads was dereplicated while keeping the coverage information (number of reads merged). All sequences < 100 bp were discarded. Eventually, every sequence that we included in our dataset was > 380 bp long. We retrieved 3,148 sequences through this MiSeq sequencing procedure. We completed this dataset with 1,345 sequences that were retrieved from Sanger sequencing. We sequenced 30 samples using both methods, which led to identical sequences. For Sanger sequencing, PCR were conducted in a final volume of 25 µL each containing 2 µL of DNA template, 14.36 Mq water, 5 µL of 10 × PCR Buffer, 1.25 µL of each primer, 1.67 µL of MgCL2, 0.5 µL of dNTPs and 0.22 µL of GoTaq (Promega). The PCR conditions were as follows: 8 cycles of denaturation (45 s at 94°C), annealing (60 s at 46°C) and elongation (90 s at 72°), followed by 22 cycles of denaturation (45 s at 94°C), annealing (60 s at 50°C) and elongation (90 s at 72°C). We used the same N16F and N16R primers that we used for the MiSeq sequencing. Sanger sequencing was performed by Genoscreen.

2.2 | Molecular-based species delimitation

As all the samples used for the sequencing procedure have been initially assigned to a genus based on morphology, we grouped all our sequences at the genus level. The resulting clusters of sequences were investigated into operational taxonomic units (OTUs), applying an unsupervised method of distance-based DNA barcodes clustering, the Automatic Barcode Gap Discovery (ABGD) species delimitation method (Puillandre, Lambert, Brouillet, & Achaz, 2012). This method recursively searches for gaps in the distribution of pairwise divergences, and partitioning of data is repeated until no further splitting occurs. It is now widely used for this application because it is robust to slight changes in parameter assumptions (Fontaneto, Kaya, Herniou, & Barraclough, 2009; Pentinsaari, Vos, & Mutanen, 2017; Pons et al., 2006), and produced conservative results relative to taxonomy, with the highest number of merges and lowest number of splits compared with other delimitation methods like GMYC or PTP (Lin, Stur, & Ekrem, 2015; Pentinsaari et al., 2017). We performed ABGD analyses with default settings (Jukes-Cantor sequence substitution model, a prior on the intraspecific pairwise distance defined by Pmin: 0.001, Pmax: 0.1, an iterative procedure set at 10 steps and 20 bins). The number of clusters found for each genera using 2%, 3%, 4% and 6% p-distance thresholds indicates that 3%–4% seems the most reasonable choice, given a lower threshold (2%) clearly over-splits compared with current taxonomy, while a higher threshold (6%) lumps together many phenotypically distinct described species (Appendix S3). We plotted the distribution of the thresholds found for each genus and found that the median p-distance threshold was 3% (Figure S2). We therefore used the 3% partitioning in subsequent analyses. The resulting clusters were considered as OTUs. Nevertheless, in 24 species pairs (17 concerning Amazonian taxa), distinct taxa (Jungfer, et al., 2013; Noonan & Gaucher, 2005) were lumped as unique OTUs because of shallow pairwise mtDNA divergence between them (notably among *Atelopus* spp. and *Osteocephalus* spp.). In these situations, we retained the accepted names rather than lumping previously accepted taxa.

For visualization purposes, we generated an unrooted Maximum Likelihood (ML) tree in RAxML v.8.2.4 (Stamatakis, 2014) of the 11,164 16S sequences that were aligned with MAFFT v.7 (Katoh & Standley, 2013). For the ML analysis, we used the GTR + Γ+I and we investigated support of nodes with 1,000 bootstrap replicates using the fast bootstrap algorithm.

2.3 | Distribution data and bioregions

We then used the amphibian occurrence data set to partition natural bioregions. To infer the presence or absence of the OTUs in a grid composed of 500 1° × 1° cells, we computed their ranges as polygons, a simple approximation which is comparable to the IUCN data, using the 'sp' package implemented in R (Pebesma & Bivand, 2005; R

Development Core Team, 2016). To perform this step, we removed 657 OTUs (out of 1,252) with less than three occurrences, resulting in a reduced data set. Eventually, 595 OTUs belonging to 50 genera were used in the species distribution analyses. Finally, we excluded cells with less than five OTUs from the spatial analysis.

We also reanalysed the IUCN Red List data for comparison. Species ranges (polygons) for 440 species were downloaded from the IUCN data portal http://www.iucnredlist.org/technical-documents/spatial-data#amphibians (accessed June 2016). We excluded 21 genera from the molecular data set because they were associated with the Cerrado, the Atlantic Forest or the Andean foothills and thus only overlapped marginally with the study area, i.e. less than 10 occurrence points within cells at the margin of the study area (Appendix S4) or because no corresponding DNA sequences were available (*Metaphryniscus*, *Dischidodactylus*). We also removed two introduced species (*Eleutherodactylus johnstonei* and *Lithobates catesbeianus*).

We generated amphibian bioregions by Latent Dirichlet Allocation (LDA; Blei, Ng, & Jordan, 2003; Valle, Baiser, Woodall, & Chazdon, 2014). One advantage of LDA over classic distance-based clustering is that it models gradual clines in taxonomic composition, assigning a probability for each cell to belong to one bioregion. LDA fits a probabilistic model to the community matrix (i.e. the matrix listing the species present in each grid cell) that assumes the coexistence of several assemblages of species over the study area. The number K of species assemblages is fixed, but can be optimized by AIC minimization. The assemblages may partially overlap in taxonomic composition, and a given grid cell may either be dominated by one assemblage or contain a mixture of assemblages. The estimated value of the mixing parameter α indicates whether the cells tend to be dominated by a single assemblage (case $\alpha < 1$). We followed the approach described in details in Sommeria-Klein et al. (2019) and summarized below. We fitted LDA using the Variational Expectation Maximization algorithm in the R package *topicmodels* (Grün & Hornik, 2011), with convergence threshold of 10^{-6} for the EM step and 10^{-8} for the variational step. We assessed the robustness of the solution by replicating the analysis a hundred times for random initial assemblages. The best bioregionalization was assumed to be that corresponding to the realization with the highest likelihood value. We optimized the number of assemblages by AIC minimization. All the scripts used for this procedure are available on github: https://github.com/guilhemSK/eDNA_LDA.

Finally, we explored the congruence of the obtained bioregions with environmental variables by applying a random forest classification model. We first compiled 30 environmental variables for Amazonia from different sources (Appendix S5), all reanalysed at 5-km resolution. These variables capture essential environmental features in the tropics, such as precipitation, elevation, soil moisture, forest structure, land surface temperature and bioclimatic variables (Appendix S5, S11, S12). Intra- and inter-annual variability for some environmental features was also calculated when time-series observations were available (Appendix S5, S11 and S12). We then built a multi-classes random forest classification model using the 30 features as explanatory variables and the bioregions as targeted classes (Appendix S13). To further inspect the environmental uniqueness of the bioregions within the Guiana Shield, we built a two-class random forest model for each of the bioregions in the Guiana Shield by labelling the target bioregion as class 1 and all other bioregions as class 2 (Appendix S14). The random forest model helped evaluate the contribution of each environmental feature to the definition of the bioregions. The *randomForest* package implemented in R was used for random forest modelling (Liaw & Wiener, 2002).

3 | RESULTS

3.1 | Revising species richness estimates in Amazonian amphibians

Our DNA barcoding species delimitation (ABGD) yielded a total of 1,486 OTUs [1,024-3,534, corresponding to the upper 6% and lower 2% ABGD thresholds]. Within the focal area, we retrieved 746 OTUs, among which only 310 could be assigned to taxa listed in the Red List. These results were overall in accordance with the ML topology with a few exceptions (Appendix S15). In nine instances within the ML topology, well-phenotypically diagnosable species are recovered embedded within other well-phenotypically diagnosable species while ABGD recovered these species as distinct clusters. These incongruences are likely due to the low number of informative sites for certain branches using ML that increase uncertainty in tree building (Reid & Carstens, 2012; Machado, Castroviejo-Fisher, & Grant, 2019). Because the Red List draws on a larger database than this study, it contains 130 additional taxa not included in our data set. Therefore, our revised estimate for the number of anurans in the focal area reaches 876 species (310 species from both our data and the Red List, 130 from the Red List data set only and 436 from our data set only). This totalizes twice as many species as currently reported in the Red List. As our average pairwise genetic distance (barcoding gap threshold) was >3% in 85% of the splits (and even >6% in 39% of them) and as several OTUs in our analysis have been confirmed to correspond to new valid species (Fouquet et al., 2015, 2016, 2018; Rojas et al., 2018), we assume that the majority of the OTUs detected here most likely represent genuine species, implying that our estimate of 876 species of anurans in the focal area is reasonable.

Our results demonstate that the known distribution range of many taxa has to be extensively revised. Most of the OTUs actually display much narrower distributions than the species listed in the Red List. For example, 246 highly genetically divergent OTUs are distributed at the margin or outside the core distribution of the currently known species to which they supposedly correspond.

Few genera are endemic to the Guiana Shield (*Otophryne* and *Anomaloglossus*). On the contrary, the vast majority of the OTUs included here are nested within widespread Amazonian or lowland Neotropical clusters (Figure 2a). Most of these clusters display deep divergence among populations (e.g. >6% in *Leptodactylus petersii*—16 OTUs in the ABGD barcoding analysis, Figure 2a). Finally, only 45 (6%) of the 746 OTUs identified in the focal area have a broad distribution

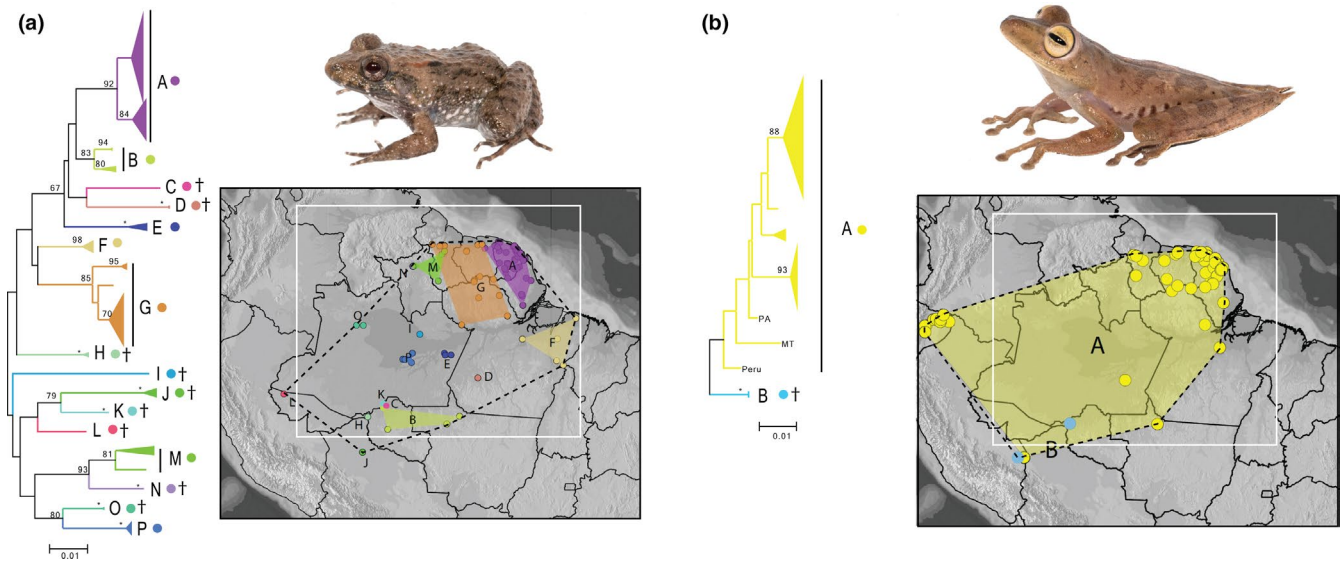

FIGURE 2 Examples of genetic and geographic structures for two panamazonian taxa. (a) The genetic structure analysis of *Leptodactylus petersii* by the Automatic Barcode Gap Discovery analysis splits the taxon into 16 OTUs with largely non-overlapping geographic distributions. (b) In contrast, the genetic analysis of *Boana calcarata* splits the clade into only two OTUs. The colours of the lineages on the tree correspond to the colours of the occurrence points and areas on the map. Symbol † indicates taxa with less than three locality records, discarded from the bioregionalization analysis. The white square on the maps corresponds to our focal area [Colour figure can be viewed at wileyonlinelibrary.com]

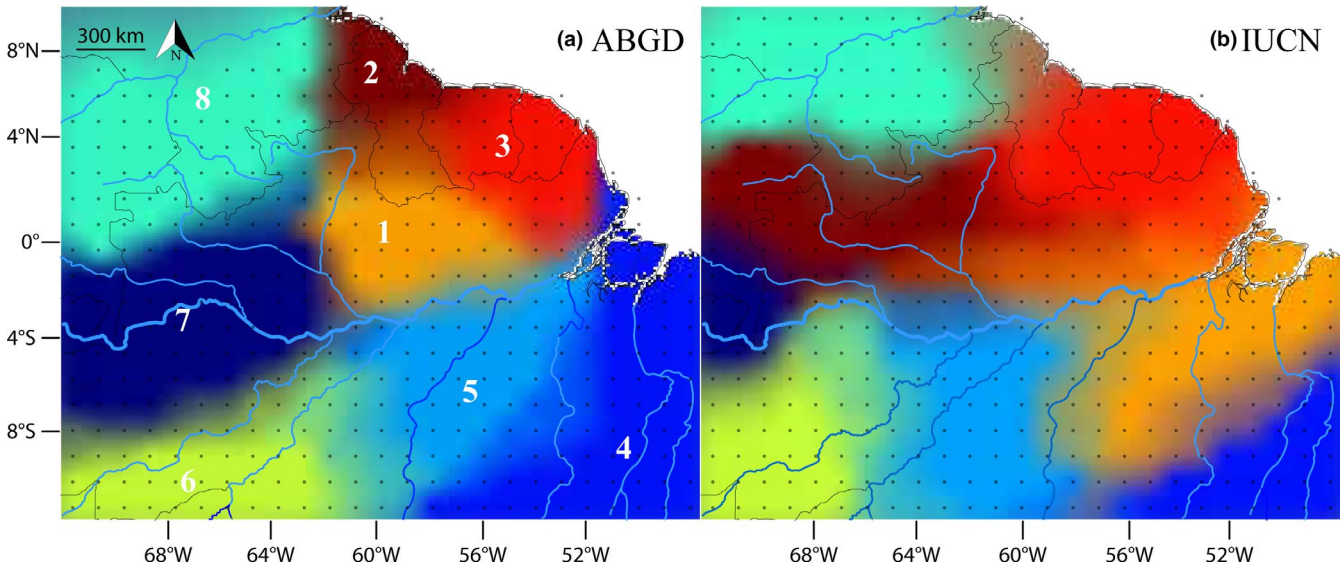

FIGURE 3 Maps of the bioregions based on species occurrence data for Amazonian anurans as obtained from (a) DNA-based species delimitation; (b) IUCN data. Colours represent affiliations of cells to bioregions as predicted by Latent Dirichlet Allocation (LDA) decomposition, followed by kriging (function gstat, R package *gstat*, with exponentially decreasing weights), for $K = 8$ bioregions labelled with numbers. In cells where several bioregions coexist, colours are mixed according to the relative probability of affiliation to bioregions [Colour figure can be viewed at wileyonlinelibrary.com]

(over 1 million km^2, e.g. *Boana calcarata*, Figure 2b). This is in stark contrast with the previously reported figure of 142 widespread species out of 440 (32%) based on the Red List.

3.2 | Bioregionalization and endemism

The optimal number of bioregions recovered by Latent Dirichlet Allocation (LDA) modelling of our data set (ABGD) was eight (Figure S6). Spatial segregation was strong among bioregions ($\alpha < 0.02$) and the solution was stable across runs (Figure S7). The Amazon River and the Pantepui region represent obvious landscape boundaries of the eastern Guiana Shield region, which is subdivided into three clearly delimited bioregions (1–Southern; 2–North-western; 3–Eastern; Figure 3a and Figure S8a). Bioregion 1 extends across the southern part of Guyana, Roraima and the Northern parts of Pará and Amazonas states in Brazil. Bioregion 2 comprises the northern part of Guyana and adjacent Venezuela. Finally, bioregion 3 covers the

state of Amapá (Brazil), French Guiana and Suriname. Outside of the Eastern Guiana Shield, the boundary between bioregions 6 and 7 roughly follows the course of the Purus River. The limits among bioregions 4, 5 and 6 are less sharp as they are displaying shared probability of presence across five cells (roughly 250 km^2). Intra-annual variability in precipitation and precipitation in the coldest quarter both had high correlations with the distribution of bioregions throughout the focal area for both ABGD and IUCN data sets (Figure S9). Within the Eastern Guiana Shield, the variables that contribute the most to the ABGD bioregions are elevation and precipitation, while soil moisture contributes also extensively to Bioregion 2 and forest structure and moisture to Bioregion 3, thus emphasizing an environmental heterogeneity matching bioregions.

Our ABGD data set yielded a very different pattern compared to the Red List data set (Figure 3b). Using the Red List data set, the Guiana Shield formed a single bioregion, excluding the north-western part of Guyana and including adjacent areas of Amapá and Pará (Brazil). This discrepancy appears to be due to the inaccuracy of many distribution ranges in the Red List data set as discussed above.

Finally, using our revised delimitation of bioregions of the Eastern Guiana Shield, we assessed the number of OTUs (including singletons) restricted to each one of these three bioregions (i.e. endemic). Bioregion 2 had the highest number of endemic OTUs in our data set, reaching 75% endemism, followed by bioregion 3 (59%) and bioregion 1 (35%). Overall, 82% of the 250 OTUs occurring in the Eastern Guiana Shield were endemic to this area. This was again in stark contrast with the Red List data, which listed only 28% of the 119 species occurring there as endemic (Appendix S10).

4 | DISCUSSION

We assessed the taxonomic diversity in anuran amphibians based on a very large data set of DNA-barcoded and geotagged individuals sampled across Amazonia. Our species delineation analysis corroborates previous suggestions that the current number of anuran species occurring in Amazonia remains vastly underestimated (Ferrão et al., 2016; Fouquet et al., 2007; Funk et al., 2012; Motta et al., 2018). Even though species delimitation solely based on the divergence of small mitochondrial DNA (mtDNA) can overestimate the actual number of species in some cases (false positives) or underestimate it in others (false negatives) (Goldstein & DeSalle, 2011; Hickerson, Stahl, & Lessios, 2006; Krishnamurthy & Francis, 2012; Sukumaran & Knowles, 2017; Yu, Rao, Matsui, & Yang, 2017), such an approach can turn out to be reliable overall when applied to large assemblages of tropical frogs as confirmed by species delimitation based on integrative taxonomy (Vieites et al., 2009). As it was not realistic to gather morphological or bioacoustic data for most of the species included in this study, we could not adopt an integrative approach and evaluate the phenotypic divergence among the delimited OTUs. Nonetheless, recent works on species limits in Amazonian anurans using integrative taxonomy strongly suggest that a mtDNA divergence of at least 3% is frequently associated with phenotypic or acoustic differentiation (Ferrão et al., 2016; Fouquet et al., 2013, 2015, 2016; Funk et al., 2012; Jansen, Bloch, Schulze, & Pfenninger, 2011; Kok, Dezfoulian, Means, Fouquet, & Barrio-Amorós, 2018; Motta et al., 2018; Orrico et al., 2017; Ortega-Andrade et al., 2015; Rojas et al., 2018). Therefore, as most of the OTUs recovered herein were beyond that threshold, we postulate that many of them likely correspond to phenotypically distinct species. It is largely admitted that distribution data currently used in meta-analyses based on the Red List are inaccurate as many species are lumped together and distribution ranges are certainly overestimated (Mayer et al., 2019). Even though such analyses recently reported that Amazonian amphibians are greatly affected by upcoming threats, including climate change (Foden et al., 2013), it would be necessary to reassess these modelling results based on more solid species occurrence data (Ficetola, Rondinini, Bonardi, Baisero, & Padoa-Schioppa, 2015; Herkt, Skidmore, & Fahr, 2017; Holt et al., 2013; Hurlbert & Jetz, 2007; Rocchini et al., 2011).

Latent Dirichlet Allocation is an emerging method in the field of biogeography, which allows the identification of both smooth and sharp transitions in community composition across space (Valle, Albuquerque, Zhao, Barberan, & Fletcher, 2018; White, Dey, Mohan, Stephens, & Price, 2019). Our bioregionalization reveals the existence of a sharply delimited biodiversity structure within the Eastern Guiana Shield, with three well-delimited bioregions. Such a pattern strikingly differs from the current definition of bioregions based on bird data (Naka et al., 2012; Oliveira, Vasconcelos, & Santos, 2017) and from the previous definitions based on anuran distribution data (Godinho & da Silva, 2018). Outside the Eastern Guiana Shield, we found delimitations of bioregions to be fuzzier and less robust. This is likely due to scarcer sampling in these areas, and thus increased noise in the spatial structure. The LDA model explicitly accounts for uneven sample size across cells, and as a consequence we did not rarefy the data so as to make the best use of the available data. Nevertheless, we cannot exclude that sampling heterogeneity might introduce bias in the retrieved patterns of bioregionalization where sampling remains scarce. Overall, the high stability of the solution across runs of the algorithm for random initial conditions indicates that the retrieved bioregionalization is robust with regard to the available data. Models similar to LDA and specifically designed for species occurrence data (i.e. presence–absence) have recently been proposed (Valle et al., 2018; White et al., 2019), highlighting the growing interest for these methods in biogeography.

Within the Eastern Guiana Shield, the south-western limits of the three bioregions coincide with the Amazon River and the Branco/Negro Rivers as found in birds (Naka et al., 2012; Oliveira et al., 2017), but the other boundaries do not correspond to any conspicuous landscape features, except for the eastern limit between bioregions 1 and 3 that seems to coincide with the interfluve between the Parú and Jari Rivers (Figure 3a). The Jari River does not seem to constitute a biogeographic barrier for small terrestrial vertebrates though (Silva et al.,

2018). Therefore, we hypothesize that the biogeographic patterns we recovered in the Eastern Guiana Shield could mainly result from current and/or past environmental heterogeneity (Fouquet et al., 2012). This hypothesis is supported by the fact that these biogeographic patterns matched patterns in rainfall seasonality and soil moisture (Figure S9). Actually, the Eastern Guiana Shield is nowadays characterized by a pronounced climatic heterogeneity, with a dry corridor observed from southern Guyana to the state of Pará (Mayle & Power, 2008) where relict savannahs are found (e.g. Sipaliwini). Neotropical climatic fluctuations during the Miocene and Pliocene are thought to have played a crucial role for in-situ diversification (Antonelli et al., 2010) through landscape modifications that have acted as barriers to dispersal (Carnaval & Bates, 2007).

Our analysis strongly suggests that estimates of regional endemism for the Guiana Shield need to be extensively revised. We estimated that as many as 82% of the OTUs are endemic to the Eastern Guiana Shield. This figure is almost three times higher than the comparable estimates previously reported for anurans excluding the Pantepui endemics (Señaris & MacCulloch, 2005). By comparison, only 8% of birds are reported to be endemic to the Guiana Shield, 29% of reptiles and 11% of mammals (Hollowell & Reynolds, 2005); for reptiles and small mammals, endemism might also be vastly underrated in the Guiana Shield (Burgin, Colella, Kahn, & Upham, 2018; Guedes et al., 2017).

The classically reported number of Amazonian anuran species ranges from 427 to 577 (Da Silva, Rylands, & Fonseca, 2005; Godinho & da Silva, 2018), but our results suggest that this number is vastly underestimated. Considering that many parts of Amazonia are still undersampled (Motta et al., 2018), it is likely that our estimate remains in the low range. If we speculate that endemism and species richness in the other five Amazonian bioregions are similar to that of the Eastern Guiana Shield (82% endemism among 250 OTUs) and apply a rough extrapolation to all the bioregions of our focal area, we hypothesize that over 2,000 species of anurans might occur in the focal area. This number is four to five times larger than that currently reported for the entire Amazonia (Da Silva et al., 2005).

Today, the Eastern Guiana Shield has few protected areas (except bioregion 3). Northern Guyana harbours the highest endemism (bioregion 2), yet it only includes two protected areas and is currently heavily impacted by gold mining (Rahm et al., 2017). In the future, priority should also be placed on extending this line of research to the more threatened areas of Amazonia (Kalamandeen et al., 2018; Tracewski et al., 2016; Vedovato, Fonseca, Arai, Anderson, & Aragão, 2016).

ACKNOWLEDGEMENTS

We thank T.C. Avila Pires, D. Baudin, S. Cally, E.A. Courtois, A. Lorenzini, S. Manzi, A. Pelozuelo, A.L.C. Prudente , P. Solbès, Juliana Naomi Tanaka, B. Villette and L. Zinger for contribution and advice. Fieldwork was made possible by: “Our Planet Revisited” expedition (Muséum national d'Histoire naturelle, ProNatura international; APA n° 973-1), with support from Conseil régional de Guyane, Conseil général de Guyane, FEDER funds, Parc Amazonien de Guyane and DEAL Guyane; the Iwokrama International Centre and R. Thomas, GFC-PRDD for work in Guyana (permission from EPA Guyana). We thank the Nature Conservation Division and the STINASU for collecting permits in Suriname, and the Région Guyane and Parc Amazonien de Guyane for an Access and Benefit Sharing Agreement (APA n° 973-23-1) to collect in French Guiana. We acknowledge support from an 'Investissement d'Avenir' grant managed by *Agence Nationale de la Recherche* (CEBA, ref. ANR-10-LABX-25-01; TULIP, ref. ANR-10-LABX-0041; ANAEE-France: ANR-11-INBS-0001). RE was supported by a research grant from the German Research Foundation (DFG ER 589/2–1). PJRK was supported by research grants from the Fonds voor Wetenschappelijk Onderzoek (FWO12A7614N/FWO12A7617N). MTR thanks Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2003/10335-8 and 2011/50146-6)), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Dimensions of Biodiversity Program [FAPESP (BIOTA, 2013/50297-0), NSF (DOB 1343578), and NASA]. PMSN was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq fellowship number 313622/2018-3). Finally, we thank the following research facilities for lending us samples for this study: Departamento de Zoologia – Instituto de Biociências, University of Sao Paulo (Brasil), Museu Paraense Emilio Goeldi, Belem (Brasil), Museum of Zoology, Senckenberg Natural History Collections Dresden (Germany), Institut Royal des Sciences Naturelles de Belgique, Bruxelles (Belgium), The University of Mississippi (USA) and all the contributors who collected samples in the field (see Appendix S16).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available as a Targeted Locus Study project that has been deposited at DDBJ/ENA/GenBank under the accession KDQF00000000. The version described in this study is the first version, KDQF01000000 (see Appendix S1).

ORCID

Jean-Pierre Vacher https://orcid.org/0000-0002-9655-183X
Jérôme Chave https://orcid.org/0000-0002-7766-1347
Francesco Gentile Ficetola https://orcid.org/0000-0003-3414-5155
Guilhem Sommeria-Klein https://orcid.org/0000-0002-5331-3639
Shengli Tao https://orcid.org/0000-0002-2145-4736
Christophe Thébaud https://orcid.org/0000-0002-8586-1234
Agustín Camacho https://orcid.org/0000-0003-2978-792X
Timothy J. Colston https://orcid.org/0000-0002-1127-1207
Maël Dewynter https://orcid.org/0000-0002-5288-3429
Philippe J. R. Kok https://orcid.org/0000-0001-9743-9738
Brice P. Noonan https://orcid.org/0000-0001-7079-6129
Pedro M. Sales Nunes https://orcid.org/0000-0002-2635-9703
Paul Ouboter https://orcid.org/0000-0003-3114-8676
Renato Recoder https://orcid.org/0000-0001-9931-9238
Miguel Trefaut Rodrigues https://orcid.org/0000-0003-3958-9919
Andrew Snyder https://orcid.org/0000-0001-5588-9319
Sérgio Marques-Souza https://orcid.org/0000-0002-7137-8524

Antoine Fouquet https://orcid.org/0000-0003-4060-0281

REFERENCES

Antonelli, A. (2017). Biogeography: Drivers of bioregionalization. *Nature Ecology & Evolution*, *1*, 0114.

Antonelli, A., Quijada-Mascareñas, A., Crawford, A. J., Bates, J. M., Velazco, P. M., & Wüster, W. (2010). Molecular studies and phylogeography of Amazonian tetrapods and their relation to geological and climatic models. In C. Hoorn & F. Wesselingh (Eds.), *Amazonia, landscape and species evolution* (1st ed., pp. 386–404). Chichester, West Sussex, UK: Wiley-Blackwell.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning*, *3*, 993–1022.

Bloomfield, N. J., Knerr, N., & Encinas-Viso, F. (2018). A comparison of network and clustering methods to detect biogeographical regions. *Ecography*, *41*, 1–10.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*, 176–182.

Brienen, R. J. W., Phillips, O. L., Feldpausch, T. R., Gloor, E., Baker, T. R., Lloyd, J., … Zagt, R. J. (2015). Long-term decline of the Amazon carbon sink. *Nature*, *519*, 344–348.

Burgin, C. J., Colella, J. P., Kahn, P. L., & Upham, N. S. (2018). How many species of mammals are there? *Journal of Mammalogy*, *99*, 1–14.

Calderón-Patrón, J. M., Moreno, C. E., Pineda-López, R., Sánchez-Rojas, G., & Zuria, I. (2013). Vertebrate dissimilarity due to turnover and richness differences in a highly beta-diverse region: The role of spatial grain size, dispersal ability and distance. *PLoS ONE*, *8*, e82905.

Carnaval, A. C., & Bates, J. M. (2007). Amphibian DNA shows marked genetic strucure and tracks Pleistocene climate change in Northeastern Brazil. *Evolution*, *61*, 2942–2957.

Cracraft, J. (1985). Historical biogeography and patterns of differentiation within the South American avifauna: Areas of endemism. *Ornithological Monographs*, *36*, 49–84.

Da Silva, J. M. C., Rylands, A. B., & Da Fonseca, G. A. B. (2005). The fate of the Amazonian areas of endemism. *Conservation Biology*, *19*, 689–694.

Davidson, E. A., de Araújo, A. C., Artaxo, P., Balch, J. K., Brown, I. F., C. Bustamante, M. M., … Wofsy, S. C. (2012). The Amazon basin in transition. *Nature*, *481*, 321–328.

Ferrão, M., Colatreli, O., de Fraga, R., Kaefer, I. L., Moravec, J., & Lima, A. P. (2016). High species richness of Scinax treefrogs (Hylidae) in a threatened Amazonian landscape revealed by an integrative approach. *PLoS ONE*, *11*, e0165679.

Ficetola, G. F., Rondinini, C., Bonardi, A., Baisero, D., & Padoa-Schioppa, E. (2015). Habitat availability for amphibians and extinction threat: A global analysis. *Diversity and Distributions*, *21*, 302–311.

Foden, W. B., Butchart, S. H. M., Stuart, S. N., Vié, J.-C., Akçakaya, H. R., Angulo, A., … Mace, G. M. (2013). Identifying the world's most climate change vulnerable species: A systematic trait-based assessment of all birds, amphibians and corals. *PLoS ONE*, *8*, e65427.

Fontaneto, D., Kaya, M., Herniou, E. A., & Barraclough, T. G. (2009). Extreme levels of hidden diversity in microscopic animals (Rotifera) revealed by DNA taxonomy. *Molecular Phylogenetics and Evolution*, *53*, 182–189.

Fouquet, A., Gilles, A., Vences, M., Marty, C., Blanc, M., & Gemmell, N. J. (2007). Underestimation of species richness in neotropical frogs revealed by mtDNA analyses. *PLoS ONE*, *2*, e1109.

Fouquet, A., Martinez, Q., Courtois, E. A., Dewynter, M., Pineau, K., Gaucher, P., … Kok, P. J. R. (2013). A new species of the genus *Pristimantis* (Amphibia, Craugastoridae) associated with the moderately elevated massifs of French Guiana. *Zootaxa*, *3750*, 569–586.

Fouquet, A., Martinez, Q., Zeidler, L., Courtois, E. A., Gaucher, P., Blanc, M., … Kok, P. J. R. (2016). Cryptic diversity in the *Hypsiboas semilineatus* species group (Amphibia, Anura) with the description of a new species from the eastern Guiana Shield. *Zootaxa*, *4084*, 79–104.

Fouquet, A., Noonan, B. P., Rodrigues, M. T., Pech, N., Gilles, A., & Gemmell, N. J. (2012). Multiple quaternary refugia in the Eastern Guiana Shield revealed by comparative phylogeography of 12 frog species. *Systematic Biology*, *61*, 461–489.

Fouquet, A., Orrico, V. G. D., Ernst, R., Blanc, M., Martinez, Q., Vacher, J.-P., … Ron, S. (2015). A new *Dendropsophus* Fitzinger, 1843 (Anura: Hylidae) of the *parviceps* group from the lowlands of the Guiana Shield. *Zootaxa*, *4052*, 39–64.

Fouquet, A., Vacher, J.-P., Courtois, E. A., Villette, B., Reizine, H., Gaucher, P., … Kok, P. J. R. (2018). On the brink of extinction: Two new species of *Anomaloglossus* from French Guiana and amended definitions of *Anomaloglossus degranvillei* and *A. surinamensis* (Anura: Aromobatidae). *Zootaxa*, *4379*, 1–23.

Funk, W. C., Caminer, M., & Ron, S. R. (2012). High levels of cryptic species diversity uncovered in Amazonian frogs. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1806–1814.

Gibson, L., Lee, T. M., Koh, L. P., Brook, B. W., Gardner, T. A., Barlow, J., … Sodhi, N. S. (2011). Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature*, *478*, 378–381.

Godinho, M. B. C., & da Silva, F. R. (2018). The influence of riverine barriers, climate, and topography on the biogeographic regionalization of Amazonian anurans. *Scientific Reports*, *8*, 3427.

Goldstein, P. Z., & DeSalle, R. (2011). Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. *BioEssays*, *33*, 135–147.

Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic Models. *Journal of Statistical Software*, *1*(13), 2011.

Guedes, T. B., Sawaya, R. J., Zizka, A., Laffan, S., Faurby, S., Pyron, R. A., … Antonelli, A. (2017). Patterns, biases and prospects in the distribution and diversity of Neotropical snakes. *Global Ecology and Biogeography*, *27*, 14–21.

Hazzi, N. A., Moreno, J. S., Ortiz-Movliav, C., & Palacio, R. D. (2018) Biogeographic regions and events of isolation and diversification of the endemic biota of the tropical Andes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), 7985–7990. 201803908.

Herkt, K. M. B., Skidmore, A. K., & Fahr, J. (2017). Macroecological conclusions based on IUCN expert maps: A call for caution. *Global Ecology and Biogeography*, *26*, 930–941.

Hickerson, M. J., Stahl, E. A., & Lessios, H. A. (2006). Test for simultaneous divergence using approximate bayesian computation. *Evolution & Development*, *60*, 2435–2453.

Hollowell, T., & Reynolds, R. P. (2005). Checklist of the terrestrial vertebrates of the Guiana Shield. *Bulletin of the Biological Society of Washington*, *13*, 1–98.

Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., … Rahbek, C. (2013). An update of Wallace's zoogeographic regions of the world. *Science*, *339*, 74–78.

Hurlbert, A. H., & Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 13384–13389.

Jansen, M., Bloch, R., Schulze, A., & Pfenninger, M. (2011). Integrative inventory of Bolivia's lowland anurans reveals hidden diversity. *Zoologica Scripta*, *40*, 567–583.

Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences*, *110*, E2602–E2610.

Jungfer, K.-H., Faivovich, J., Padial, J. M., Castroviejo-Fisher, S., Lyra, M. M., Berneck, B. V. M., … Wheeler, W. (2013). Systematics of

spiny-backed treefrogs (Hylidae: *Osteocephalus*): An Amazonian puzzle. *Zoologica Scripta*, *42*, 351–380.

Kalamandeen, M., Gloor, E., Mitchard, E., Quincey, D., Ziv, G., Spracklen, D., … Galbraith, D. (2018). Pervasive rise of small-scale deforestation in Amazonia. *Scientific Reports*, *8*, 1600.

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple sequence alignment software Version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780.

Kok, P. J. R., Dezfoulian, R., Means, D. B., Fouquet, A., & Barrio-Amorós, C. L. (2018). Amended diagnosis and redescription of *Pristimantis marmoratus* (Boulenger, 1900) (Amphibia: Craugastoridae), with a description of its advertisement call and notes on its breeding ecology and phylogenetic relationships. *European Journal of Taxonomy* , *397*, 1–30.

Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, *37*, 2029–2053.

Krishnamurthy, P. K., & Francis, R. A. (2012). A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation*, *21*, 1901–1919.

Lambin, E. F., & Meyfroidt, P. (2011). Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 3465–3472.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*, 18–22.

Lin, X., Stur, E., & Ekrem, T. (2015). Exploring genetic divergence in a species-rich insect genus using 2790 DNA barcodes. *PLoS ONE*, *10*, e0138993.

Lomolino, M. V., Riddle, B. R., & Whittaker, R. J. (2016). *Biogeography* (5th ed.). Sunderland, MA, USA: Sinauer.

Lujan, N. K., & Armbruster, J. W. (2011). The Guiana Shield. In J. Albert & R. Reis (Eds.), *Historical biogeography of neotropical freshwater fishes* (pp. 211–224). Berkeley, CA, USA: University of California Press.

Machado, D. J., Castroviejo-Fisher, S., & Grant, T. (2019). Evidence of absence treated as absence of evidence: The effects of variation in the number and distribution of gaps treated as missing data on the results of standard maximum likelihood analysis. *bioRxiv*, 1–8.

Malhi, Y., Roberts, J. T., Betts, R. A., Killeen, T. J., Li, W., & Nobre, C. A. (2008). Climate change, deforestation, and the fate of the Amazon. *Science*, *319*, 169–172.

Mayer, M., da Fonte, L. F. M., & Lötters, S. (2019). Mind the gap! A review of Amazonian anurans in GenBank. *Salamandra*, *55*, 89–96.

Mayle, F. E., & Power, M. J. (2008). Impact of a drier Early–Mid-Holocene climate upon Amazonian forests. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 1829–1838.

Motta, J., Menin, M., Almeida, A. P., Hrbek, T., & Farias, I. P. (2018). When the unknown lives next door: A study of central Amazonian anurofauna. *Zootaxa*, *4438*, 79.

Naka, L. N., Bechtoldt, C. L., Magalli, H. L. P., & Brumfield, R. T. (2012). The role of physical barriers in the location of avian suture zones in the Guiana Shield, northern Amazonia. *The American Naturalist*, *179*, E115–E132.

Noonan, B. P., & Gaucher, P. (2005). Phylogeography and demography of Guianan harlequin toads (*Atelopus*): Diversification within a refuge. *Molecular Ecology*, *14*, 3017–3031.

Oliveira, U., Vasconcelos, M. F., & Santos, A. J. (2017). Biogeography of Amazon birds: Rivers limit species composition, but not areas of endemism. *Scientific Reports*, *7*, 2992.

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. *BioScience*, *51*, 933–938.

Orrico, V. G. D., Nunes, I., Mattedi, C., Fouquet, A., Lemos, A. W., Rivera-Correa, M., … Haddad, C. F. B. (2017). Integrative taxonomy supports the existence of two distinct species within *Hypsiboas crepitans* (Anura: Hylidae). *Salamandra*, *53*, 99–113.

Ortega-Andrade, H. M., Rojas-Soto, O. R., Valencia, J. H., Espinosa de los Monteros, A., Morrone, J. J., Ron, S. R., & Cannatella, D. C. (2015). Insights from integrative systematics reveal cryptic diversity in *Pristimantis* frogs (Anura: Craugastoridae) from the Upper Amazon Basin. *PLoS ONE*, *10*, e0143392.

Pebesma, E., & Bivand, R. (2005). Classes and methods for spatial data in R. *R News*, *5*(2), 9–13.

Pentinsaari, M., Vos, R., & Mutanen, M. (2017). Algorithmic single-locus species delimitation: Effects of sampling effort, variation and nonmonophyly in four methods and 1870 species of beetles. *Molecular Ecology Resources*, *17*, 393–404.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., … Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, *55*, 595–609.

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, *21*, 1864–1877.

Qian, H. (2009). Global comparisons of beta diversity among mammals, birds, reptiles, and amphibians across spatial scales and taxonomic ranks. *Journal of Systematics and Evolution*, *47*, 509–514.

R Development Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rahm, M., Thibault, P., Shapiro, A., Smartt, T., Paloeng, C., Crabbe, S., …Joubert, P. (2017). Monitoring the impact of gold mining on the forest cover and freshwater in the Guiana Shield.Reference year 2015.

Reid, N. M., & Carstens, B. C. (2012). Phylogenetic estimation error can decrease the accuracy of species delimitation: A Bayesian implementation of the general mixed Yule-coalescent model. *BMC Evolutionary Biology*, *12*, 196.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jiménez-Valverde, A., Ricotta, C., … Chiarucci, A. (2011). Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography*, *35*, 211–226.

Rojas, R. R., Fouquet, A., Ron, S. R., Hernández-Ruz, E. J., Melo-Sampaio, P. R., Chaparro, J. C., … Hrbek, T. (2018). A Pan-Amazonian species delimitation: High species diversity within the genus Amazophrynella (Anura: Bufonidae). *PeerJ*, *6*, e4941.

Rueda, M., Rodríguez, M. Á., & Hawkins, B. A. (2010). Towards a biogeographic regionalization of the European biota. *Journal of Biogeography*, *37*, 2067–2076.

Saatchi, S., Asefi-Najafabady, S., Malhi, Y., Aragão, L. E. O. C., Anderson, L. O., Myneni, R. B., & Nemani, R. (2013). Persistent effects of a severe drought on Amazonian forest canopy. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 565–570.

Salducci, M. D., Marty, C., Fouquet, A., & Gilles, A. (2005). Phylogenetic relationships and biodiversity in Hylids (Anura: Hylidae) from French Guiana. *Comptes Rendus Biologies*, *328*, 1009–1024.

Señaris, J. C., & MacCulloch, R. D. (2005). Amphibians: Checklist of the terrestrial vertebrates of the Guiana shield. *Bulletin of the Biological Society of Washington*, *13*, 9–23.

Silva, C. R., Ribas, C. C., Da Silva, M. N. F., Leite, R. N., Catzeflis, F., Rogers, D. S., & De Thoisy, B. (2018). The role of Pleistocene climate change in the genetic variability, distribution and demography of *Proechimys cuvieri* and *P. guyannensis* (Rodentia: Echimyidae) in northeastern Amazonia. *PLoS ONE*, *13*, e0206660.

Silva, F. R., Almeida-Neto, M., do Prado, V. H. M., Haddad, C. F. B., & de Cerqueira Rossa-Feres, D. (2012). Humidity levels drive reproductive

modes and phylogenetic diversity of amphibians in the Brazilian Atlantic Forest. *Journal of Biogeography*, *39*, 1720–1732.

Sommeria-Klein, G., Zinger, L., Coissac, E., Iribar, A., Schimann, H., Taberlet, P., & Chave, J. (2019). Latent Dirichlet Allocation reveals spatial and taxonomic structure in a DNA-based census of soil biodiversity from a tropical forest. *Molecular Ecology Resources*, *20*(2), 371–386.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313.

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 1607–1612.

Tracewski, Ł., Butchart, S. H. M., Di Marco, M., Ficetola, G. F., Rondinini, C., Symes, A., … Buchanan, G. M. (2016). Toward quantification of the impact of 21st-century deforestation on the extinction risk of terrestrial vertebrates. *Conservation Biology*, *30*, 1070–1079.

Vacher, J.-P., Kok, P. J. R., Rodrigues, M. T., Lima, J. D., Lorenzini, A., Martinez, Q., … Fouquet, A. (2017). Cryptic diversity in Amazonian frogs: Integrative taxonomy of the genus *Anomaloglossus* (Amphibia: Anura: Aromobatidae) reveals a unique case of diversification within the Guiana Shield. *Molecular Phylogenetics and Evolution*, *112*, 158–173.

Valle, D., Albuquerque, P., Zhao, Q., Barberan, A., & Fletcher, R. J. (2018). Extending the Latent Dirichlet Allocation model to presence/absence data: A case study on North American breeding birds and biogeographical shifts expected from climate change. *Global Change Biology*, *24*(11), 5560–5572.

Valle, D., Baiser, B., Woodall, C. W., & Chazdon, R. (2014). Decomposing biodiversity data using the Latent Dirichlet Allocation model, a probabilistic multivariate statistical method. *Ecology Letters*, *17*, 1591–1601.

Vasconcelos, T. S., Prado, V. H. M., da Silva, F. R., & Haddad, C. F. B. (2014). Biogeographic distribution patterns and their correlates in the diverse frog fauna of the Atlantic Forest hotspot. *PLoS ONE*, *9*, e104130.

Vedovato, L. B., Fonseca, M. G., Arai, E., Anderson, L. O., & Aragão, L. E. O. C. (2016). The extent of 2014 forest fragmentation in the Brazilian Amazon. *Regional Environmental Change*, *16*, 2485–2490.

Vences, M., Thomas, M., van der Meijden, A., Chiari, Y., & Vieites, D. R. (2005). Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Frontiers in Zoology*, *2*, 1–12.

Vieites, D. R., Wollenberg, K. C., Andreone, F., Köhler, J., Glaw, F., & Vences, M. (2009). Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences of the USA*, *106*, 8267–8272.

Vilhena, D. A., & Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nature Communications*, *6*, 6848.

White, A. E., Dey, K. K., Mohan, D., Stephens, M., & Price, T. D. (2019). Regional influences on community structure across the tropical-temperate divide. *Nature Communications*, *10*, 2646.

Yu, G., Rao, D., Matsui, M., & Yang, J. (2017). Coalescent-based delimitation outperforms distance-based methods for delineating less divergent species: The case of Kurixalus odontotarsus species group. *Scientific Reports*, *7*, 16124.

Zeisset, I., & Beebee, T. J. C. (2008). Amphibian phylogeography: A model for understanding historical aspects of species distributions. *Heredity*, *101*, 109–119.

BIOSKETCH

Antoine Fouquet is a researcher at the French National Center for Scientific Research. His research interests include systematics, biogeography and evolution of amphibians and reptiles, with a focus on frogs of the Guiana Shield.

Author contributions: J.P.V. and A.F. wrote the manuscript, conceived the study design, performed the molecular work, performed the ABGD analysis and wrote the manuscript with input from all other authors. J.C. helped to write the manuscript. F.F. helped with the biogeographic analysis and contributed to the manuscript. G.S.K. performed the LDA analyses and contributed to the manuscript. S.T. performed the random forest analyses and contributed to the manuscript. C.T. helped edit the manuscript. M.B., A.C., J.C., T.J.C., M.D., R.E., P.G., J.O.G., R.J., P.J.R.K., J.D.L., Q.M., C.M., B.P.N., P.N., P.O., R.R., M.T.R., A.S. and S.M.S. helped supply the samples and edit the manuscript.

SUPPORTING INFORMATION

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

How to cite this article: Vacher J-P, Chave J, Ficetola F, et al. Large-scale DNA-based survey of frogs in Amazonia suggests a vast underestimation of species richness and endemism. *J Biogeogr.* 2020;47:1781–1791. https://doi.org/10.1111/jbi.13847