

## COMPLETENESS OF DIGITAL ACCESSIBLE KNOWLEDGE OF PLANTS OF BENIN AND PRIORITIES FOR FUTURE INVENTORY AND DATA DISCOVERY

JEAN COSSI GANGLO\* AND SUNDAY BERLIOZ KAKPO

*Laboratory of Forest Sciences, Faculty of Agricultural Sciences, University of Abomey-Calavi, 03 BP 393, Cotonou, Benin. \*Corresponding author: [ganglocj@gmail.com](mailto:ganglocj@gmail.com).*

*Abstract.*—Discovery of and access to primary biodiversity data are critical components in informed decision-making regarding sustainable use of biological resources and conservation of biodiversity. Primary biodiversity data are increasingly available from Benin, but information about completeness of this information across the country is still lacking for most groups. This study analyzed the Digital Accessible Knowledge regarding the plants of Benin to identify gaps in both geographic and environmental dimensions. Many gaps exist in plant data for Benin, particularly in the northern most departments; central and southern Benin are better known, but some gaps remain even there. The resulting view of Beninese Digital Accessible Knowledge can guide future inventory and data discovery efforts.

*Key words.*—Benin, inventory completeness, data cleaning, digital accessible knowledge, gaps, GBIF.

Primary biodiversity data are data that combine three essential attributes: geographic location, collection date, and taxonomic identification (Johnson, 2007; Soberón and Peterson, 2009; Sousa-Baena et al. 2013). Digital accessible knowledge (DAK) is the part of the existing primary biodiversity data that is (1) in digital format, (2) published and accessible worldwide for free, and (3) integrated into the broader global storehouse of biodiversity information (in essence, a transformation of data into “knowledge”; Sousa-Baena et al. 2013). DAK is essential in enabling decision-making on natural resource management, especially biodiversity conservation (Peterson et al. 2000; Peterson et al. 2001; Peterson et al. 2004; Kremen et al. 2008; Morris et al. 2013; Gaiji et al. 2013). The Global Biodiversity Information Facility (GBIF<sup>1</sup>) is by far the largest initiative assembling and sharing DAK on biodiversity, with the aim of sustaining scientific research, conservation, and sustainable development (Gaiji et al. 2013, Peterson et al. 2015). GBIF was created in 2001, and—as of May 2016—provides access to >640M records from more than 1.6M species.

Although plant collections have accumulated from Benin since the 1780s (Akoegninou et al., 2006), digitization and publication of primary occurrence data within the country began only recently (2010), in the framework of the activities of the Global Biodiversity Information Facility (GBIF) through GBIF-Benin<sup>2</sup>. Some DAK for

Benin was available previously through other sources<sup>3</sup>: indeed, institutions in 22 countries including Benin presently contribute data about Beninese biodiversity, with 170 occurrence datasets and >190,000 records. DAK on the biodiversity of Benin is of great importance to guiding and shaping priorities of decision-makers and engaging them in sustainable management of the limited and increasingly stressed natural resources of the country.

### *Biodiversity Assessments versus DAK Availability in Benin*

Many studies have addressed biodiversity, in terms of its collections, patterns, and conservation, at various levels across Benin. Adomou (2005) studied distributional patterns of vegetation types of Benin, and divided the country into 10 phytogeographic districts, and described the major vegetation types and conservation priorities. Various individual vegetation types have been studied in terms of plant communities and species composition (Ganglo et al. 1999; Ganglo and Lejoly 1999; Sokpon et al. 2001; Ganglo 2004; Ganglo 2005; Ganglo and De Foucault 2006; Tohngodo et al. 2006; Awokou et al. 2009; Noumon and Ganglo 2005; Noumon et al. 2009; Aoudji et al. 2011; Yévidé et al. 2011), resulting in classifications of forest sites, and useful recommendations for sustainable management. The diversity of neglected and underutilized species was assessed by Dansi et al. (2012), who found 41 species of high importance in terms of nutrient content, medicinal value,

<sup>1</sup> <http://www.gbif.org>.

<sup>2</sup> <http://gbif-benin.org>.

<sup>3</sup> <http://www.gbif.org/country/BJ/about/datasets>.

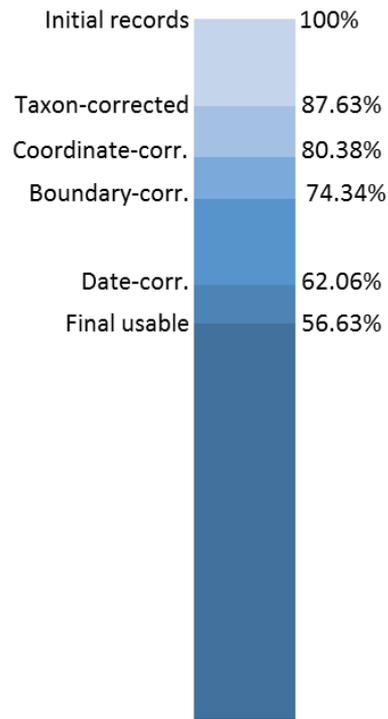


Figure 1. Loss of DAK records during the data-cleaning process for data regarding occurrences of the plants of Benin.

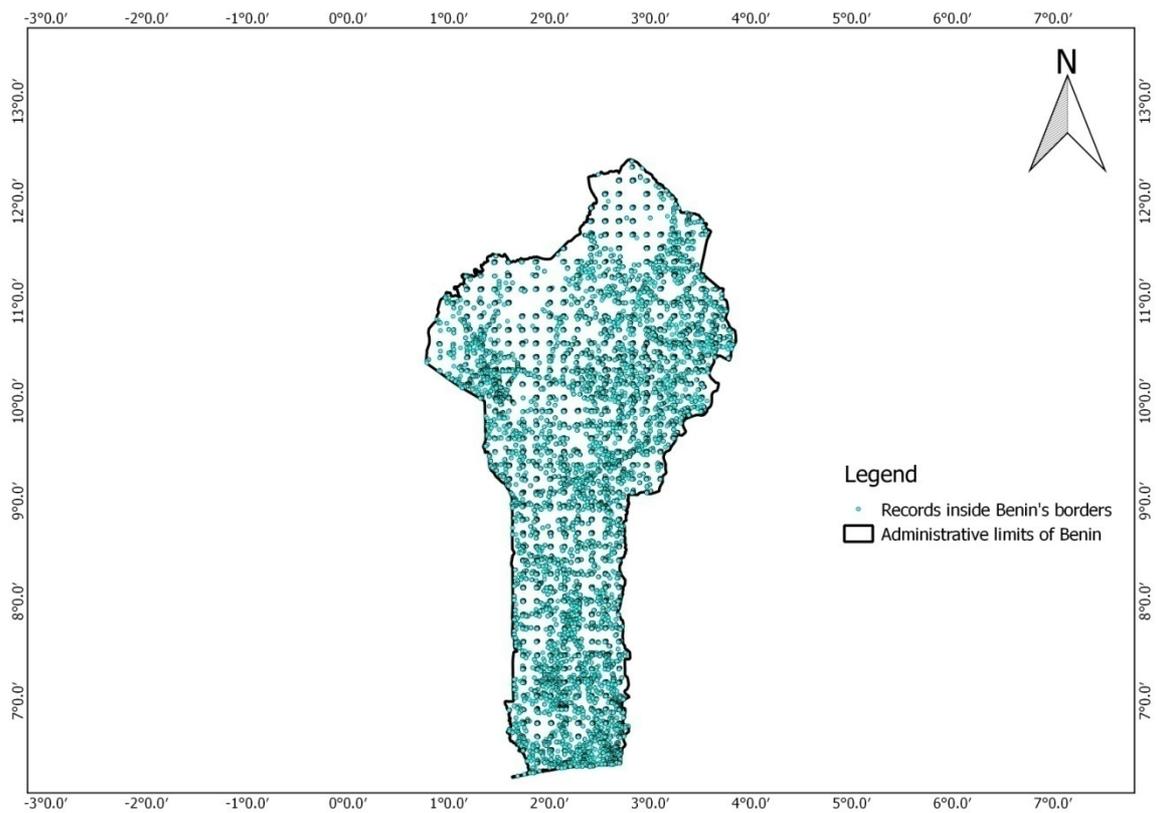


Figure 2. Spatial distribution of georeferenced plant records of Benin.

contribution to household income, poverty reduction, extent of consumption, degree of consumption, extent of production, availability during the year, contribution to empowerment of women, market value, and market use. Several stands and populations of forest species, medicinal plants, and multipurpose species have been studied in Benin in terms of structural characteristics, ecology, and usefulness (Sokpon and Biao 2002; Sokpon et al. 2006; Gouwakinou et al. 2009; Yêhouénu Tessi et al. 2012; Koura et al. 2011; Koura et al. 2013a; Koura et al. 2013b). The spatial genetic structure of baobab (*Adansonia digitata*) populations from West African agroforestry systems was evaluated by Kyndt et al. (2009): 11 populations from Benin, Ghana, Burkina Faso, and Senegal had comparable levels of genetic diversity, and the organization of genetic diversity appeared to result essentially from spatially restricted gene flow, with some influences of human seed exchange.

Although of importance in advancing knowledge of vegetation types and structure, and their determinants and conservation priorities, these studies did not address representativeness of sampling of plant diversity across the country, which could help identify knowledge gaps, data resource limitations, and possible biases (Sousa-Baena et al. 2013). The single exception is the palms (Arecaceae), which were analyzed in detail in terms of completeness across Benin by Idohou et al. (2015). Moreover, apart from scientific collections (Ariño, 2010), which covered the whole country during the elaboration of the flora of Benin (Akoegninou et al. 2006), most of the studies stated above only addressed limited sectors of the country. Worse yet, few, if any, of the occurrence data from those surveys are openly available to the scientific community.

A survey of biodiversity data holders and users within Benin (GBIF Benin, 2011) showed that the most important collection in terms of number of specimens is the national herbarium (50,000 specimens), but <30% of its specimens have been digitized and published by Beninese institutions via GBIF-Benin and GBIF. Therefore, the representativeness of the DAK with regard to the geography and species diversity of the country remains unclear. This study aims to develop a detailed assessment of DAK for Benin's plants to assess its fitness for use in detailed analyses.

## METHODS

### *Data cleaning*

Data used in this paper were downloaded from the GBIF site in January 2015 using the filter "Plantae" in the scientific name field on the Benin page<sup>4</sup>. To 'clean' the data, to make optimal use of information available, we used an iterative series of cleaning steps, as follows. (1) We created lists of unique taxon names in each dataset in Microsoft Excel, and inspected them for multiple versions of the same taxonomic concepts: misspellings, name variants, different versions of authority information, synonyms, etc. Such name variants were flagged, and checked with independent sources; a field was created to hold standardized preferred scientific names that correctly corresponded to single taxa. We used the list-matching service of the Catalogue of Life<sup>5</sup> and PROTA<sup>6</sup>, both of which were accessed in 2015. (2) We checked for and removed incomplete or suspicious geographic coordinates (e.g., one or both of the coordinates either missing or falling outside of Benin in spite of being referred to that country). (3) Within the country, we checked for consistency between textual descriptions of departments (administrative divisions) and the position of geographic coordinates. In each case, where problems were detected, we created a corrected version of the data records; where no clear correction was possible, we discarded records, recording data losses at each step in the process. Finally, (4) we discarded data records for which full information on year, month, and day of collection was lacking; we created a unique 'stamp' of time as year\_month\_day.

### *Inventory completeness*

Next, we aggregated point-based occurrence data to 0.5° spatial resolution across the country. This spatial resolution was the product of detailed analyses of balancing the benefits of aggregating data (i.e., larger sample sizes) versus the loss of spatial resolution that accompanies broader aggregation areas and can make imperceptible important geographic features (i.e., 0.5° resolution is a square ~55 km on a side). Those steps focused on choice of an optimal spatial resolution for analysis are detailed in another paper (Ariño et al., in preparation).

We produced the aggregation grid shapefiles in the Vector Grid module of QGIS, version 2.6,

<sup>4</sup> <http://www.gbif.org/occurrence/search?country=BJ#>.

<sup>5</sup> <http://www.catalogueoflife.org/listmatching>.

<sup>6</sup> <http://www.prota4u.info/protaindex.asp>.

Table 1. Inventory completeness in the departments of Benin.

Departments	Percentage of grid cells or objects with no data (a)	Percentage of grid cells or objects not well sampled with $N < 200$ and $C < 0.5$ (b)	Percentage of grid cells or objects not well sampled with $N \geq 200$ and $C < 0.5$ (c)	Total (a) + (b) + (c)	Percentage of grid cells or objects well sampled with $0.5 \leq C < 0.6$ (d)	Percentage of grid cells or objects well sampled with $0.6 \leq C < 0.7$ (e)	Percentage of grid cells or objects well sampled with $0.7 \leq C < 0.8$ (f)	Percentage of grid cells or objects well sampled with $C \geq 0.8$ (g)
Alibori	1.27	9.28	0.00	10.55	0.00	0.00	1.27	2.95
Borgou	1.27	5.91	0.00	7.18	0.00	0.00	6.33	2.53
Atakora	0.42	5.06	0.00	5.48	0.00	0.00	5.91	2.11
Kouffo	0.00	2.11	0.00	2.11	0.00	0.00	1.27	2.53
Plateau	2.11	0.00	0.00	2.11	0.00	0.00	1.27	1.69
Collines	0.00	1.69	0.00	1.69	0.00	1.69	2.11	3.80
Donga	0.00	1.27	0.00	1.27	0.00	0.00	2.11	3.38
Zou	0.00	0.84	0.00	0.84	0.00	0.00	2.11	5.06
Mono	0.00	0.42	0.00	0.42	0.00	1.27	1.69	2.53
Atlantique	0.00	0.00	0.00	0.00	0.00	2.11	1.69	5.06
Ouémé	0.00	0.00	0.00	0.00	0.00	1.27	0.42	3.38
Littoral	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.00

Table 2. Inventory completeness in the municipalities of Benin.

Department	Municipality	Percentage of grid cells or objects with no data (a)	Percentage of grid cells or objects not well sampled with $N < 200$ and $C < 0.5$ (b)	Percentage of grid cells or objects not well sampled with $N \geq 200$ and $C < 0.5$ (c)	Total (a) + (b) + (c)	Percentage of grid cells or objects well sampled with $0.5 \leq C < 0.6$ (d)	Percentage of grid cells or objects well sampled with $0.6 \leq C < 0.7$ (e)	Percentage of grid cells or objects well sampled with $0.7 \leq C < 0.8$ (f)	Percentage of grid cells or objects well sampled with $C \geq 0.8$ (g)
Alibori	Banikoara	0.00	2.11	0.00	2.11	0.00	0.00	0.00	0.42
	Gogounou	0.00	2.11	0.00	2.11	0.00	0.00	0.00	0.42
	Kandi	0.00	1.69	0.00	1.69	0.00	0.00	0.42	0.42
	Karimama	0.42	1.69	0.00	2.11	0.00	0.00	0.00	0.42
	Malanville	0.42	0.84	0.00	1.26	0.00	0.00	0.42	0.84
	Segbana	0.42	0.84	0.00	1.26	0.00	0.00	0.42	0.42
Atakora	Boukoumbé	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.42
	Cobly	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.00
	Kérou	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.00
	Kouandé	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.42
	Matéri	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.00
	Natitingou	0.00	0.00	0.00	0.00	0.00	0.00	1.27	0.42
	Péhunco	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.42
	Tanguiéta	0.42	0.00	0.00	0.42	0.00	0.00	0.84	0.00
Toucountouna	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42	
Atlantique	Abomey-Calavi	0.00	0.00	0.00	0.00	0.00	0.84	0.00	0.84
	Allada	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	Kpomassè	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.42
	Ouidah	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.42
	Sô-Ava	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.42
	Toffo	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.42

	Tori-Bossito	0.00	0.00	0.00	0.00	0.00	0.00	0.42	1.27
	Zè	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.84
Borgou	Bembéréké	0.00	1.27	0.00	1.27	0.00	0.00	0.00	0.42
	Kalalé	0.84	1.27	0.00	2.11	0.00	0.00	0.00	0.42
	N'Dali	0.00	0.84	0.00	0.84	0.00	0.00	1.69	0.42
	Nikki	0.42	0.42	0.00	0.84	0.00	0.00	0.42	0.42
	Parakou	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.00
	Pèrèrè	0.00	0.42	0.00	0.42	0.00	0.00	0.84	0.00
	Sinendé	0.00	0.84	0.00	0.84	0.00	0.00	0.42	0.42
	Tchaourou	0.00	0.84	0.00	0.84	0.00	0.00	2.11	0.42
Collines	Bantè	0.00	0.84	0.00	0.84	0.00	0.42	0.42	0.00
	Dassa-Zoumè	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.84
	Glazoué	0.00	0.00	0.00	0.00	0.00	0.42	0.84	1.27
	Ouèssè	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.42
	Savalou	0.00	0.84	0.00	0.84	0.00	0.42	0.00	0.42
	Savè	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.84
Donga	Bassila	0.00	0.84	0.00	0.84	0.00	0.00	0.84	0.84
	Copargo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.84
	Djougou	0.00	0.42	0.00	0.42	0.00	0.00	0.84	1.27
	Ouaké	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.42
Kouffo	Aplahoué	0.00	0.84	0.00	0.84	0.00	0.00	0.42	0.42
	Djakotomey	0.00	0.42	0.00	0.42	0.00	0.00	0.42	0.00
	Klouékanmè	0.00	0.42	0.00	0.42	0.00	0.00	0.00	0.42
	Lalo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.84
	Toviklin	0.00	0.42	0.00	0.42	0.00	0.00	0.42	0.84
Littoral	Cotonou	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.00
Mono	Athiémé	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.42
	Bopa	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42

	Comè	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.42
	Dogbo	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.42
	Grand-Popo	0.00	0.42	0.00	0.42	0.00	0.42	0.42	0.42
	Houéyogbé	0.00	0.00	0.00	0.00	0.00	0.42	0.42	0.42
Ouémé	Adjarra	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.42
	Adjohoun	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	Aguégués	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.42
	Akpro-Missérété	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	Avrankou	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	Bonou	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.42
	Dangbo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	Porto-Novo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	Sèmè-Kpodji	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.42
Plateau	Adja-Ouèrè	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.42
	Ifangni	0.42	0.00	0.00	0.42	0.00	0.00	0.00	0.42
	Kétou	0.84	0.00	0.00	0.84	0.00	0.00	0.42	0.42
	Pobè	0.42	0.00	0.00	0.42	0.00	0.00	0.42	0.00
	Sakété	0.42	0.00	0.00	0.42	0.00	0.00	0.00	0.42
Zou	Abomey	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	Agbangnizoun	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	Bohicon	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	Covè	0.00	0.00	0.00	0.00	0.00	0.00	0.42	1.27
	Djidja	0.00	0.84	0.00	0.84	0.00	0.00	0.00	0.84
	Ouinhi	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.00
	Za-Kpota	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.42
	Zagnanado	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.42
	Zogbodomey	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.84

added the coarse-resolution grid identification codes to each occurrence datum, and aggregated each datum into the coarse-resolution grid squares. In Excel, we explored relations between data on species identity, time (i.e, taking days as a unit of sampling effort), and aggregation grid square. We calculated (1) total number of records available from each grid square (termed  $N$ ), (2) total number of distinct species recorded from each grid square ( $S_{obs}$ ), (3) number of species detected on exactly one day ( $a$ ), and (4) number of species detected on exactly two days ( $b$ ). Via equations provided by Chao (1987), we calculated the expected number of species ( $S_{exp}$ ), as

$$S_{exp} = S_{obs} + \frac{a^2}{2b}$$

and inventory completeness ( $C$ ) as  $C = S_{obs} / S_{exp}$ .

We explored plots of  $C$  versus  $N$  to establish appropriate and adequate definitions of relatively completely versus incompletely inventoried grid squares. Once we had established criteria under which grid squares would be considered as well-sampled, in QGIS, we linked the table with the grid square statistics (i.e.,  $N$ ,  $S_{obs}$ ,  $S_{exp}$ ,  $C$ ) to the aggregation grid, and saved this file as a shapefile. We created a shapefile of well-sampled grid squares, which we in turn converted to raster (geotiff) format using custom scripts in R. This raster coverage was the basis for our identification of gaps in coverage of geographic and environmental spaces, as follows.

We used the Proximity (Raster Distance) function in QGIS to summarize geographic distance across the country to any well-sampled area, at  $0.05^\circ$  spatial resolution. To create a parallel view of environmental (climatic) difference from well-sampled areas, we used a detailed protocol, as follows. We plotted 5000 random points across the country, and linked each point to the geographic distance raster and to raster coverages ( $2.5'$  spatial resolution) summarizing annual mean temperature and annual precipitation drawn from the WorldClim climate data archive (Hijmans et al., 2005) via the Point Sampling Tool in QGIS.

We exported the attributes table associated with the random points, and imported it into Excel. We standardized values of each environmental variable to the overall range of the variable among the random points as  $(x_i - x_{min}) / (x_{max} - x_{min})$ , where  $x_i$  is the particular observed value in question. We then created a matrix of

Euclidean distances in the two-dimensional climate space, relating all of the points with a geographic distance  $>0$  to all of the points with geographic distance of zero; the latter represent points falling in well-sampled regions, whereas the former are scattered across the entire country. Points in well-sampled regions were assigned (by definition) environmental distances of zero.

Finally, environmental distances were imported into QGIS, and linked back to the random point's shapefile. This shapefile provided a broad sampling across the country, with a  $z$ -value that is the environmental distance associated with that point. To convert this vector-format dataset to raster format, with values across the entire region, we used inverse distance weighting, with a distance coefficient of 2.0. We explored these results further via relating them to Benin's municipalities to provide local contexts for future inventories.

We further explored impacts of roads, waterways, and protected areas on data completeness of the country. This exploration was achieved by intersecting completeness ( $C$ ) at  $0.5^\circ$  spatial resolution with data layers summarizing distributions of roads, waterways, and protected areas across Benin. These data layers were downloaded from open sources for roads, waterways<sup>7</sup>, and protected areas<sup>8</sup>. The attributes tables of intersection layers were used to calculate respectively the length of roads and waterways as well as the surface of protected areas associated to each grid cell. We could then calculate the coefficients of correlation between  $C$  values and each of the parameters calculated. Finally, we explored, in a preliminary manner, the impacts of absence of roads, waterways, and protected areas on data completeness.

## RESULTS

Initial numbers of plant records downloaded from GBIF comprised 148,944 primary occurrence records. After cleaning and removing duplicates and records of exotic species, we were left with 84,350 records (56.6% of the original data set) corresponding to 3188 species (Figure 1). Of the original total of records, 87.6% were identified to the species level, and 80.4% had adequate geographic coordinates. After checking coordinates and displaying the records against the administrative limits of Benin (Figure 2), only 74.3% of records were located within Benin's

<sup>7</sup> <http://www.diva-gis.org/gdata>.

<sup>8</sup> <http://www.protectedplanet.net/country/BJ#>.

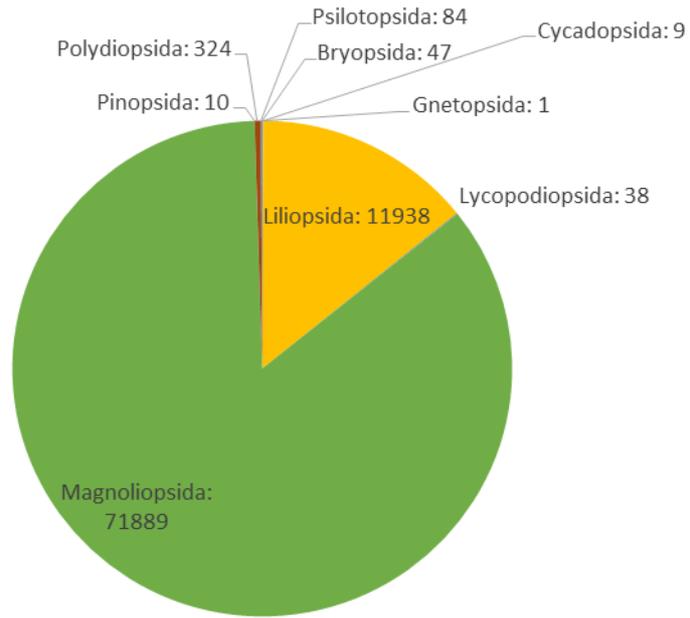


Figure 3. Floristic composition of DAK for plants of Benin in terms of representation of plant classes.

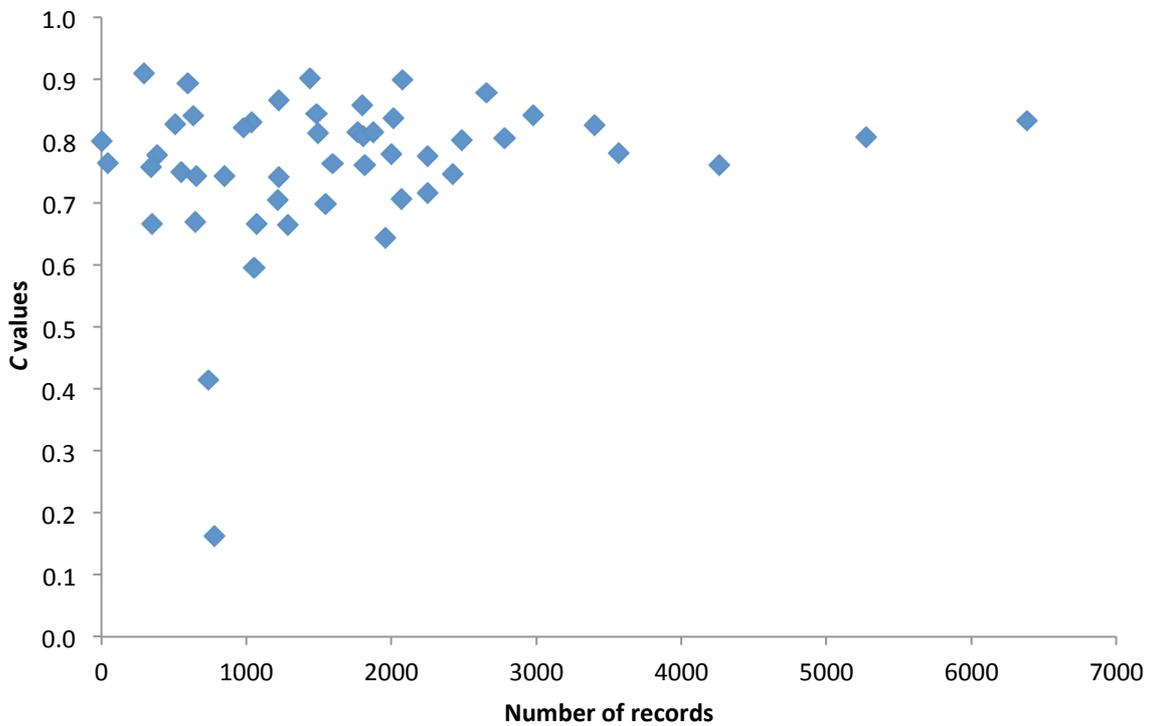


Figure 4. Inventory completeness (*C* values) as a function of numbers of records available within 0.5° grid squares across Benin.

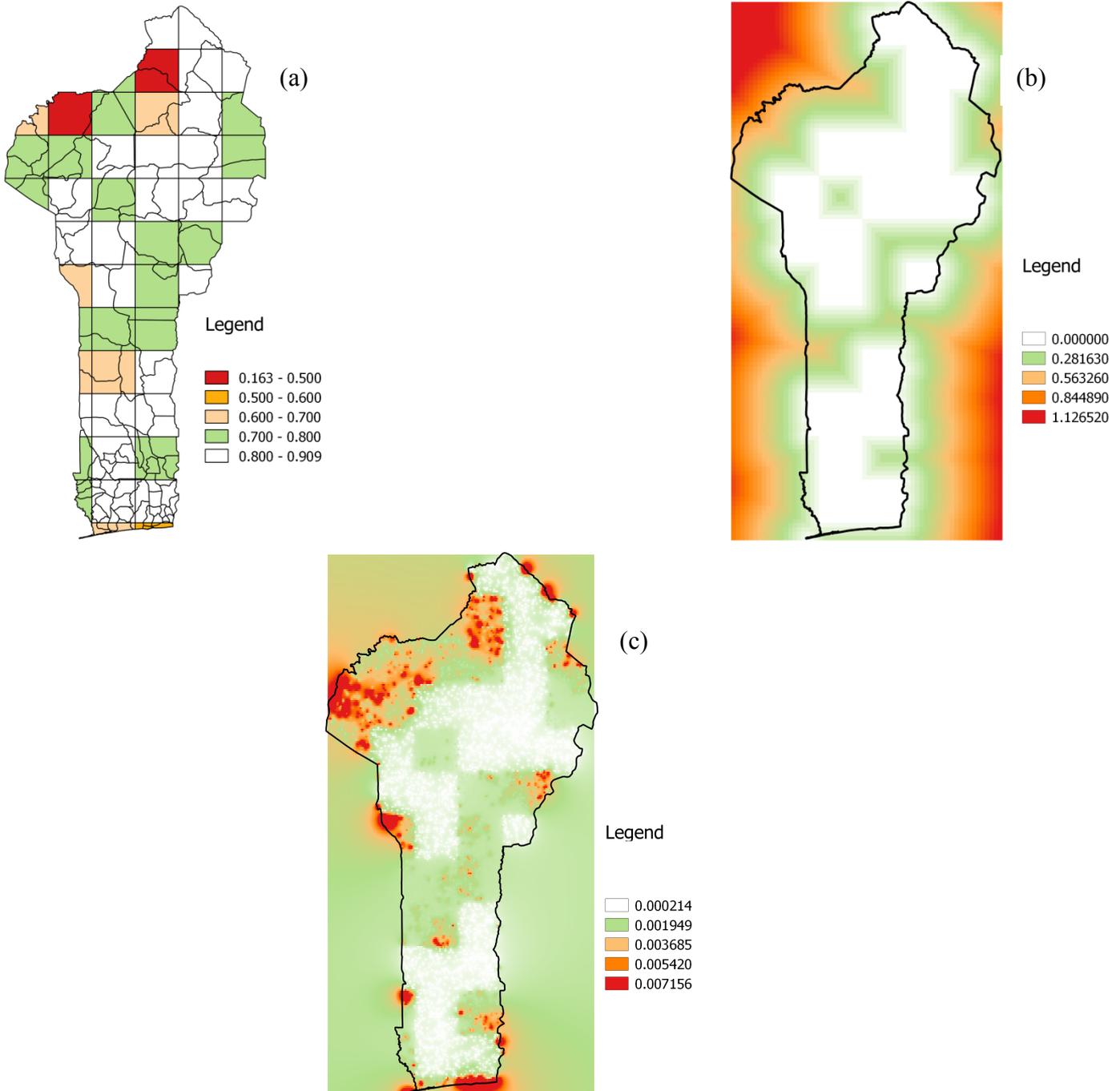


Figure 5. Geographic patterns of inventory completeness across Benin (a) 0.5° spatial resolution; (b) geographic distance to the most well-known cells ( $N \geq 200$  and  $C \geq 0.80$ ); (c) environmental distance to the most well-known cells.

borders and 62.1% had full date information. Most species belonged to the classes Magnolipsida (85.2% of records) and Liliopsida (14.2%; Figure 3). The occurrence data corresponded to species in 188 families, principally Fabaceae (25.1% of records), Combretaceae (9.8%), Rubiaceae (6.6%), Sapotaceae (6.3%), and Poaceae (5.1%).

We inspected the relationship between  $C$  and numbers of records ( $N$ ) at  $0.5^\circ$  spatial resolution (Figure 4). The correlation was weak and not significant ( $R = 0.203$ ,  $P = 0.157$ ). To account for artifactual  $C$  values, (grid cells with high  $C$  values (i.e.,  $>0.7$ ) for few record numbers), and taking into account the relationship between  $C$  values and  $N$  (Figure 4), we considered grid cells as well-sampled when  $N \geq 200$  records and  $C \geq 0.5$ .

The analysis of inventory completeness across Benin using the well-sampled criteria set above ( $N \geq 200$  records and  $C \geq 0.5$ ) showed that only two grid cells in northern west Benin are not well known (Figure 5a). To be more consistent with the status and reality of data completeness of the country, we considered grid cells as best-known when  $N \geq 200$  records and  $C \geq 0.8$ . Using the latter criteria we achieved the distribution patterns of inventory completeness at  $0.5^\circ$  spatial resolution with the corresponding distribution of geographic and environmental distances (Figure 5).

The analysis of inventory completeness at the level of departments shows that most departments of northern Benin (Alibori, Borgou, Atakora) had the largest percentages of grid cells with no data or not well-sampled (5-11%), whereas the best-known departments ( $C \geq 0.8$ ) were in southern and central Benin (Table 1, Figure 5). Analysis of inventory completeness at the level of municipalities (a lower-level administrative subdivision nested within departments) showed that the municipalities of Banikoara, Gogounou, Karimama, Kandi, Mallanville, and Segbana (department of Alibori), and Kalalé and Bembéréké (department of Borgou) had the most grid cells with no data or not well-sampled (Table 2, Figure 5). Those municipalities thus represent priorities for future inventory efforts to fill the picture of inventory completeness across Benin. The best-known municipalities ( $C \geq 0.8$ ) were Djougou (department of Donga, northern Benin), Tori-Bossito (Atlantique, southern Benin), Glazoué (Collines, central Benin), and Covè (Zou, central Benin; Table 2, Figure 5).

The intersection of inventory completeness with roads, waterways, and protected areas suggested positive impacts of all three on plant data completeness of Benin (Table 3, Figure 6). Among grid cells holding roads, only 28.5% had no data or were not well-sampled, compared with 50% across all grid cells; presence of roads in grid cells also increased the percentage of well-known grids cells (34.8%) against only 22.4% across the country (Table 3, Figure 6a). However, the correlation between road length within grid cells and  $C$  values was weak and not significant ( $R = -0.027$ ;  $P = 0.501$ ).

The intersection of the waterways data layer with grid cells revealed that 35.8% of the latter had no data or are not well-sampled against 50% for the whole grid cells across the country; the presence of waterways in grid cells also increased the percentage of best-known grids cells (36.8%) against only 22.4% across the country (Table 3, Figure 6b). Absence of waterways also coincided with grid cells lacking data, although the correlation between the length of waterways and  $C$  values is weak and not significant ( $R = -0.033$ ;  $P = 0.342$ ). However, grid cells containing waterways are also those containing roads, so the relative impacts of the two factors remain unclear.

The intersection of protected areas layer with the grid cells revealed that 35.0% of the latter had no data or were not well-sampled, compared with 50% for the whole grid cells across the country. Presence of protected areas in grid cells also increased the percentage of well-known grids cells (39.4%) against only 22.4% across the country (Table 3, Figure 6c). However, the correlation between protected area and  $C$  was significant ( $R = -0.253$ ;  $P = 0.035$ ), but was negative in sign. Again, grid cells containing protected areas also held roads and waterways, so the relative impacts of the three factors are difficult to evaluate.

## DISCUSSION

### *Completeness at different spatial levels*

We considered a grid cell as well-sampled if it was represented by  $\geq 200$  records and if completeness was  $\geq 0.5$ . Although this threshold is low compared to the 1000 records per grid cell, at the same spatial resolution, used by Sousa-Baena et al. (2013) for analyses of plant communities in Brazil, our threshold is higher than that of Funk et al. (2005), who used 40 records per 50 x 50 km grid cell in their analyses across Guyana. The additional criterion of  $C \geq$

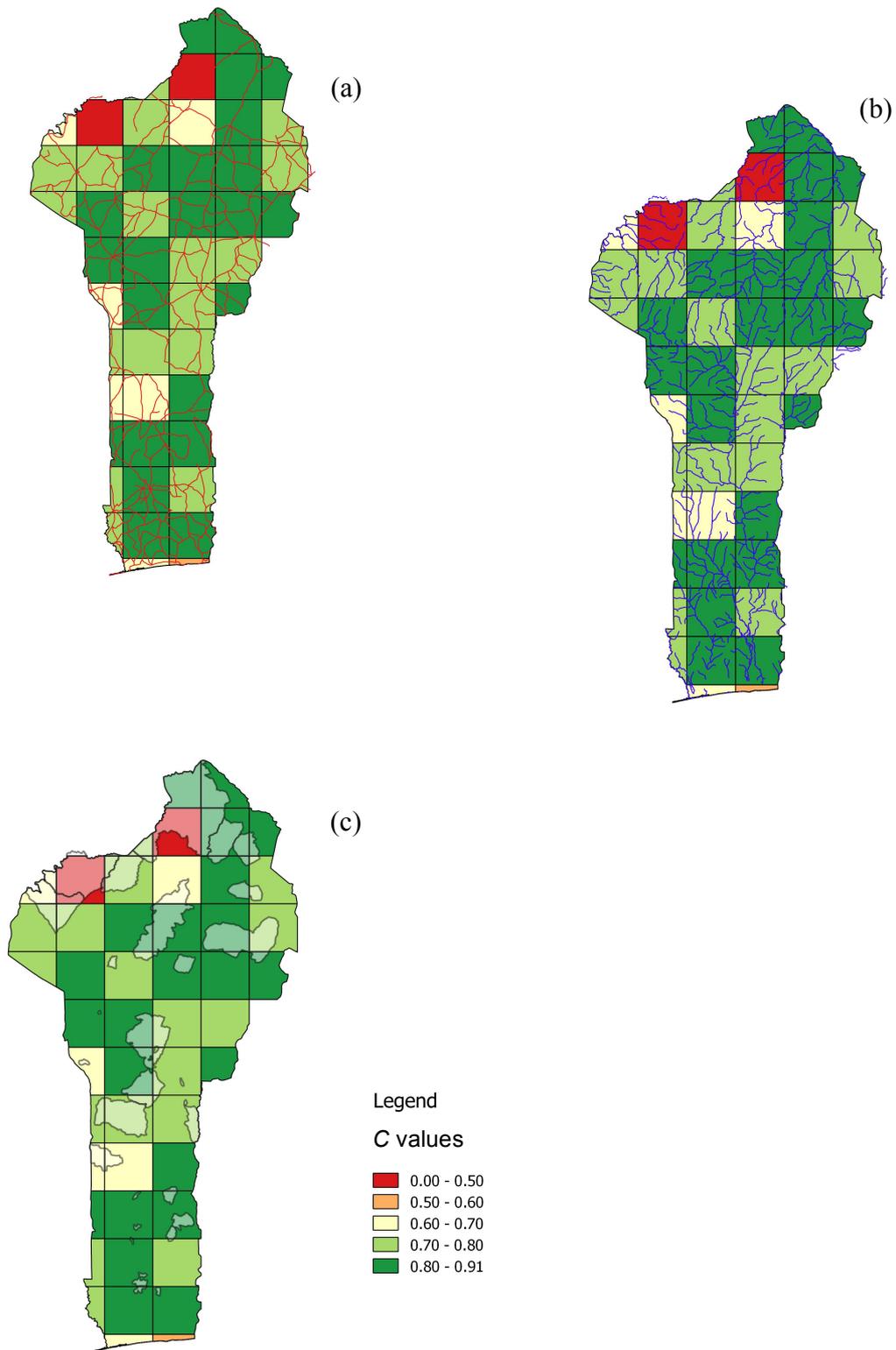


Figure 6. Geographic patterns of inventory completeness across Benin at 0.5° spatial resolution overlaid on spatial patterns of (a) roads, (b) waterways, and (c) protected areas.

**Table 3:** Impacts of protected areas, waterways, and roads on data completeness in Benin

Type of layer information	Percentage of grid cells or objects with no data (a)	Percentage of grid cells or objects not well sampled with $N < 200$ and $C < 0.5$ (b)	Percentage of grid cells or objects not well sampled with $N \geq 200$ and $C < 0.5$ (c)	Total (a) + (b) + (c)	Percentage of grid cells or objects well sampled with $0.5 \leq C < 0.6$ (d)	Percentage of grid cells or objects well sampled with $0.6 \leq C < 0.7$ (e)	Percentage of grid cells or objects well sampled with $0.7 \leq C < 0.8$ (f)	Percentage of grid cells or objects well sampled with $C \geq 0.8$ (g)
Half degree grid cells	12.07	37.93	0.00	50.00	1.72	5.17	20.69	22.41
Half degree grid cells inter protected areas	1.41	33.80	0.00	35.21	0.00	1.41	23.94	39.44
Half degree grid cells inter waterways	2.30	33.49	0.00	35.79	0.00	3.99	23.46	36.76
Half degree grid cells inter roads	1.77	26.77	0.00	28.54	0.00	5.65	30.97	34.84

0.5 for a grid cell to be considered as well inventoried coincides with the threshold used by Sousa-Baena et al. (2013) in their study of inventory completeness of plants of Brazil.

The analysis of inventory completeness for different departments and municipalities revealed that most departments of northern Benin, and—more precisely—the municipalities of Banikoara, Gogounou, Karimama, Kandi, Mallanville, and Segbana in the department of Alibori, and the municipalities of Kalalé and Bembéréké in the department of Borgou, were poorly known. These gaps stand in contrast to southern and central Benin, where municipalities were better-known ( $N \geq 200$ ,  $C \geq 0.8$ ). The municipalities of northern Benin listed above are therefore of highest priority for future inventory work in the country. Lack of infrastructure and institutions in the past in that region could explain our results in northern Benin, as well as insufficiency of publishing of occurrence data from the studies of vegetation in that part of the country. Indeed, in the region, the University of Parakou was created only in 2001, and its Faculty of Agronomy has initiated biodiversity studies only since that time.

Exploring inventory completeness at  $0.5^\circ$  spatial resolution, we observed a positive impact of roads on data completeness. These results coincide with those of Kadmon et al. (2004), who affirmed that bias in distributional data is common in the form of high concentrations of collection sites along roads. Souza-Baena et al. (2013) also found road bias effects on plant data completeness across Brazil, and Hijmans et al. (2000) reported that most gene bank accessions for wild potatoes in Bolivia were collected within 2 km of roads, 3-fold greater than random expectations. Ballesteros-Mejia et al. (2013) reported positive effects of infrastructure (traffic access, road density, tourism) on inventory completeness for tropical insects of Sub-Saharan Africa.

Analyses of distribution of DAK with respect to waterways also showed a positive impact of the latter on data completeness. Our further exploration revealed that the grid cells concerned here are served both by waterways and roads so that the positive impact of waterways on data completeness could be induced by the presence of roads instead.

Considering geographic patterns of inventory completeness in relation to protected areas, we found a significant impact of the latter on data completeness. The correlation of area protected and  $C$  was significant and negative, suggesting

that data completeness was more effective in protected areas with smaller surfaces. We inferred that efforts of data inventories were more and more diluted when the surface of protected areas increases. Our results support those of Ballesteros-Mejia et al. (2013), who reported positive effects of protected areas status on inventory completeness of sphinx moths (Sphingidae) of Sub-Saharan Africa.

#### *DAK Fitness for Use*

After cleaning, our data set corresponded to 3188 species. The total species richness of plant species of Benin has been estimated at 3200 species (Adjanohoun et al., 1989), but the most complete and recent flora of Benin included only 2807 species (Akoegninou et al., 2006). Our result can therefore complement the list of the flora of Benin of Akoegninou et al. (2006), via identification of several hundred species that have been recorded within the country.

Data fitness for use revolves around data precision, accuracy, and authenticity for specific uses (Faith et al. 2013). As such, an important initial observation is that  $>40\%$  of initial records downloaded from GBIF proved unavailable or incomplete, and thus were not included in our analyses. This amount is considerable in terms of loss of data—we emphasize that the data records lost were already DAK, and yet were not available for analysis owing to gaps or inconsistencies in content in the form of incomplete taxonomic determination, inadequate geographic coordinates, and missing time information. This result confirmed data concerns raised in feedback received by GBIF from its community of data users: the recent (2010) survey by GBIF's Content Needs Assessment Task Group (CNATG) revealed major concerns in terms of geographic and taxonomic gaps in data coverage, as well as the need for data quality assurance (Faith et al. 2013).

For Benin, 12.4% of records lacked usable species names, higher than the 9.5% found across the broader GBIF network (Gaiji et al., 2013), although we perhaps used a stricter set of criteria for our filtering. Records lacking coordinates in Benin (19.6%) were comparable to the 18.5% found by CNATG as of December 2010, but higher than the 14.1% reported in February 2012 (Gaiji et al. 2013). The percentage of records falling outside of Benin's borders (7.5% of georeferenced records) was high compared to the 3.6% reported by Gaiji et al. (2013). Numbers of records lacking full temporal information for

Benin (37.9%) was high compared to the 23.1% found by Sousa-Baena et al. (2013) for plant data in Brazil, and the 30.8% found by Gaiji et al. (2013) for GBIF-mediated data globally.

To improve DAK quality for the biodiversity of Benin, rigorous data capture protocols and detailed error detection and data cleaning workflows need to be implemented. These steps may depend in large part on sound capacity-building, such that data managers and publishers of data relevant to Benin can work more effectively. For instance, error rates in taxonomic names can be reduced massively by using authority data held in authority lists like the Catalogue of Life and PROTA as controlled vocabularies. Geographic coordinates can be resolved using tools like GEOLocate and BioGeomancer to add coordinates to well-described localities where data have been collected (Gaiji et al., 2013).

#### ACKNOWLEDGMENTS

The JRS Biodiversity Foundation has generously supported biodiversity informatics activities in Benin. We address our sincere gratitude to Town Peterson of the University of Kansas, who facilitated writing this paper through training in a course on national biodiversity diagnoses, and revised the manuscript profoundly. We also thank Kate Ingenloff and Lindsay Campbell for their assistance in analyses.

#### REFERENCES

- Adjanooun, E. J., V. Adjakidjè, M. R. A. Ahyi, L. Ake Assi, A. Akoegninou, J. d'Almeida, F. Apovo, K. Boukef, M. Chadare, G. Cusset, K. Dramane, J. Eyme, J.-N. Gassita, N. Gbaguidi, E. Goudoté, S. Guinko, P. Hounnon, L. Issa, A. Keita, H. V. Kiniffo, D. Koné-Bamba, A. Musampa Nseyya, M. Saadou, T. Sogandji, S. de Souza, A. Tchabi, C. Zinsou Dossa and T. Zohoun. 1989. Contribution aux Études Éthnobotaniques et Floristiques en République Populaire du Bénin. Agence de Coopération Culturelle et Technique (ACCT), Paris, France.
- Adomou, A. C., 2005. Vegetation Patterns and Environmental Gradients in Benin: Implications for Biogeography and Conservation. Ph.D. thesis, Wageningen University, Wageningen.
- Aoudji, A. K. N., I. S. A. Yévidé, J. C. Ganglo, G. Atindogbé, S. M. Toyi, C. De Canniere, H. A. Azontondé, V. Adjakidjè, B. De Foucault and A. B. Sinsin. 2011. Structural characteristics and forest sites identification in Pahou Forest Reserve, South-Benin. Bois et Forêts des Tropiques 308:47-58.
- Akoegninou, A., W. J. Van der Burg, L., J., G., Van der Maesen, V. Adjakidjè, J. P. Essou, B. Sinsin and H. Yédomonhan. 2006. Flore Analytique du Bénin. Backhuys Publishers. Wageningen.
- Ariño, A. H. 2010. Approaches to estimating the universe of natural history collections data. Biodiversity Informatics 7: 81-92.
- Awokou, K. S., J. C. Ganglo, H. A. Azontondé, V. Adjakidjè and B. De Foucault. 2009. Caractéristiques structurales et écologiques des phytocénoses forestières de la forêt classée d'Ichèdè (Département du Plateau, Sud-est Bénin). Science & Nature 6: 125-138.
- Ballesteros-Mejia, L., I. J. Kitching, W. Jetz, P. Nagel and J. Beck. 2013. Mapping the biodiversity of tropical insects: Species richness and inventory completeness of African sphingid moths. Global Ecology and Biogeography 22: 586-595.
- Chao, A. 1987. Estimating the population size for capture recapture data with unequal catchability. Statistica Sinica 10: 227-246.
- Chapman, A. D., M. E. S. Muñoz and I. Koch. 2005. Environmental information: Placing biodiversity phenomena in an ecological and environmental context. Biodiversity Informatics 2: 24-41.
- Collen, B., M. Ram, T. Zamin and L. McRae. 2008. The tropical biodiversity data gap: Addressing disparity in global monitoring. Tropical Conservation Science 1: 75-88.
- Dansi, A., R. Vodouhè, P. Azokpota, H. Yedomonhan, P. Assogba, A. Adjatin, Y. L. Loko, I. Dossou-Aminon and K. Akpagana. 2012. Diversity of the neglected and underutilized crop species of importance in Benin. Scientific World Journal 2012: 932947.
- Faith, D. P., B. Collen, A. H. Ariño, P. Koleff, J. Guinotte, J. Kerr and V. Chavan. 2013. Bridging biodiversity data gaps: Recommendations to meet users' data needs. Biodiversity Informatics 8: 41-58.
- Funk, V. A., K. S. Richardson and S. Ferrier. 2005. Survey-gap analysis in expeditionary research: Where do we go from here? Biological Journal of the Linnean Society 85: 549-567.
- Gaiji, S., V. Chavan, D. Hobern, A. H. Ariño, R. Sood, J. Otegui and E. Robles. 2013. Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. Biodiversity Informatics 8: 94-172.
- Ganglo, C. J., J. Lejoly and T. Pipar. 1999. Le teck (*Tectona grandis* L. f.) au Bénin, gestion et perspectives. Bois et Forêts des Tropiques 261: 17-27.
- Ganglo, C. J. and J. Lejoly. 1999. Biotope et valeur indicatrice écologique de l'association à *Lecaniodiscus cupanioides* et *Landolphia calabarica* dans le sous-bois naturel des teckeraies du sud-Bénin. Acta Botanica Gallica 146: 227-245.

- Ganglo, C. J. 2004. Phytosociologie Appliquée à l'Aménagement des Forêts: Cas du Périmètre Forestier de Toffo (Sud-Bénin, Département de l'Atlantique). Rapport Scientifique, Université d'Abomey-Calavi, Cotonou.
- Ganglo J. C. 2005. Groupements de sous-bois, identification et caractérisation des stations forestières: Cas d'un bois au Bénin. Bois et Forêts des Tropiques 285: 35-46.
- Ganglo, J. C. and B. de Foucault. 2006. Plant communities, forest site identification and classification in Toffo reserve, South-Benin. Bois et Forêts des Tropiques 288: 25-38.
- GBIF-Benin. 2011. Enquête sur les Détenteurs et les Utilisateurs de Données et d'Informations sur la Biodiversité au Bénin. GBIF Bénin, Cotonou.
- Gouwakinnou, N. G., Kindomihou, V., Assogbadjo, E. A. and B. Sinsin. 2009. Population structure and abundance of *Sclerocarya birrea* (A. Rich) Hochst subsp. *birrea* in two contrasting land-use systems in Benin. International Journal of Biodiversity and Conservation 1: 194-201.
- Johnson, N. F. 2007. Biodiversity informatics. Annual Review of Entomology 52: 421-38
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.
- Hijmans, R. J., K. A. Garrett, Z. Huaman, D. P. Zhang, M. Schreuder and M. Bonierbale. 2000. Assessing the geographic representativeness of gene bank collections: The case of Bolivian wild potatoes. Conservation Biology 14: 1755-1765.
- Idohou, R., A. H. Ariño, A. E. Assogbadjo, R. L. Glèlè Kakai and B. Sinsin. 2015. Diversity of wild palms (Arecaceae) in the Republic of Benin: Finding gaps in the national inventory combining field and digital accessible knowledge. Biodiversity Informatics 10: 45-55.
- Kadmon, R. and A. Danin. 2003. A systematic analysis of factors affecting the performance of climatic envelope models. Ecological Applications 13: 853-867.
- Kadmon, R., O. Farber and A. Danin. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications 14: 401-413.
- Koura, K., J. C. Ganglo, A. E. Assogbadjo and C. Agbangla. 2011. Ethnic differences in use values and use patterns of *Parkia biglobosa* in northern Benin. Journal of Ethnobiology and Ethnomedicine 7: 42.
- Koura, K., E. F. Dissou and J. C. Ganglo. 2013a. Caractérisation écologique et structurale des parcs à néré [*Parkia biglobosa* R. Br.] du département de la Donga au Nord-Ouest du Bénin. International Journal of Biological and Chemical Sciences 7: 726-738.
- Koura, K., Y. Mbaidé and J. C. Ganglo. 2013b. Caractéristiques phénotypique et structurale de la population de *Parkia biglobosa* R. Br. du Nord-Bénin. International Journal of Biological and Chemical Sciences.7: 2317-2327
- Kremen, C., J. Dransfield, Jr. E. Louis, M. Vences, A. Cameron, B. L. Fisher, R. A. Nussbaum, D. R. Vieites, A. Moilanen, F. Glaw, T. C. Good, C. J. Raxworthy, P. C. Wright, S. J. Phillips, G. J. Harper, C. D. Thomas, R. J. Hijmans, A. Razafimpahanana, M. L. Zjhra, H. Beentje, D. C. Lees and G. E. Schatz. 2008. Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. Science 320: 222-226.
- Kyndt, T., E. A. Assogbadjo, J. Olivier, J. O. Hardy, R. Glèlè Kakai, B. Sinsin, P. Van Damme and G. Gheysen. 2009. Spatial genetic structuring of baobab (*Adansonia digitata*, Malvaceae) in the traditional agroforestry systems of West Africa. American Journal of Botany 96: 950-957.
- Martínez-Meyer, E. 2005. Climate change and biodiversity: Some considerations in forecasting shifts in species' potential distributions. Biodiversity Informatics 2: 42-55
- Morris, R. A., V. Barve, M. Carausu, V. Chavan, J. Cuadra, C. Freeland, G. Hagedorn, P. Leary, D. Mozzherin, A. Olson, G. Riccardi, I. Teage and G. Whitbread. 2013. Discovery and publishing of primary biodiversity data associated with multimedia resources: The Audubon Core strategies and approaches. Biodiversity Informatics 8: 185-197.
- Noumon, J. C. and J. C. Ganglo. 2005. Phytosociologie appliquée à l'aménagement des forêts: Cas du périmètre forestier de Koto (Département du Zou, Centre-Bénin). Acta Botanica Gallica 152: 421-426.
- Noumon, J. C., J. C. Ganglo, H. A. Azontondé, B. De Foucault and V. Adjakidjè. 2009. Ecological and silvicultural indicator value of plant-communities of Koto Forest Reserve (Centre-Benin). International Journal of Biological and Chemical Sciences 3: 367-377.
- Peterson, A.T., S. L. Egbert, V. Sánchez-Cordero and K. P. Price. 2000. Geographic analysis of conservation priorities using distributional modeling and complementarity: Endemic birds and mammals in Veracruz, Mexico. Biological Conservation 93:85-94.
- Peterson, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor 103: 599-605.
- Peterson, A. T., M. A. Ortega-Huerta, J. Bartley, V. Sanchez-Cordero, J. Soberón, R. H. Buddemeier and D. R. B. Stockwell. 2002. Future projections for Mexican faunas under global climate change scenarios. Nature 416: 626-629.
- Peterson, A.T., E. Martínez-Meyer, C. González-Salazar and P. W. Hall. 2004. Modeled climate

- change effects on distributions of Canadian butterfly species. *Canadian Journal of Zoology* 82: 851-858.
- Peterson, A. T., J. Soberón and L. Krishtalka. 2015. A global perspective on decadal challenges and priorities in biodiversity informatics. *BMC Ecology* 15: 15.
- Soberón, J., R. Jiménez, J. Golubov and P. Koleff. 2007. Assessing completeness of biodiversity databases at different spatial scales. *Ecography* 30: 152-160.
- Soberón, J. and A. T. Peterson. 2009. Monitoring biodiversity loss with primary species-occurrence data: Toward national-level indicators for the 2010 Target of the Convention on Biological Diversity. *AMBIO* 38: 29-34.
- Sokpon, N., T. Sinadouwirou, F. Gbaguidi and H. S. Biaou. 2001. Les forêts édaphiques hygrophiles du Bénin. *Belgian Journal of Botany* 134: 79-93.
- Sokpon, N. and S. H. Biaou. 2002. The use of diameter distributions in sustained-use management of remnant forests in Benin: Case of Bassila Forest Reserve in North Benin. *Forest Ecology and Management* 161: 13-25.
- Sokpon, N., S. H. Biaou, C. Ouinsavi and O. Hunhyet. 2006. Bases techniques pour une gestion durable des forêts claires du Nord-Bénin: Rotation, diamètre minimal d'exploitabilité et régénération. *Bois et Forêts des Tropiques* 287: 45-57.
- Sousa-Baena, M. S., L. C. Garcia and A. T. Peterson. 2013. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions* 20: 369-381.
- Tohngodo, C., J. C. Ganglo, K. E. Agbossou, B. de Foucault and V. Adjakidjè. 2006. Caractéristiques structurales, identification et caractérisation des stations forestières de la forêt classée de Bonou (Sud-est Bénin). *Sciences & Nature* 3: 39-47.
- Yêhouénou Tessi, R. D., S. G. Akouèhou and J. C. Ganglo. 2012. Caractéristiques structurales et écologiques des populations de *Antiaris toxicaria* (Pers.) Lesch et de *Ceiba pentandra* (L.) Gaertn dans les forêts reliques du Sud-Bénin. *International Journal of Biological and Chemical Sciences* 6: 5056-5067.
- Yêvidé, I. S. A., J. C. Ganglo, K.A. Aoudji, S. M. Toyi, C. De Cannière, B. De Foucault, J.-L. Devineau and B. Sinsin. 2011. Caractéristiques structurelles et écologiques des phytocénoses de sous-bois des plantations privées de teck du département de l'Atlantique (Sud-Bénin, Afrique de l'Ouest). *Acta Botanica Gallica* 158: 263-283.