

Gap analysis: collections-based data for conservation actions: engaging decision-making actors to save globally threatened epiphytes in Colombia- Data mobilization and a framework of actions to promote further data mobilization

Daihana Arango

María Alexandra Guerra

Jennifer Calderón

Ana Maria Benavides

Botanical Garden Foundation "Joaquín Antonio Uribe" of Medellín

June 2022 v.2 ¹

The fundamental question we seek to contribute is what the conservation status of 400 species of epiphytes and lichens and 150 species of floral visitors is. To answer this question, we have first evaluated the list of epiphytes for Colombia and the status of the collections and records published in GBIF. Which allows us to understand where the data mobilization gaps are and guides us in directing the efforts of mobilization. Furthermore, this has allowed us to generate the first checklist of epiphytes for Colombia, which could be an additional product of the project as a data paper. Next, we describe the information gap analysis processes and the agreements with the herbaria partners.

Reconstructing the species checklist of vascular epiphytes of Colombia

This is the description of how the epiphyte species checklist of Colombia has been generated bases on five databases. A preliminary list was created based on three databases. First, using epiphytes and hemiepiphytes as keywords in growth habit a species list was produced in Catalog of Plants of Antioquia (Idárraga Piedrahíta et al. 2011) and Catalog of Plants and Lichens of Colombia (Bernal et al. 2015). In addition, a list of epiphytic ferns provided by Alejandra Vasco, a specialist in the group was included. With these databases a list was created in which duplicates were joined and eliminated. Next, we included the list of epiphytes for the world, EpiList 1.0 (Zotz et al. 2021), including those species distributed in Colombia using rgbif package (Chamberlain 2017). Additionally, epiphyte species for Colombia from GBIF were included in the list. In GBIF every record in Colombia that was called epiphyte was searched for both in Spanish as in English (epiphyte, trepadora, hemiepiphyte, epiphyte, hemiepiphyte), duplicates were removed.

In addition to consolidate a single list of species of the five lists, the name of the species was corroborated, and the synonyms were deleted, with lcvplant package (Freiberg et al. 2020, Figure 1). In the case of lichens, a GBIF list of the Peltigeraceae and Hygrophoraceae families was obtained, where the names were corroborated in TNRS (Boyle et al. 2013), and the number of records for Colombia was subsequently searched in GBIF.

In total for the five checklists, 6.810 species in 73 families were found for vascular plants. At least 1.848 species are endemic to Colombia. This preliminary list of Colombian epiphyte species will be refined and verified with specialists, to corroborate the epiphytic habit, clarifying in which cases the growth on a host tree has been accidentally-

¹ The most notorious differences since the previous report correspond to (a) two additional new families of ferns (Aspleniaceae and Hymenophyllaceae), (b) the search in GBIF was refined with epiphytic habit, so species list was reduced by more than 1000 species.

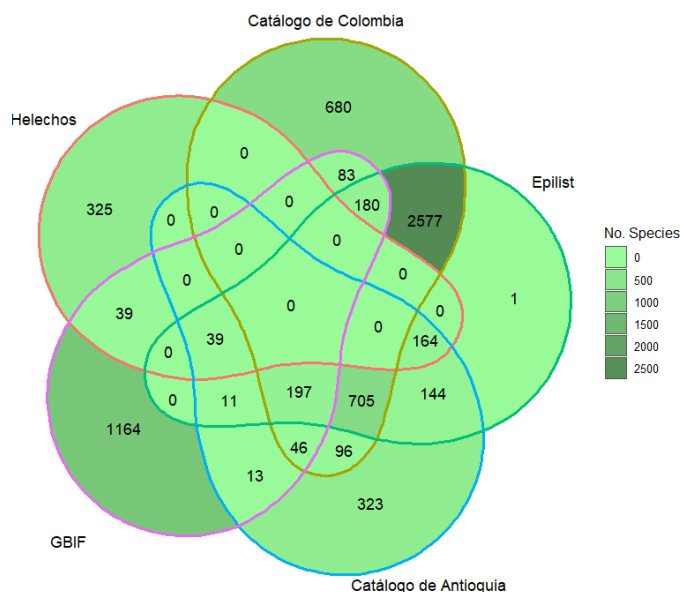


Figure 1 Venn diagram of the five databases used to make the unique list of epiphyte species for Colombia

Epiphyte species of Colombia represented in GBIF

Records of each species in Colombia, including both, records with coordinates and without coordinates were obtained with the rgbif package (Chamberlain 2017). Subsequently were selected the columns: ownerInstituion, phylum, kingdom, family, genus, scientificName, decimalLongitude, decimalLatitude, county, country, locality, municipality, occurrunceRemarks, stateProvince, institutionCode, catalogNumber.

This downloaded database was cleaned for consistency and technical errors (Table 1). Following the following workflow: (1) incorrectly entered fields, (2) several names were associated with the same entity or (3) department. As a result, the records for Colombia were obtained, which allows us to analyze the status of mobilization of epiphyte records in Colombia.

Table 1 Types of errors found in the databases downloaded from GBIF.

Column	Error	Type	Number of Excluded Records
locality	Some locations are misspelled, and fields are empty	Technical Consistency /	0
stateProvince	Some departments are misspelled, incomplete or do not correspond to a department that corresponds to Colombia Empty fields	Technical Consistency /	0
institutionCode	Some institutions are incomplete, or several names correspond to the same institution and empty fields	Technical Consistency /	0

decimalLatitude	Empty fields.	Technical	0
deciamLongitude	Empty fields	Technical	0
county	Some names are misspelled, and cells are empty	Technical Consistency	/ 0
ownerInstitutionCode	Empty fields	Technical	0

Based on 208.369 records in GBIF there were 64 % records with coordinates (133.334 records with coordinates) accounting for 4420 species. These records predominated in the Andean region of the country (Figure 2). Records without coordinates (75.035 records) accounted for 4.251 species. For lichens, a total of 14 species were found without coordinates and 86 species with coordinates. So far, most important problem detected in the GBIF records was the absence of coordinates for both lichens and vascular plants. The lack of georeferencing in biological records weakens research processes (Gomez et al. 2013). Accordingly, an effort should be made to contact the institutions to improve the quality of the data (including the coordinates).

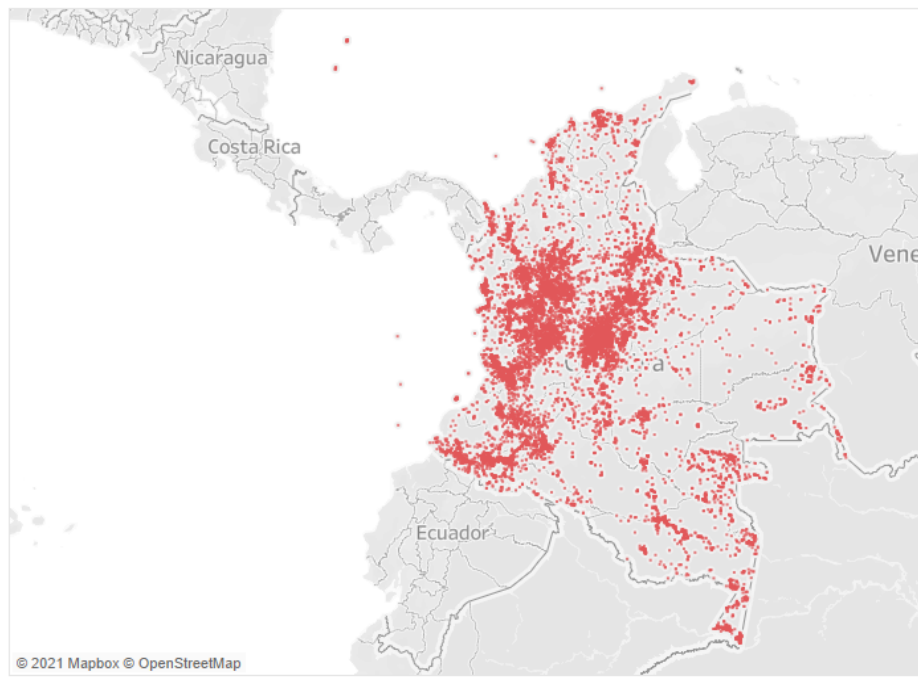


Figure 2 Map of distribution of epiphyte records in GBIF for Colombia.

To prepare the information for the development phase, in which extinction risk will be assess for the focused groups (vascular epiphytes: Cyclanthaceae, Araceae, Gesneriaceae, Piperaceae and ferns Dryopteridaceae, Lomariopsidaceae, Oleandraceae and Polypodiaceae), and lichens: (Hygrophoraceae, and Lobariaceae -now Peltigeraceae-). A preliminary analysis has been done in the whole checklist. A total of 85.876 records were obtained for plant families prioritized with coordinates and 49.605 records for plant families without coordinates. For lichens, 1.213 records with coordinates and 248 records without coordinates (Table 2).

Table 2 . Plant and lichen families prioritized with the number of records in GBIF

Families	Records with coordinates	No. species	Records without coordinates	No. species
Araceae	19762	571	5969	454
Aspleniaceae	3532	82	2620	82
Bromeliaceae	11219	413	5005	409
Cyclanthaceae	1761	64	690	63
Dryopteridaceae	3886	182	2266	182
Gesneriaceae	7047	212	4286	199
Hymenophyllaceae	3265	143	2503	144
Lomariopsidaceae	198	8	94	9
Oleandraceae	144	10	104	9
Orchidaceae	20232	2399.	16426	2982
Piperaceae	5037	217	3212	221
Polypodiaceae	9793	415	6430	414
Hygrophoraceae	211	21	99	7
Peltigeraceae (Lobariaceae)	1002	32	149	3
Total	87089		49853	

Based on partner herbaria (and institutions), Colombian Herbarium (COL) and Alexander von Humboldt Scientific Research Institute (FMB) were the institutions with highest number of published records (Figure 3). On the other hand, the Joaquín Antonio Uribe herbarium (JAUM) presented no published records (Table 3,

Figure 4). The Orchidaceae family was the most representative in all GBIF records associated herbaria followed by Araceae. The departments (provinces) in Colombia with the highest number of records were Antioquia, Boyacá, Cundinamarca, and Chocó. The lowest records were associated with Atlántico, Cesar, and La Guajira (Figure 5). Plant families with fewer records, such as Cyclanthaceae, Lomariopsidaceae, Oleandraceae and Piperaceae, need a greater systematization effort in the herbaria JAUM, CUVC and in the UDFJC (Figure 3). At the international level, the largest records are found in the COL, Missouri Botanical Garden (MO), SINCHI and NA', which means that there is an enormous number of records that are not associated with an institution. It should be clarified that, although these records have coordinates, during the validation of these, most of them did not correspond to their localities and also the records are not represented in all the departments of the country.

Table 3 Number of records for partner institutions and the status of herbarium collections according to the needs of the focused families

Institutions	Records with coordinates in GBIF (% of the total)	Records without coordinates in GBIF	Prioritized families in each herbarium	Already systematized	Records per families published in GBIF	
COL	16.659 (64%)	8.716	Bromeliaceae Cyclanthaceae Piperaceae	Orchidaceae Gesneriaceae	Araceae	4.375
					Aspleniaceae	1.863
					Bromeliaceae	1.935
					Cyclanthaceae	453

					Dryopteridaceae	1.384
					Gesneriaceae	2.824
					Hymenophyllaceae	1356
					Lomariopsidaceae	57
					Oleandraceae	71
					Orchidaceae	5.832
					Piperaceae	1.011
					Polypodiaceae	4.220
					Hygrophoraceae	-
					Peltigeraceae (Lobariaceae)	-
FMB	4.289 (78%)	1.347		Everything is systematized until 2017	Araceae	898
					Aspleniaceae	261
					Bromeliaceae	812
					Cyclanthaceae	69
					Dryopteridaceae	298
					Gesneriaceae	417
					Hymenophyllaceae	146
					Lomariopsidaceae	16
					Oleandraceae	13
					Orchidaceae	1476
					Piperaceae	451
					Polypodiaceae	740
					Hygrophoraceae	-
					Peltigeraceae (Lobariaceae)	-
HUA	3.486 (99%)	6	Ferns	Araceae	Araceae	1545
				Orchidaceae,	Aspleniaceae	-
				Bromeliaceae,	Bromeliaceae	519
				Gesneriaceae,	Cyclanthaceae	324
				Cyclanthaceae	Dryopteridaceae	-
				Lichens	Gesneriaceae	964
				Piperaceae	Hymenophyllaceae	-
					Lomariopsidaceae	-
					Oleandraceae	-
					Orchidaceae	102
					Piperaceae	37
					Polypodiaceae	-
					Hygrophoraceae	-
					Peltigeraceae (Lobariaceae)	-
CUVC	128 100%)	0	Pteridohytes		Araceae	57
			Araceae		Aspleniaceae	-
			Gesneriaceae,		Bromeliaceae	10
			Orchidaceae		Cyclanthaceae	-
					Dryopteridaceae	6
					Gesneriaceae	-
					Hymenophyllaceae	-
					Lomariopsidaceae	-
					Oleandraceae	-

					Orchidaceae	48
					Piperaceae	6
					Polypodiaceae	1
					Hydrophoraceae	-
					Peltigeraceae (Lobariaceae)	-
JAUM	0	0	Araceae, Bromeliaceae Cyclanthaceae	Orchidaceae Gesneriaceae Lichens Ferns	Without published records	
Total	21.698	8.939				

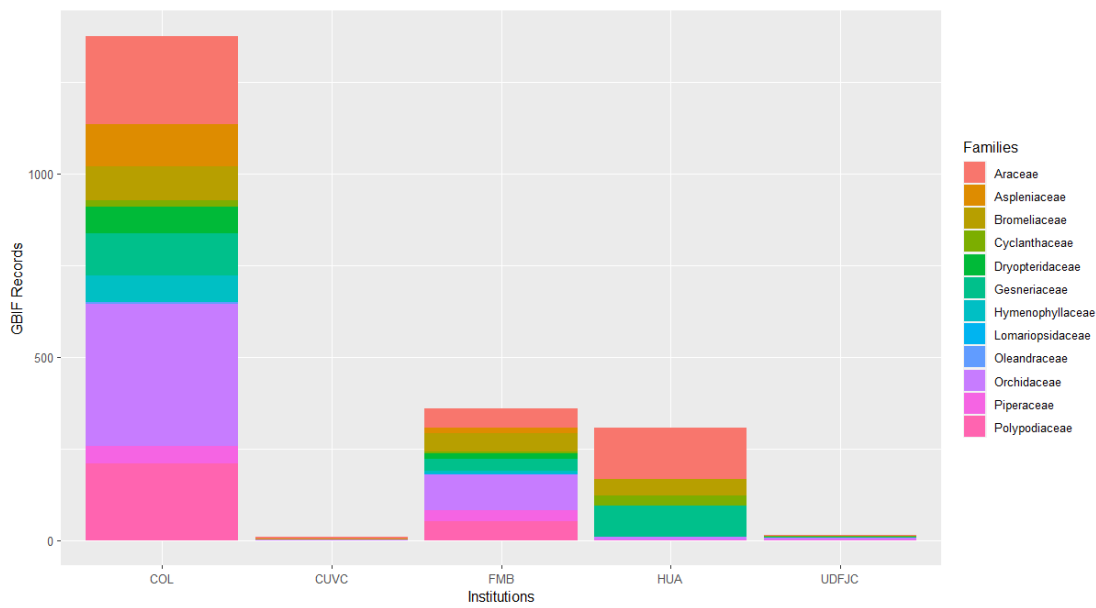


Figure 3. Number of records with coordinates by prioritized families in partner institutions

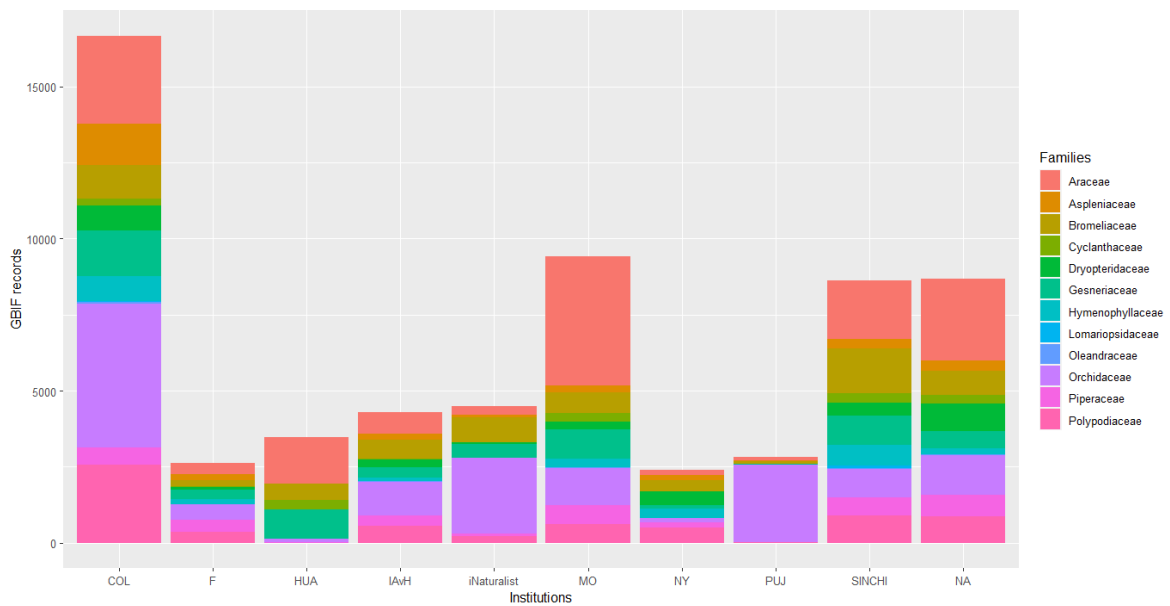


Figure 4. Number of records in GBIF with coordinates in institutions with records greater than 1000. COL: Colombian Herbarium, F: Field Museum of Natural History, IAvH: Instituto de Investigaciones Científicas

Alexander von Humboldt, HUA: Universidad de Antioquia Herbarium, MO: Missouri Botanical Garden, NY: The New York Botanical Garden, OBC: Oleoducto Bicentenario, PUJ: Pontificia Universidad Javeriana, SINCHI: Scientific Research Institute of the Amazon, NA: Not Applicable.

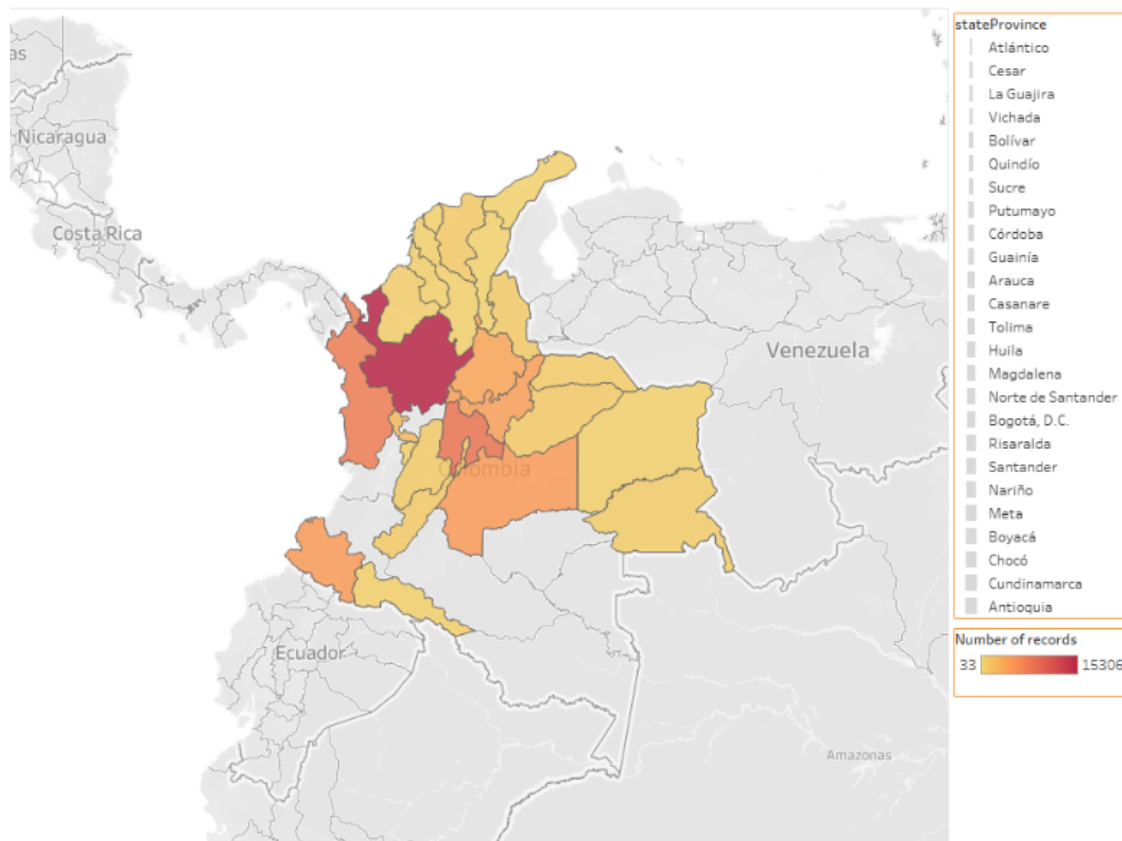


Figure 5. Number of records with coordinates by department for prioritized families

For records without coordinates, the institution with the highest number of records in herbarium corresponds to COL and IAvH (Figure 6). We noticed a greater number of records in the departments (provinces) of Meta, Cundinamarca, Antioquia, Chocó, and Valle del Cauca. It should be noted that the large number of records pertaining to Meta may be because the centroid of Colombia corresponds to this department. While the departments with the fewest records were from Arauca, Atlántico and Sucre (Figure 7), This may be since most of the epiphytes of the neotropics inhabit tropical rainforests distributed in the Amazon, Pacific and Andean regions. (Etter 1998; Arévalo 2003) and the necessary investigation in the most forgotten places in Colombia. Regarding the number of records in general, it can be observed that foreign institutions were more predominant than Colombian institutions, with the Smithsonian Institution (US) being the institution with most records without coordinates, followed by the COL (Figure 8).

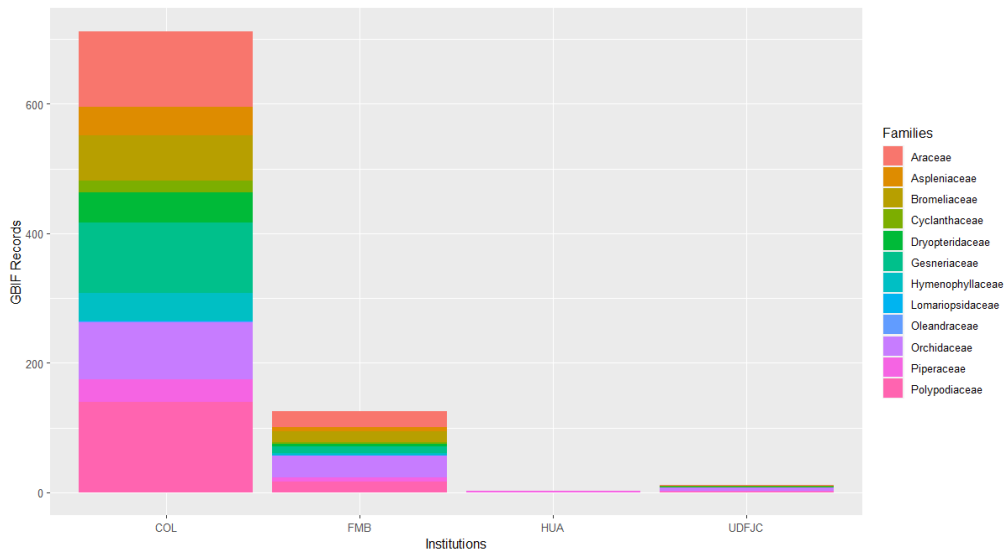


Figure 6 Number of records without coordinates by prioritized families in partner institutions.

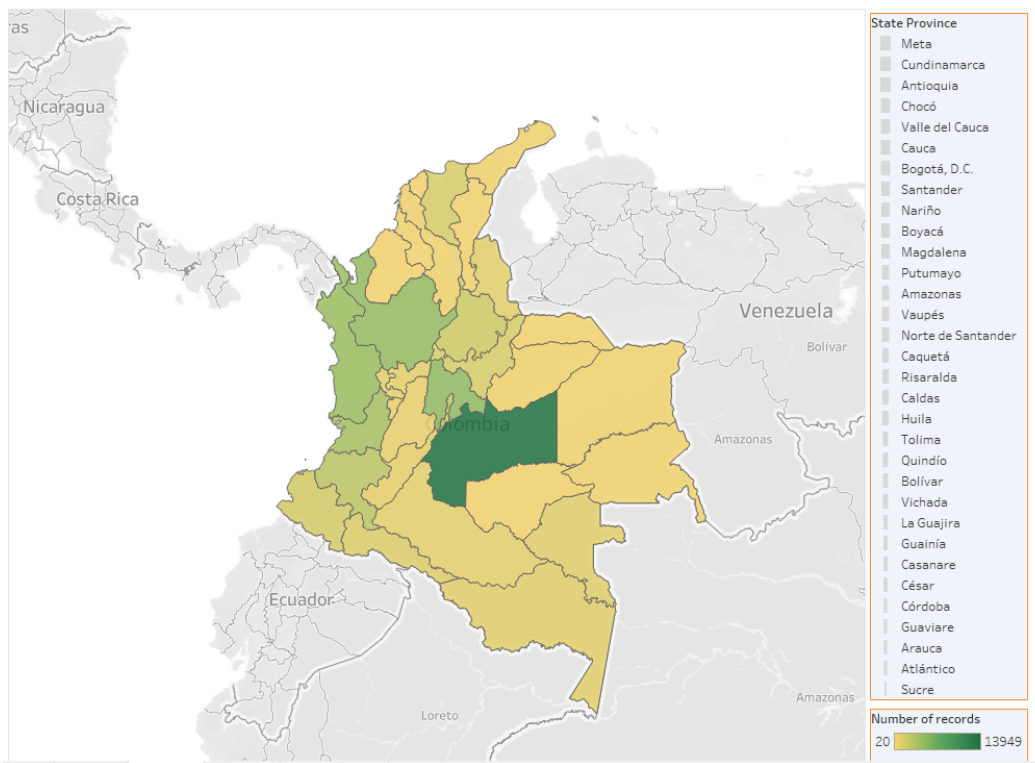


Figure 7. Records without coordinates by department in the prioritized families

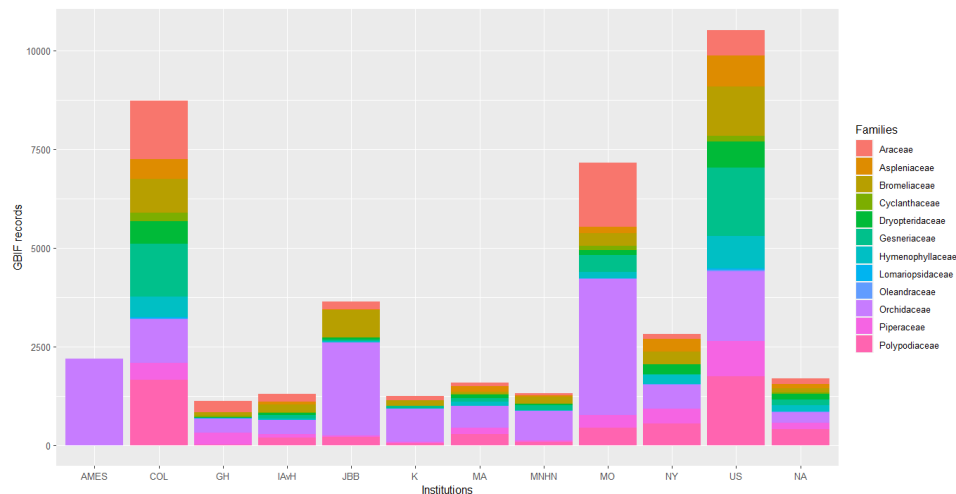


Figure 8. Number of records in GBIF without coordinates for institutions with records greater than 1.000. AMES: Oakes Ames Orchid Herbarium Harvard University, GH: Gray Herbarium Harvard University, IAvH: Alexander von Humboldt Scientific Research Institute, COL: Colombian Herbarium, JBB: Bogotá Botanical Garden, K: Royal Botanic Gardens, MA: Real Jardín Botánico, MNHN: Museum National d'Histoire Naturelle, MO: Missouri Botanical Garden, NY: The New York Botanical Garden, US: Smithsonian Institution, NA: Not Applicable.

About lichen records, all the downloaded records do not have coordinates, a large proportion of records do not have an associated institution and most records are published by foreign institutions. It should also be noted that the Hygrophoraceae family has a reduced number of records compared to Peltigeraceae (now Lobariaceae) (Figure 9).

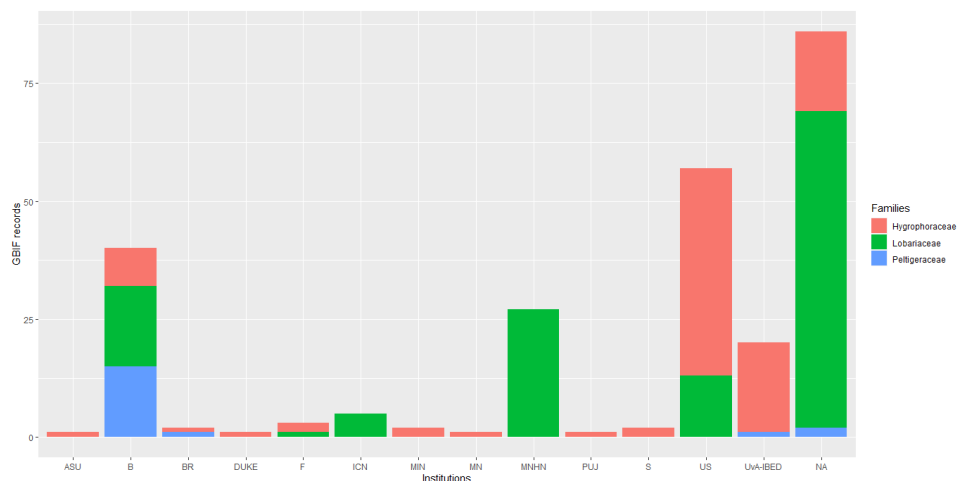


Figure 9. Number of records in GBIF without coordinates for lichens in the institutions that have been published. ASU: Arizona State University Biocollections, B: Botanic Garden and Botanical Museum Berlin, BR: Meise Botanic Garden, DUKE: Duke University Herbarium, F: Field Museum of Natural History, ICN: Instituto de Ciencias Naturales de la Universidad Nacional de Colombia, MA: Real Jardín Botánico, MIN: University of Minnesota Bell Museum, Herbário do Museu Nacional, MNHN: Museum National d'Histoire Naturelle, NY: The New York Botanical Garden, O: University of Oslo, PUJ: Pontificia Universidad Javeriana, S: Swedish Museum of Natural History, TNS: National Museum of Nature and Science, UFRGS: Universidade Federal do Rio Grande do Sul, US: Smithsonian Institution, UvA-IBED: University of Amsterdam / IBED, NA: Not Applicable.

Regarding the records with coordinates and according to dialogues established with each herbarium director, we have defined three principles to the process of mobilization: (1) the mobilization process and in particular the record systematization has included a family approach (not only epiphytic species), seeking to fill the information gaps of each herbarium. (2) each herbarium has prioritized a set of families, according to the state of systematization, focused on those with low or no systematization (Table 3) Each herbarium will maintain its ownership of the data, and the project team will facilitate the process to be made public through SiB Colombia.

First guidelines to promote further data mobilization and update information resources developed by the National Epiphytes Consortium

This project will contribute to the accomplishment of the policy frameworks in Colombia for conserving epiphytes and their floral visitors, which have identified the need to increase knowledge and assess the extinction risk of these groups of species. The analysis of information gaps and workflow, in addition, with conversations held with officials from the Colombian Ministry of the Environment, directors of the herbariums and specialists have allowed us to identify a guideline to promote further data mobilization and extinction risk analysis.

Guidelines to guarantee the mobilization of data in institutions, herbariums or biological collections

According with our experience,

- To have a good data quality each institution or herbarium, must incorporate standards and process of validation on real time to the reconstruction of coordinates, taxonomic categories, catalog number and duplicates. Using systematically tools like the Gbif data validator and the SiB Colombia validation routines for both Qgis and open refine have been useful to find errors and complete information in the data sets. Tools like datacanadensys for the conversion of coordinates and dates have been of great help.
- For institutions, a format with mandatory fields is recommended to detect errors more easily and have complete information.
- It is recommended that institutions have a person to verifier data in which errors are frequent such as collection dates that do not coincide with determination dates, coordinates that do not correspond to locality descriptions, errors even in the specimen description.
- It is important that collectors consider aspects such as: datum and associated error. In case when collectors do not have a GPS or cell phone in the field, be very specific in the description of the location.

References

- Arévalo R (2003) Composición y distribución de epífitas vasculares en cuatro bosques de la cuenca de Puerto Abeja (Chiribiquite, Amazonia Colombiana). 60
- Etter A (1998) Ecosistemas de Bosque Húmedo Tropical. Inf Nac sobre el Estado la Biodivers Colomb 106–133
- Bernal R, Gradstein SR, Celis M (2015) Catalog of Plants and Lichens of Colombia (Institute of Natural Sciences, National University of Colombia, Bogotá, Colombia). Available: <http://catalogoplantasdecolombia.unal.edu.co/es/>
- Boyle B, Hopkins N, Lu Z, et al (2013) The taxonomic name resolution service: An online tool for automated standardization of plant names. BMC Bioinformatics 14: <https://doi.org/10.1186/1471-2105-14-16>
- Chamberlain S (2017) rgbif: Interface to the Global “Biodiversity” Information Facility “API”. R package version 0.9.8. <https://CRAN.R-project.org/package=rgbif>

- Escobar D, Jojoa L, Díaz S, et al (2016) Georeferencing of locations: A reference guide for biological collections. 144
- Freiberg M, Winter M, Gentile A, et al (2020) LCVP, The Leipzig catalog of vascular plants, a new taxonomic reference list for all known vascular plants. *Sci Data* 7: <https://doi.org/10.1038/s41597-020-00702-z>
- Gomez SE, Torres S, Cáseres G, et al (2013) Georeferencing of the Museo de La Plata database using the Point-Radio method
- Idárraga Piedrahíta Á, Ortíz R del C, Callejas Posada R, Merello M (2011) Flora de Antioquia. Catalog of vascular plants vol. II List of vascular plants in the department of Antioquia. In: University of Antioquia
- Zotz G, Weigelt P, Kessler M, et al (2021) EpiList 1.0: a global checklist of vascular epiphytes. *Ecology* 102: <https://doi.org/10.1002/ecy.3326>