



**HAL**  
open science

# Mining rare Earth elements: Identifying the plant species most threatened by ore extraction in an insular hotspot

Guillaume Lannuzel, Léa Pouget, David Bruy, Vanessa Hequet, Shankar Meyer, Jérôme Munzinger, Gildas Gâteblé

## ► To cite this version:

Guillaume Lannuzel, Léa Pouget, David Bruy, Vanessa Hequet, Shankar Meyer, et al.. Mining rare Earth elements: Identifying the plant species most threatened by ore extraction in an insular hotspot. *Frontiers in Ecology and Evolution*, 2022, 10, 10.3389/fevo.2022.952439 . hal-03740190

**HAL Id: hal-03740190**

**<https://hal.inrae.fr/hal-03740190>**

Submitted on 29 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



## OPEN ACCESS

## EDITED BY

Corrado Marceno,  
Masaryk University, Czechia

## REVIEWED BY

Guillaume Delhaye,  
Royal Botanic Gardens, Kew,  
United Kingdom  
Maria M. Romeiras,  
University of Lisbon, Portugal

## \*CORRESPONDENCE

Guillaume Lannuzel  
guillaume.lannuzel@gmail.com

## SPECIALTY SECTION

This article was submitted to  
Biogeography and Macroecology,  
a section of the journal  
Frontiers in Ecology and Evolution

RECEIVED 25 May 2022

ACCEPTED 12 July 2022

PUBLISHED 29 July 2022

## CITATION

Lannuzel G, Pouget L, Bruy D,  
Hequet V, Meyer S, Munzinger J and  
Gâteblé G (2022) Mining rare Earth  
elements: Identifying the plant species  
most threatened by ore extraction  
in an insular hotspot.  
*Front. Ecol. Evol.* 10:952439.  
doi: 10.3389/fevo.2022.952439

## COPYRIGHT

© 2022 Lannuzel, Pouget, Bruy,  
Hequet, Meyer, Munzinger and  
Gâteblé. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Mining rare Earth elements: Identifying the plant species most threatened by ore extraction in an insular hotspot

Guillaume Lannuzel 1,2\*, Léa Pouget<sup>1,3</sup>, David Bruy 4,5,  
Vanessa Hequet 4,5, Shankar Meyer<sup>2</sup>, Jérôme Munzinger 4  
and Gildas Gâteblé 1,6

<sup>1</sup>Institut Agronomique Néo-Calédonien, Équipe ARBOREAL, Nouméa, New Caledonia, <sup>2</sup>Endemia, New Caledonia Plants Red List Authority, Nouméa, New Caledonia, <sup>3</sup>Université de Montpellier, Place Eugène Bataillon, Montpellier, France, <sup>4</sup>AMAP, Université de Montpellier, IRD, CIRAD, CNRS, INRAE, Montpellier, France, <sup>5</sup>AMAP, IRD, Herbier de Nouvelle-Calédonie, Nouméa, New Caledonia, <sup>6</sup>INRAE, UE 1353, Unité Expérimentale Villa Thuret, Antibes, France

Conservation efforts in global biodiversity hotspots often face a common predicament: an urgent need for conservation action hampered by a significant lack of knowledge about that biodiversity. In recent decades, the computerisation of primary biodiversity data worldwide has provided the scientific community with raw material to increase our understanding of the shared natural heritage. These datasets, however, suffer from a lot of geographical and taxonomic inaccuracies. Automated tools developed to enhance their reliability have shown that detailed expert examination remains the best way to achieve robust and exhaustive datasets. In New Caledonia, one of the most important biodiversity hotspots worldwide, the plant diversity inventory is still underway, and most taxa awaiting formal description are narrow endemics, hence by definition hard to discern in the datasets. In the meantime, anthropogenic pressures, such as nickel-ore mining, are threatening the unique ultramafic ecosystems at an increasing rate. The conservation challenge is therefore a race against time, as the rarest species must be identified and protected before they vanish. In this study, based on all available datasets and resources, we applied a workflow capable of highlighting the lesser known taxa. The main challenges addressed were to aggregate all data available worldwide, and tackle the geographical and taxonomic biases, avoiding the data loss resulting from automated filtering. Every doubtful specimen went through a careful taxonomic analysis by a local and international taxonomist panel. Geolocation of the whole dataset was achieved through dataset cross-checking, local botanists' field knowledge, and historical material examination. Field studies were also conducted to clarify the most unresolved taxa. With the help of this method and

by analysing over 85,000 data, we were able to double the number of known narrow endemic taxa, elucidate 68 putative new species, and update our knowledge of the rarest species' distributions so as to promote conservation measures.

#### KEYWORDS

nickel-mining threat, biodiversity hotspots, narrow endemic taxa, new species, ultramafic substrates, data cleaning, conservation planning, primary biodiversity data

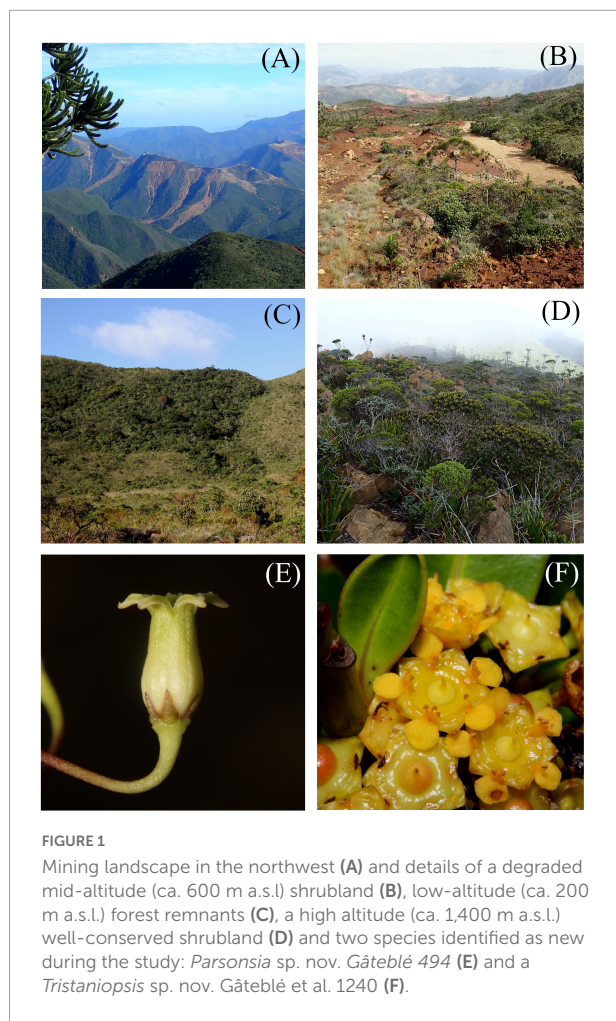
## Introduction

Although biodiversity knowledge is critical to conservation planning, the lack of knowledge about the diversity and distribution of species, known, respectively, as the Linnean and the Wallacean shortfalls, is more than patent in highly diverse regions (e.g., [Bini et al., 2006](#); [Brito, 2010](#)). To overcome these biases, more and more predictive computing methods are being developed, published and applied to larger and larger biodiversity datasets. Some methods can provide insightful results and conclusions at the global scale even with imperfect large datasets, while other automated methods can greatly improve the quality of datasets before analysis ([Bayraktarov et al., 2019](#); [Panter et al., 2020](#); [Heberling et al., 2021](#)). These methods are, however, mostly based on data filtering, thus resulting in more accurate but truncated datasets ([Zizka et al., 2020](#)).

As a result, this increasing amount of online biological data has contributed to a vast variety of studies worldwide in recent decades ([Nelson and Ellis, 2018](#); [Ball-Damerow et al., 2019](#)). However, these studies proved to inherit some recurrent biases from global datasets, mostly taxonomic (misidentification, synonymy), or geographical issues (wrong geolocation or lack of geolocation) ([Meyer et al., 2016](#); [Ball-Damerow et al., 2019](#)). A set of automated methods has been developed to filter the geographic mismatch ([Zizka et al., 2020](#)). Taxonomic bias on the other hand had been widely discussed in the form of a so-called “taxonomic chauvinism,” by which certain taxonomic groups were over-represented in large biological datasets, to the detriment of others ([Troutet et al., 2017](#); [Phaka et al., 2022](#)). Taxonomic inaccuracy on the other hand, is rarely addressed (but see [Anderson et al., 2016](#)) and mostly again by means of a filtering that enhances the accuracy of the overall dataset ([Smith et al., 2016](#)). However, filtering leads to the loss of a great part of the dataset, which is unacceptable if the goal is to shed light on the hidden anomalies, such as new species or new occurrences of known species. Focused taxonomic studies, limited to one taxon (i.e., at genus or family levels) are also signalled to lack of taxonomic checking ([Meyer et al., 2016](#); [Freitas et al., 2020](#)). [Ball-Damerow et al. \(2019\)](#) also surprisingly found that most studies analysed less than a hundred taxa,

a counter-intuitive result in the light of the vast amount of data gathered online. A huge challenge that remains in the use of such large datasets is to produce studies covering a large range of taxonomic groups on a significant area. In this task, the principal hindrance is obviously the amount of time and knowledge needed to clean the data with the least possible filtering, in order to obtain datasets that tend to be as whole and accurate as possible. Expert knowledge on primary datasets here appears to be the best way to achieve the most reliable results at smaller taxonomic and distribution scales ([Maldonado et al., 2015](#)).

In New Caledonia, a southwest Pacific biodiversity hotspot ([Mittermeier et al., 2011](#)), the situation is depressingly similar to other highly diverse regions. On the bright side, according to [Kier et al. \(2009\)](#), it is a territory boasting one of the highest plant endemism rates worldwide. Recent estimates show that almost 1% of the world's vascular plant species are present in New Caledonia (based on [Christenhusz and Byng, 2016](#); [Munzinger et al., 2022](#)), of which 75.5% are endemic, for a territory covering barely 0.01% of the world's terrestrial area. Furthermore, the territory is supposed to be part of the still under debate OCBIL (old, climatically buffered, infertile landscapes) theory, often metal-rich areas, where species are more extinction-prone under anthropogenic disturbance ([Hopper et al., 2016](#); [Pillon et al., 2021](#)). These areas, represented by ultramafic substrates, cover roughly 30% of Grande Terre, the main island of the archipelago, and are distributed among areas known locally as “the great southern massif” covering the southern third with a northward projection on the east coast, and northwest massifs scattered on the non-ultramafic plains of the west coast. In these areas, some vegetation types can show a 97% endemism rate ([Isnard et al., 2016](#)). Ultramafic substrates are characterised by a high metal content, notably nickel ore, and as such are subject to increasing degradation by the nickel-mining industry ([Figure 1](#)), with about 25 active open-cast mines ([Losfeld et al., 2015](#)). This high endemism rate, and high level of threat, have already led to the identification of micro- of nano-hotspots ([Wulff et al., 2013](#); [Gâteblé et al., 2018](#)), but also showed how deep our lack of taxonomic knowledge is. [Gâteblé et al. \(2018\)](#) for instance, stated the high rate of species discovery in New Caledonia (about 1 sp. described per month since 2000 on



average), and predicted that this trend would become even more pronounced in the future.

The last territory-scale study (Wulff et al., 2013), conducted with limited datasets, had already identified about 20% of the endemic species as narrow-endemics, and also elucidated several narrow-endemism hotspots, especially on ultramafic substrate. The ongoing redlisting work of the local Red List Authority (RLA-NC) which had 1,837 species evaluated (Meyer et al., 2021) identified 44% of them as threatened, mainly by bushfires, invasive species, and mining activities. This last threat represents a particular challenge because, unlike the first two, it affects biodiversity but also the possibilities of natural regeneration, because of soil removal. Despite the advances made, Meyer et al. (2021) recognised that the RLA-NC work ran into a kind of Linnean shortfall, due to the knowledge gap in many families and genera. Mine-oriented studies have also been conducted to bypass this obstacle, making it possible to list threatened species on some mined areas (Lowry and Munzinger, unpublished, Lannuzel et al. (2021), unpublished) and plan conservation, but remained limited to

several mining-areas. An updated work, based on all available data and up-to-date taxonomic knowledge, was thus needed to arrest the biodiversity erosion due to the mining industry at a larger scale.

To do so, a methodology was needed to account for every occurrence available for the study area, and improve their value without losing part of the dataset through automated filtering methods. Several issues identified in global primary biodiversity databases thus needed to be addressed. The aggregation of all occurrences available in the study area, even the non- or mis-geolocated ones, the synonymy-harmonisation issue, and the misidentified occurrences were the most obvious ones. And finally, regarding the significant proportion of unidentified occurrences, and their inclination to hide new narrow endemic taxa in New Caledonia (Gâteblé et al., 2018), their identification was critical. To tackle these potential biases, we chose to build a methodology that relies on the aggregation of all botanical occurrences available, the involvement of as many taxonomic experts as possible, as well as local non-professional botanists for their field-knowledge. In doing so, we aimed at (1) producing an updated list of plant taxa threatened by mining activities, (2) updating the narrow-endemism hotspots established by Wulff et al. (2013), (3) showing the added value that resides in the use of all available botanical resources.

## Materials and methods

### Study area

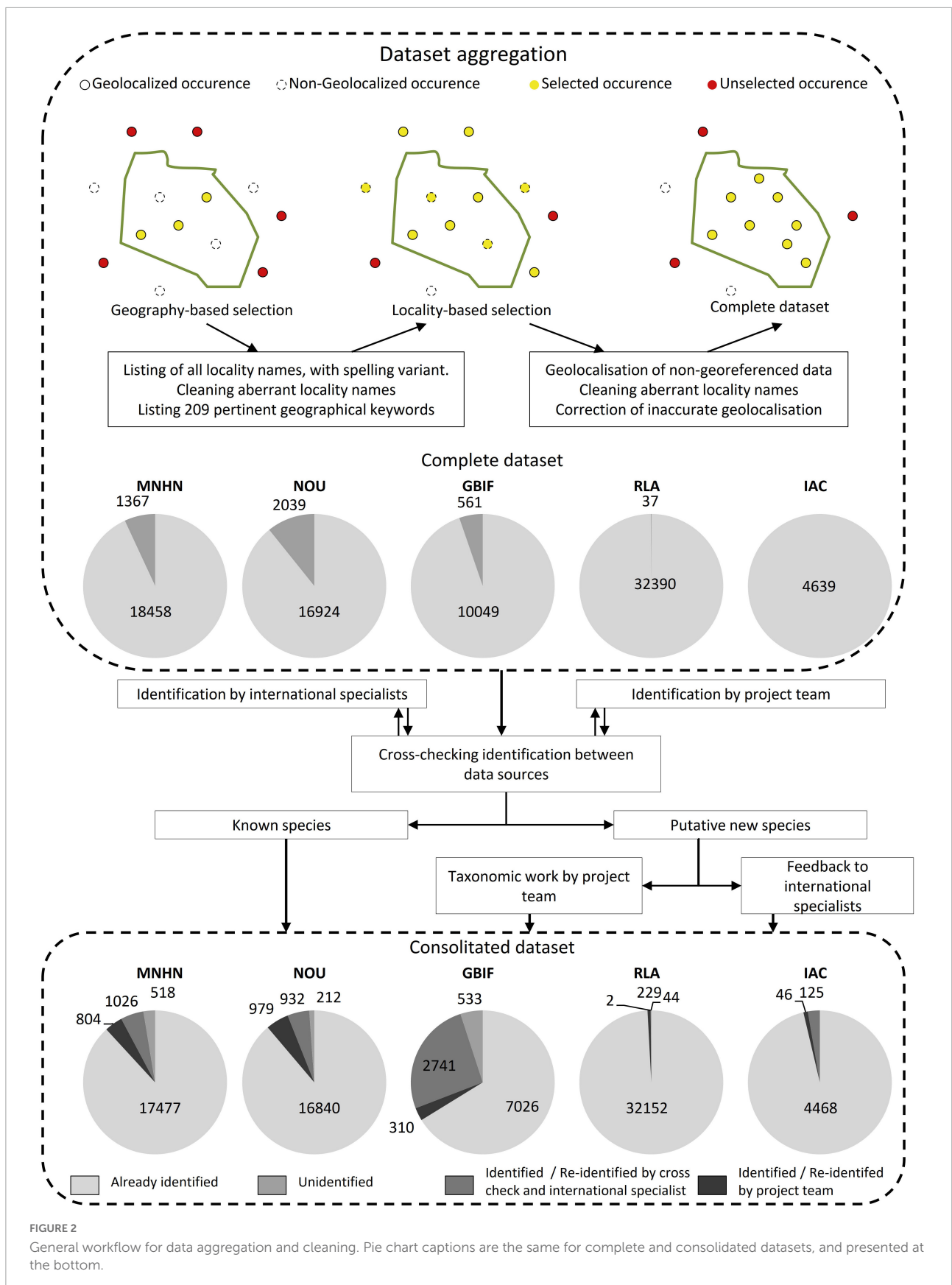
The study area was defined by including (i) every mining concession which includes an active mining area, (ii) every concession adjacent to one included in (i), and (iii) a 1,000 m buffer to account for possible one arc-minute fuzzing in the data.

This results in a 3,100 km<sup>2</sup> study area covering about 55% of all ultramafic substrate in New Caledonia, from the great southern massif, to Poum mountain in the North-West of Grande Terre. The study area was then divided for further analysis into regions representing locally accepted mining entities. Every geographical analysis was computed using QGIS 3.16.

### Data aggregation

Data treatment was computed using a dedicated PostgreSQL 13.2 database, with a pgAdmin 4 user interface and a postGIS 3.1 link with the geography software.

Species presence data were gathered from local and international datasets for all vascular plants. The full dataset from the NOU herbarium (Bruy et al., 2021) was provided, as well as a complete dataset from the P herbarium





(Le Bras et al., 2017), corresponding to New Caledonia. Data was also extracted from GBIF<sup>1</sup> based on the field “Country or Area = “New Caledonia.” Within the GBIF dataset, entries from the herbarium P were removed, as they were obtained directly from this institution. Locally, the RLA-NC database, which comprises both herbaria and field observation data was included, such as the Institut Agronomique néo-Calédonien (IAC) database on rare and threatened species. Some additional field trips were carried out on areas with a lack of data or for specific taxa (see [Supplementary Appendix 2](#)).

Geographical filtering was applied to extract all occurrences present in the study area. The list of all locality names, as filled in the dataset, was created from this first extraction, with all spelling variants kept. This list was then used to query again the initial dataset and extract every occurrence with a corresponding locality name, thus including those with no, or inaccurate, georeference. Finally, a list of significant geographical keywords was created (209 words) from this list and used to re-query the initial dataset. At each step, every occurrence was integrated in our dataset without duplicates. The general workflow for data aggregation is presented in [Figure 2](#). The dataset obtained, called “complete dataset” is summarised in [Table 1](#).

## Data cleaning

Geolocation of the whole dataset was achieved through successive steps. At first, locality names were harmonised through different sql functions, in order to eliminate most spelling errors and allow comparison and grouping of occurrences on a locality-name basis. New Caledonian toponymy is in the middle of the ford (Gay, 2017) since the Kanak toponyms are still being inventoried with an orthography that is not always consistent. The harmonisation process was therefore carefully handled, with the help of multiple local references and discussions with local knowledge holders. Second, cross-checking was performed between different sources to recover geographical coordinates. These cross-checkings were done on the basis of a collector and collection number correspondence, then on the correspondence of locality name and altitude. RLA and NOU datasets were used as references, as they had gone through local procedures for location determination (see Meyer et al., 2021 for the RLA-NC procedure). The NOU herbarium geolocation is based on the coordinates indicated on herbarium labels, or, when absent, is related to the MacKee gazetter and based on the field journal of H.S. MacKee, the most prolific collector in New Caledonia (Morat, 2010). The MacKee gazetter, long available online, is not available anymore since 31/12/2020.

<sup>1</sup> <https://doi.org/10.15468/dl.wnppq>

**TABLE 1** Summary of the original data aggregated per dataset with N, occurrence number; Ngeoloc, number of geolocated data; Nindet, number of occurrence not identified at species or lower level.

	N	Ngeoloc	Nindet
P	19,825	16,341 (82%)	1,367 (7%)
NOU	18,958	18,862 (99%)	2,039 (11%)
GBIF (hors P)	11,174	8,419 (79%)	561 (5%)
Other (RLA + IAC)	32,424 + 4,981	37,066 (100%)	37 (< 0.1%)
Project field mission	371	371 (100%)	66 (18%)

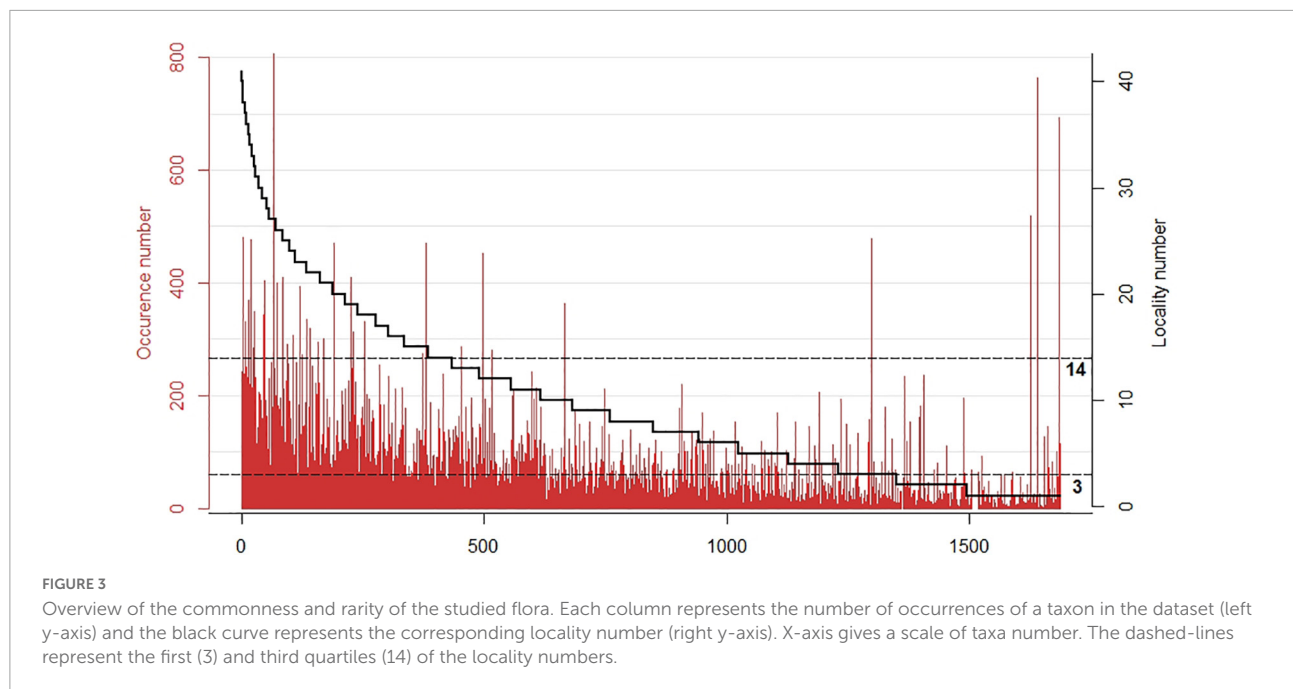
Remaining locality names were located manually using the team’s field knowledge and various historical items, including maps (Balansa, 1873; Laporte, 1903, 1939) and field botanists’ journals, either held in NOU herbarium, or published (Meunier and Tessereau, 2017). Local members of the RLA-NC were also called upon at some point to confirm the most obscure localities.

Finally, the whole dataset was filtered one-by-one on every 209 locality keywords to reveal the obviously erroneous geolocations and correct them. When correcting the locations, occurrences collected since 2007 were treated very carefully, as they are supposed to be based on GPS device location, and consequently more truthful than earlier collections. A sampling effort index was computed considering the number of occurrences from NOU, P, and GBIF datasets per square kilometre to assess the potential sampling bias. Other datasets were not considered for this index because rare and threatened species are often overrepresented in local databases, mainly influenced by conservation-oriented grants or studies, and only in certain regions. Including it in the analyses thus would have biased the results concerning the repartition of sampling effort in the study area.

The first step in taxonomic cleaning was taxon name harmonisation. The reference considered was Florical (Munzinger et al., 2022), the most advanced taxonomy reference for New Caledonia. Identifications were then cross-checked between collection duplicates in different herbaria. Reference identification was considered when the identifier field was filled and the identifier was recognised as relevant for the taxon concerned.

Remaining unidentified herbarium specimens were then distributed among the project members, and taxonomy specialists in the world, according to their respective field of expertise. A systematic taxonomy literature review served as reference at this point. Depending on the specialists’ conclusions, some infra-specific taxa were kept and others were discarded from the dataset.

Finally, non-endemic taxa and hybrids were excluded and some groups remained unresolved and were noted as “taxonomic revision needed,” including *Adenodaphne* (Lauraceae), *Alectryon* (Sapindaceae), *Alyxia* (Apocynaceae), *Arthrochianthus* (Fabaceae), *Balanops*



(Balanopaceae), *Canarium* (Burseraceae), *Casearia* (Salicaceae), *Coronanthera* (Gesneriaceae), *Cryptocarya* (Lauraceae), *Dianella* (Asphodelaceae), *Endiandra* (Lauraceae), *Eriocaulon* (Eriocaulaceae), *Eugenia* (Myrtaceae), *Ficus* (Moraceae), *Garcinia* (Clusiaceae), *Guioa* (Sapindaceae), *Homalium* (Salicaceae), *Korthalsella* (Santalaceae), *Lethedon* (Thymelaeaceae), *Litsea* (Lauraceae), *Meiogyne* (Annonaceae), *Myrsine* (Primulaceae), *Myrtopsis* (Rutaceae), *Peperomia* (Piperaceae), *Smilax* (Smilacaceae), *Tapeinosperma* (Primulaceae), *Vitex* (Lamiaceae), *Xylosma* (Salicaceae), and *Zygogynum* (Winteraceae). Within these genera, however, specialists were able to identify at least the robust taxa, those were kept for further analysis. At higher rank, Cyperaceae and a large part of the Pteridophyta were excluded from the analysis, due to the high number of unresolved taxa within, and the lack of experts committed to these groups.

Putative new species were identified and noted as follows “Genus sp. nov. collector collection number,” with one specimen chosen as a temporary reference. Material for the description of new species and/or genus revision were then transmitted to the recognised specialist if available. **Figure 2** gives a detailed analysis of datasets evolution by main data sources.

## Identifying narrow endemic and threatened taxa

A list of orthographic and taxonomic synonyms was established for each taxon resulting from the cleaning process. These synonyms were obtained by comparing the accepted name with the original one in the source dataset. All original

data sources were then re-queried for every taxon name and its synonyms to gather occurrences outside of our study area. In New Caledonia, narrow endemic species were defined by [Wulff et al. \(2013\)](#) as species present in three localities or less, with a locality being a group of occurrences separated by less than 10 km. Our definition differs from [Wulff et al.’s \(2013\)](#) concept of narrow-endemic species (NES) in that we here kept some infra-specific taxa, thus justifying the use of the term narrow-endemic taxa (NET). As in [Wulff et al. \(2013\)](#), the size of a locality used for the definition of NET was questioned. The final goal being to identify the most threatened taxa, locality number was computed with a 10 and a 5 km definition, and compared to IUCN status of already evaluated taxa. The dependence of NET locality number, obtained with both definitions, on IUCN status (CR, EN, or VU) was tested by a linear regression. The dependence was significant for the 10 km ( $p\text{-value}_{10\text{ km}} < 2,2 \cdot 10^{-16}$ ,  $R^2_{10\text{ km}} = 0,26$ ) and the 5 km locality definition ( $p\text{-value}_{5\text{ km}} < 2,2 \cdot 10^{-16}$ ,  $R^2_{5\text{ km}} = 0,30$ ). Further analyses showed that the 5 km definition reduced the number of taxa considered as NET (i.e., up to three localities), omitting several taxa already evaluated as threatened by the IUCN. On the other hand, the 10 km definition gave a higher sensitivity for the most threatened taxa (CR and EN), with 75.7% of already IUCN-evaluated taxa included in the NET list, than the 5 km definition (62.1%). The 10 km definition of a locality used by [Wulff et al. \(2013\)](#) was therefore kept and allows comparisons with the present results. Consequently, NET are defined as taxa present in one (NET1), two (NET2), or three (NET3) localities. Lastly, the endemism richness ([Kier et al., 2009](#)) was computed for each region of the study area, both based on the entire NET list and restricted to the NET1. This

index is computed by giving to each taxon a value of one, equally distributed across its range, based on the mapping unit. In our case, for instance, a taxon restricted to one region represents a 1 range-equivalent for this region. A taxon present in 2 regions has a value of 0,5 for each region and a value of 0,33 for a taxon present in 3 regions. The summary value is then plotted to a 10,000 km<sup>2</sup> surface area to allow comparisons with Kier et al. (2009) results.

Sampling effort was estimated for each region of the study area, dividing the total number of reliable occurrences (i.e., Herbarium specimens) by the region area. Sampling effort effect on the number of NETs, NET1s, and endemism richness was then tested *via* log-linear regressions to assess whether variation in narrow-endemism across massifs results of a survey gap or involves particular biogeographic processes.

## Results

### Consolidated dataset

The complete dataset, corresponding to the study area, comprises 87.733 occurrences. These occurrences were gathered from three primary data holders (P, NOU and IAC) and two aggregators (GBIF and RLA; see Table 1) thus covering about a hundred different original datasets. About 78.500 occurrences (91%) were obtained through geographic extraction, and 7.500 (9%) through the subsequent locality-based extraction. Among main sources, the rate of geolocation is between 79 and 100% of occurrences geolocated in the original dataset, and 0–11% were unidentified at species level or higher.

At the end of the geolocation step, 100% of the data were geolocated, and 48.800 (57%) had their original coordinates modified by more than a kilometre.

In the consolidated dataset obtained, about 7.300 new identifications were made, including above 2.350 made by the project team members, and 1.150 occurrences (1.3% of the dataset) remained unidentified at species level.

The final dataset comprises 1.686 endemic taxa, of which 1.099 had already been evaluated by RLA-NC following IUCN methodology (IUCN, 2012). Twenty-six taxa are about to be added as their revision is nearing completion. Occurrences gathered thanks to the locality-based extraction process added 102 taxa (6%) that would not have been present otherwise, and the taxonomic work also added 321 taxa (19%) to the dataset. Among the latter, 66 taxa are identified as putative new species, and another 41 taxa are not newly discovered but still unpublished taxa (ined. in Munzinger et al., 2022). Resulting locality number analysis is illustrated on Figure 3 and shows that a quarter of the identified taxa corresponds to the NET definition (three or less localities), while half are present in less than 8 localities. The corresponding occurrence numbers also show that even though rare species are globally less observed

than common ones, the observation pressure is not proportional to the locality number. The several abnormal high peaks in occurrences correspond to taxa studied by the IAC (unpublished data) during earlier conservation studies.

### Narrow-endemic taxa listing

The resulting NET-list comprises 457 taxa (Table 2)—representing 384 species and 73 infra-specific taxa—including 63% already evaluated through IUCN methodology by the RLA-NC. Detailed NET-list is given in Supplementary Appendix 1. Seventeen taxa identified as NET 1 or 2 are only present within active extractive mining areas, four of them having already been evaluated as CR by the RLA-NC. The remaining 169 NET will be evaluated in future RLA-NC workshops. It is notable that almost half of the identified NET are NET 1. It is worth noting that almost every putative new species identified during the study corresponds to the definition of NET, while 60% of the unpublished taxa don't.

The distribution of NET numbers, endemism ratio and herbarium specimen density within regions are presented in Figure 4. The endemism ratio ranges from 329 (Camp des Sapins) to 4.999 (Kaala) range equivalent per 10,000 km<sup>2</sup>, with a mean of 1.711 for the total NES values and 942 for NET1. Regions located in the great southern massif have respective mean values of 869 and 491, while regions located in the northwestern massifs have respective mean values of 2.659 and 1.449. Pinpin and Cap Bocage regions, both extremities of these entities, show medium values. The herbarium specimen density shows a clear over-surveying of the northwestern massifs with some exception in the great southern massif represented by the South and Nakety/Dothio regions. The Poro/Kiel and Camp des Sapins regions are under-surveyed but are also among the biggest and most inaccessible regions. Log-linear regression showed a significant positive effect of sampling effort on the number of NET (*p*-value = 0,002), on the number of NET1 (*p*-value = 0,003) and on the endemism richness (*p*-value < 0,001) (see Supplementary Appendix 3).

### Narrow-endemic taxa evolution from 2013 to now

Wulff et al. (2013) found 211 NES within the present study area. The results obtained here doubled this number. Among both NET lists, proportions of 1, 2, or 3 localities species are similar (nearly 40% of NET 1, 33% of NET 2, and 27% of NET 3).

Almost all of the old NES are included in the updated list (Figure 5A). 151 are currently considered as NET, of which 72 have an increased number of localities. 41



**TABLE 2** Summary of narrow endemic taxa (NET) listed with their IUCN status established by the RLA-NC.

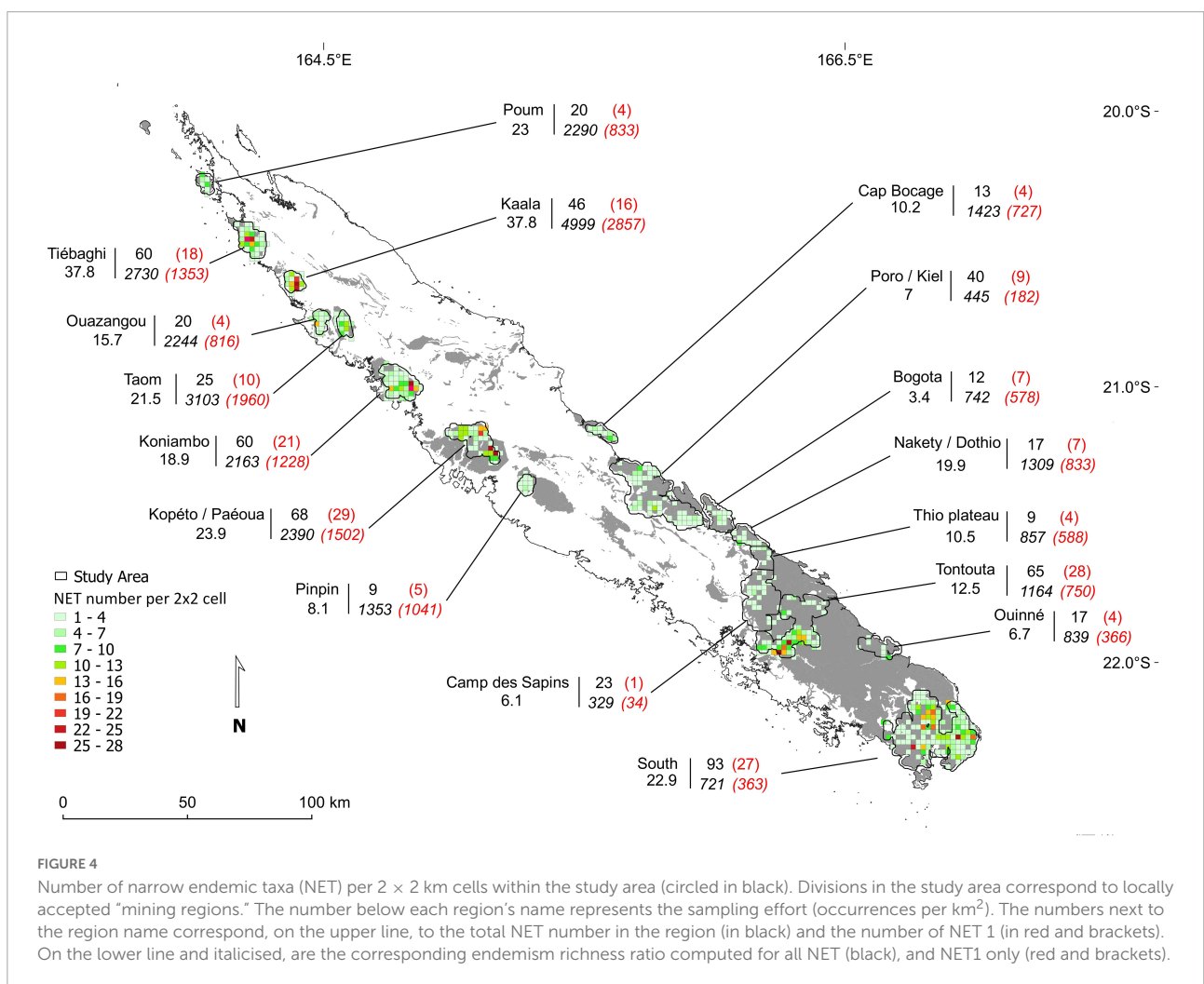
Status	NET1	NET2	NET3	Total
Endangered (CR + EN + VU)	71	67	65	203 (44%)
Non-threatened (LC + NT)	6	6	20	32 (7%)
Evaluation not made (DD + NE)	25	18	12	55 (12%)
Evaluation needed	89	54	24	167 (37%)
<i>Putative sp. nov.</i>	48	10	4	62 (14%)
Unpublished taxa	13	7	5	25 (5%)

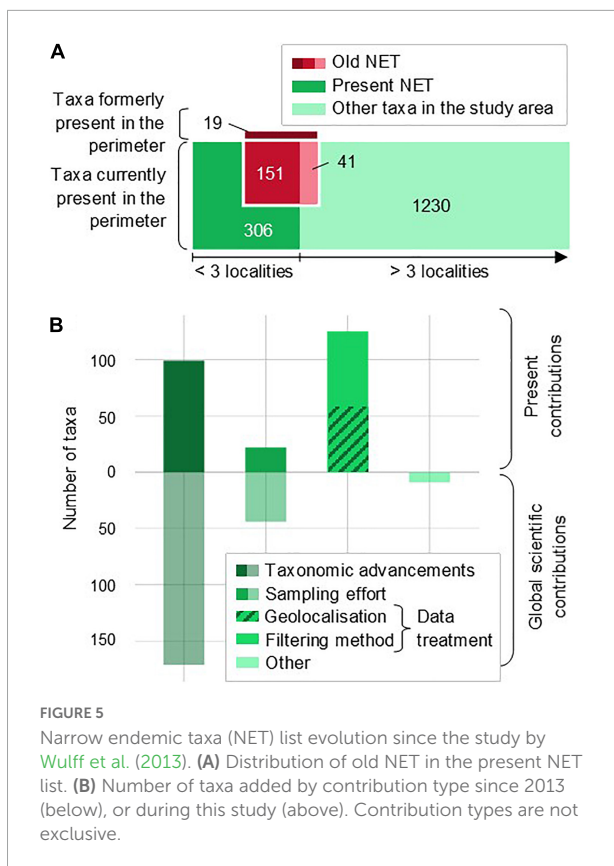
NET1–NET3, narrow endemic taxa with 1–3 localities; CR, Critically Endangered; EN, Endangered; VU, Vulnerable; DD, Data Deficient; NE, Not Evaluated. NE species went through the RLA-NC plants process but were considered not robust enough to be evaluated.

taxa included in Wulff et al.’s (2013) list now have a number of localities above the threshold of 3, excluding them from the present list. Nineteen taxa also disappeared from the total list because of taxonomic revisions (16

taxa), re-identification (1 taxon) or methodological bias (2 taxa).

The 306 taxa increase in the updated NET list is explained both by global scientific contribution and the specific methodology used in this study (Figure 5B). For the global contributions, 29 of the new NET have been described since 2013 (Gâteblé et al., 2018). Further, the RLA-NC work, along with many identifications by specialists explain the remaining “taxonomic advances”-related appearances. New herbarium collections—as “sampling effort”—explain 44 appearances. The “other” category is supposed to be linked with the increasing digitisation effort, but this is hard to prove (Figure 5B). In the same way, present taxonomic contributions are related to putative new species and collections identification realised during the study. Field trips organised during the study also added another 22 taxa in the NET list. Finally, the present method allowed the recovery of 125 taxa via geolocalisation efforts, or the consideration of infra-specific ranks. Some taxa have been considered as NET with a 10 km distance between two localities whereas they correspond to ten or more localities with





a 5 km buffer. This phenomenon affects 10 taxa: *Achilleanthus hypolasius*, *Acianthus veillonis*, *Agathis ovata*, *Austrobuxus rubiginosus*, *Gea connatistipula* ined., *Oncotheca balansae*, *Pittosporum gatopense*, *Pleioluma sebertii*, *Tristaniopsis glauca*, and *Xanthostemon myrtifolius*.

## Discussion

### Tackling the big data issue

The results obtained here show the relevance of using entire datasets in order to obtain a robust listing of the species that need urgent conservation action. Indeed, the method applied to tackle the geographical and taxonomic shortfalls led to the recovery of roughly 20% of the taxa set. However, this work represents a vast amount of work and time and was only made possible by the dedication of one person over a 3-year period, along with the involvement of RLA-NC and the large network of taxonomists who participated in this effort.

Our results confirm the identified biases in biodiversity datasets around the world. Beyond including non-geolocalised data, the cleaning method led to change the localisation of more than half of the occurrences by more than a kilometre and resulted in the addition of 57 species (12.5% of actual

NES) within the project area, a result concomitant with findings elsewhere (Zizka et al., 2020). The one kilometre or coarser resolution is often used for species modelling at a regional scale (Mod et al., 2016; Pecchi et al., 2019), one of the main uses of primary biodiversity data (Ball-Damerow et al., 2019; Heberling et al., 2021), because it fits the bioclimatic data available at a 30 arc-s resolution (Fick and Hijmans, 2017; Karger et al., 2017). Our results, obtained on the basis of locally recognised references, historical material and the involvement of local field-experts, thus reveals a hindrance for using this kind of data at a fairly high resolution. Without any kind of cleaning, such work would suffer a substantial geographical bias, adding uncertainty to the identified biases in the bioclimatic datasets (Dubos et al., 2022).

Secondly, the taxonomy issue, well described in global datasets (James et al., 2018), is particularly expressed here. As a matter of fact, identifications of herbaria from the P and NOU herbaria, which provided us with up-to-date data, remained almost unchanged, except for the unidentified specimens, and some synonyms issues. On the other hand, about a quarter of GBIF-obtained occurrences had their identification corrected during the study, while the amount of unidentified specimens remained stable. This latter statement is mainly due to the absence of online available specimen pictures. The former on the other hand points to the difference between herbaria, curated and regularly updated, and online data repositories where data curation is a complex issue (Triebel et al., 2012; Zizka et al., 2020). Here we recognise the high value of such a worldwide repository, but focus on, and advocate for, the need for feedback and/or curation of the datasets (Miller et al., 2015) if they are to be used at global scale, i.e., without local taxonomists or people aware of the local toponymy.

### The increasing knowledge availability

On the other hand, we confirmed the benefits obtained from the digitisation effort worldwide. At a global scale, Heberling et al. (2021) showed a sixfold increase of online herbarium vouchers between 2007 and 2021. No such analysis is available locally but, as an indication, the totality of the NOU herbarium sheets (ca. 90 000 specimens) have been digitised and put online during the last decade (Bruy et al., 2021), including several thousands by the project team. Furthermore, based on the numbers given by Wulff et al. (2013), we were able to calculate that their dataset was composed of around 150.000 occurrences from P, NOU, and Z herbarium for all of New Caledonia. As a comparison, our raw dataset for these three institutions comprises about 250.000 occurrences for the whole of New Caledonia, representing a 70% increase within 10 years. This increase must at least partly explain the 180% rise in the number of narrow-endemics sensu Wulff et al. (2013)—i.e., considering only the specific rank) found here. A bigger dataset might also represent a limit, due to our definition of a NET.

Our methodology is based on a 10 km distance between two occurrences to consider distinct localities. In some cases, two long known localities, separated by more than 10 km, may now be considered as one, if new observations were added between both original locations. Thus, increasing the quantity of occurrences may have the mechanical effect of reducing the number of localities, leading to the assessment of a common species as a NET in some cases. Taxa affected by this phenomenon raise some doubts about their consideration as NET. All taxa concerned are restricted to the great southern massif, except *Pittosporum gatopense*, a species considered as critically endangered (Gemmill et al., 2017) and present at the bottom of some northwestern massifs. These “anomalies” may thus reflect a distribution pattern of taxa restricted to the great southern massif, yet relatively abundant locally. In such a case, the distribution of a taxon restricted to one massif such as the great southern massif, could result in one or a few large localities (10 km definition) but more narrower localities (5 km definition). This distinction has long been stated in the studies of rarity (Rabinowitz, 1981), where rare species can have a narrow distribution and be both rare or abundant locally. The concept of NET as defined by Wulff et al. (2013) tends to identify constantly sparse and geographically restricted species. We, however, show here that it can also capture locally abundant but geographically restricted species, following Rabinowitz (1981) forms of rarity. The IUCN status of these species, for the ones already evaluated, ranges from CR to LC, depending mostly on the threats to their habitat, which affects the location count. This limit in the methodology illustrates an apparent contradiction and reveals that in some cases, geographically restricted taxa may not be critically threatened. The NET approach is thus powerful to identify taxa that are inherently vulnerable due to their narrow-range distribution, and is very informative for biogeographical understanding of the flora. It, however, needs to be complemented by the IUCN (2012) approach in order to at least complete the prioritisation process for conservation actions.

The great increase in NET numbers also stems from the continuous scientific effort toward completion of the biodiversity inventory, both in terms of taxonomic revisions and collection of specimens. Our results show the last decade's survey efforts by botanists added several dozen species to the project list and we showed a positive relationship between the number of NETs, NET1s, endemism richness and the survey effort per massif. This shows that the sampling effort is not yet sufficient in New Caledonia to fully understand the distribution of and number of species, advocating for a continuous field work. More importantly, as stated by Gâteblé et al. (2018), New Caledonia still remains on course to achieve the complete inventory of its flora. They also prophesied that further taxonomic description would concern primarily narrow range species. Our results corroborate this statement as almost all putative new species identified during our study

are considered NET in our results. Some of them are presently being described but a lot of the genera and families left aside in this study remain orphans of a specialist engaged to push this endeavour forward. We hope this study will foster trained taxonomists to fill this gap in the near future.

## The extreme nature of narrow-endemism in New Caledonia...

Previous studies on narrow-endemic species in New Caledonian flora (Wulff et al., 2013) or fauna (Caesar et al., 2017) highlighted, respectively, 21.7% of plants and 22% of animals as narrow-endemics. Our results, even slightly higher (27.1%), remain comparable and emphasise the extreme nature of narrow-endemism in New Caledonia, and the irreplaceability of the flora. Kier et al. (2009) defined the endemism richness as the range equivalent sum (proportion of the distribution of an endemic species per area unit) for 10,000 km<sup>2</sup>. They already found that New Caledonia was surpassing every other place with an endemism richness of about 1,350, the second being the Cape region in South Africa with about 750. This value can be compared with the 3,700 found by Gâteblé et al. (2018)—computed on the basis of NET1—on Ile Art, north of Grande Terre, another narrow-endemism hotspot but not yet threatened by active mining activity. We computed this ratio only on the basis of the NET and NET1 lists, making any comparison tricky. However, even with this restricted definition, we found values higher than Kier et al.'s (2009) results for the whole of New Caledonia, especially in the regions located in the northwestern massifs. Despite a similar sampling effort, the results found in the southern massif are also surprising because they yield lower values than in the northwest. They may be the result of a larger area in a single block, interpreted as a hindrance to speciation and consequently to the diversification of NETs in this region. These results, along with Gâteblé et al.'s (2018), may be seen as a validation of Isnard et al. (2016) definition of the ultramafic patches as “edaphic islands.” Edaphic islands are known as drivers of plant speciation and endemism (Rajakaruna, 2018), and it has been suggested that their endemism rate was a function of isolation and matrix permeability (McGann, 2002; Itescu, 2019). The matrix (non-ultramafic areas here) permeability issue would require more investigation to be tested soundly. However, the apparent geographical isolation of northwestern massifs in comparison with the southern massif seems to be reflected here in the superior endemism ratio found in the former. Further biogeographical work is needed to conclude on that matter, but in the meantime, the precautionary principle suggests to consider the northwestern massifs as edaphic isolated islands and plan conservation actions accordingly. Considering this, climate change is also to be considered from this point of view, as it poses

peculiar challenges in the case of edaphic endemics (Corlett and Tomlinson, 2020), and notably those having a narrow altitudinal range.

## ... and the need for adapted conservation planning

From a conservation point of view, the high narrow-endemism levels described here are a sad reminder of the urgent need for conservation actions in ultramafic areas of New Caledonia, as well as of the lack of protection (Jaffré et al., 1998; Wulff et al., 2013; Gâteblé et al., 2018; Ibanez et al., 2019). The original narrow-endemism hotspots highlighted by Wulff et al. (2013) are here highlighted again, except that far more NET are now identified in these areas. Among these NET, 86% of taxa assessed on the IUCN Red List (excluding DD and NE) are considered to be threatened. There is no doubt that the results will be roughly equivalent for the whole NET list. However, any conservation measure on ultramafic areas must account for the peculiar nature of these environments. Recently, Pillon et al. (2021), showed that ultramafic areas in New Caledonia could be considered as OCBILs (Hopper et al., 2016), a statement supported by the high endemism richness found in this study. Furthermore, the high level of irreplaceability, a point here enlightened by the 18 NET only existing within mining areas, was also recognised as a key feature of OCBILs conservation (Hopper et al., 2021). Thus, thinking of ultramafic areas in the OCBIL intellectual framework brings insight on the NET conservation on mining areas, as it cannot be planned as in other areas (Hopper et al., 2021). First, the infertile nature of these areas, largely stated by L'Huillier et al. (2010), as well as the reduced dispersability of the plant species present (Ititiaty et al., 2020; Pillon et al., 2021), is a common feature of OCBILs (Hopper et al., 2016). It results locally in a hindrance for ecological restoration (Losfeld et al., 2015)—a standard way to improve and recover both quality and quantity of species' habitats in industrial environments. We here admit that it represents a similar hindrance for species translocation—a common but still hazardous tool for species conservation (Godefroid et al., 2016)—as it may impact both establishment and growth rates. As a mirror, Hopper et al. (2021) suggested the inefficiency of both restoration and biodiversity offsets (May et al., 2017), a commonly proposed mining industry attenuation tool (ICMM, 2005) in such areas for conservation purposes. Similarly, the discussed subject of avoiding biodiversity loss by reducing habitat fragmentation (Fahrig, 2019) would make little sense here, at least at a territory scale, regarding the high NET turnover between isolated massifs. It could, however, be effective at a higher resolution, to plan reserve and conservation actions on ultramafic massifs (Justeau-Allaire et al., 2021). Consequently, if we accept the OCBIL nature of the ultramafic areas, the island-like nature of northwestern massifs, and the resulting high narrow-range

diversity stated here, conservation actions need to be scaled accordingly. We thus advocate for several urgent measures. First and foremost, the implementation of diversely scaled reserves (Ibanez et al., 2019) is urgently needed on every massif to protect a significant proportion of these taxa. Second, the long lasting evolution of species assemblages in these environments (Hopper et al., 2021) cannot be substituted by human timescale devices. The use of biodiversity offsets then shouldn't be considered as a conservation tool *per se*. And last but not least, we urge to change our prism and consider each and every ultramafic patch as an island, and build conservation and restoration plans to that scale. The present results must be of use for such an endeavour, as already proved on one massif (Lannuzel et al., 2021). But other tools developed locally (Justeau-Allaire et al., 2021) or in similar environments (e.g., Tomlinson et al., 2020) must be mobilised to ensure that biodiversity erosion is stopped in these unique environments.

## Conclusion

In this study, we first aimed at using the biggest dataset available to enhance our knowledge of taxa threatened by mining activities. Through an original workflow, this goal was reached and showed more than ever the high endemism richness found in the flora of New Caledonian ultramafic areas. As a consequence, we were able to tackle, at least partly, the Linnean and Wallacean shortfalls identified by Meyer et al. (2021) and our results will feed the future works of the RLA-NC plants. However, as often in science, the answers brought here led to more questions for the conservation of this unique flora. On one side, the recognition of far more NET, along with the recent advances regarding the ecology of these environments draws a path for actions. On the other side, we also emphasised the lack of knowledge about biodiversity and the work still needed to complete the biodiversity inventory. The scientific progress being slower than the mining exploitation, the question remains, how to preserve what is not yet known?

## Data availability statement

The original contributions presented in this study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

GL and LP mainly wrote the present manuscript. All authors participated in the comments, and review process before submission and included at every step of the study.



## Funding

This research was fully funded by a grant (CSF n°4PS2017-CNRT.IAC/ERMINE) from “CNRT Nickel et son Environnement, Nouméa, New Caledonia” to study rare species threatened by nickel mining.

## Acknowledgments

We are grateful to NOU and P herbaria for giving access to the whole collection concerning New Caledonia. All members of the RLA-NC plants are also thanked for their support and advice, both orally and in the field, as well as for the shared knowledge. This work shouldn't have been done without the involvement of every taxonomy specialist involved. We are especially grateful to R. Amice, L. Barrabé, M. Callmander, S. Knapp, C. Laudereau, P.P. Lowry, A. Mouly, L. Perrie, Y. Pillon, N. Snow, H. Vandrot, and J. Wang, who made many identifications within their group of interest. We apologise for any omission on that point. The NMC, SMGM, SMT, and SLN mining companies are thanked as well for giving access to their sites during the study. Roy Benyon is acknowledged for his help and wise advice on English writing. We are also grateful to

the two reviewers who helped improve the quality of this work.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fevo.2022.952439/full#supplementary-material>

## References

- Anderson, R., Araujo, M., Lobo, J. M., Martinez-Meyer, E., Peterson, A. T., and Soberon, J. (2016). *Final Report of the Task Group on GBIF Data Fitness for Use in Distribution Modelling*. Copenhagen: GBIF.
- Balansa, B. (1873). Carte de la Nouvelle-Calédonie Indiquant les Principaux Itinéraires Suivis par Mr B. Balansa de 1868 à 1872. Available Online at: <http://catalogue.bnf.fr/ark:/12148/cb406344157> (accessed October 22, 2019).
- Ball-Damerow, J. E., Brenskelle, L., Barve, N., Soltis, P. S., Sierwald, P., Bieler, R., et al. (2019). Research applications of primary biodiversity databases in the digital age. *PLoS One* 14:e0215794. doi: 10.1371/journal.pone.0215794
- Bayraktarov, E., Ehmke, G., O'Connor, J., Burns, E. L., Nguyen, H. A., McRae, L., et al. (2019). Do Big Unstructured Biodiversity Data Mean More Knowledge? *Front. Ecol. Evol.* 6:239. doi: 10.3389/fevo.2018.00239
- Bini, L. M., Diniz-Filho, J. A. F., Rangel, T. F. L. V. B., Bastos, R. P., and Pinto, M. P. (2006). Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. *Divers. Distrib.* 12, 475–482. doi: 10.1111/j.1366-9516.2006.00286.x
- Brito, D. (2010). Overcoming the Linnean shortfall: data deficiency and biological survey priorities. *Basic Appl. Ecol.* 11, 709–713. doi: 10.1016/j.baec.2010.09.007
- Bruy, D., Barrabé, L., Birnbaum, P., Dagostini, G., Donnat, M., Fambart-Tinel, J., et al. (2021). *L'Herbier de Nouvelle-Calédonie [Internet]*. Nouméa: UMR AMAP. IRD, CIRAD, CNRS, INRAE, Univ. Montpellier. AMAP: UMR. doi: 10.23708/HERBIER-NOUVELLE-CALEDONIE
- Caesar, M., Grandcolas, P., and Pellens, R. (2017). Outstanding micro-endemism in New Caledonia: more than one out of ten animal species have a very restricted distribution range. *PLoS One* 12:e0181437. doi: 10.1371/journal.pone.0181437
- Christenhusz, M. J. M., and Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa* 261, 201–217. doi: 10.11646/phytotaxa261.3.1
- Corlett, R. T., and Tomlinson, K. W. (2020). Climate change and edaphic specialists: irresistible force meets immovable object? *Trends Ecol. Evol.* 35, 367–376. doi: 10.1016/j.tree.2019.12.007
- Dubos, N., Augros, S., Deso, G., Probst, J.-M., Notter, J.-C., and Roesch, M. A. (2022). Here be dragons: important spatial uncertainty driven by climate data in forecasted distribution of an endangered insular reptile. *Anim. Conserv.* doi: 10.1111/acv.12775
- Fahrig, L. (2019). Habitat fragmentation: a long and tangled tale. *Glob. Ecol. Biogeogr.* 28, 33–41. doi: 10.1111/geb.12839
- Fick, S. E., and Hijmans, R. J. (2017). Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi: 10.1002/joc.5086
- Freitas, T. M. S., Montag, L. F. A., de Marco, P. Jr., and Hortal, J. (2020). How reliable are species identifications in biodiversity big data? Evaluating the records of a neotropical fish family in online repositories. *Syst. Biodivers.* 18, 181–191. doi: 10.1080/14772000.2020.1730473
- Gâteblé, G., Barrabé, L., McPherson, G., Munzinger, J., Snow, N., and Swenson, U. (2018). One new endemic plant species on average per month in New Caledonia, including eight more new species from Ile Art (Belep Islands), a major-hotspot in need of protection. *Aust. Syst. Bot.* 31, 448–480. doi: 10.1071/SB18016
- Gay, J.-C. (2017). “Un atlas pour la Nouvelle-Calédonie,” in *Entre Terres et Mers, Cartographies du Sud-Ouest de l'Océan Indien. De la Carte Ancienne aux Projets D'Atlas Contemporains*, eds C. H. Germanaz, S. et Bouchet (Saint-Denis: Epica Éditions).



- Gemmill, C., Veillon, J.-M., Amice, R., Cazé, H., Dumontet, V., Fleurot, D., et al. (2017). *Pittosporum gatopense*. *The IUCN Red List of Threatened Species* 2017: e.T35266A67764057. <https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T35266A67764057.en>. (accessed Apr 12, 2022).
- Godefroid, S., Le Pajolec, S., and Van Rossum, F. (2016). Pre-translocation considerations in rare plant reintroductions: implications for designing protocols. *Plant Ecol.* 217, 169–182. doi: 10.1007/s11258-015-0526-0
- Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., and Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2018093118. doi: 10.1073/pnas.2018093118
- Hopper, S. D., Lambers, H., Silveira, F. A. O., and Fiedler, P. L. (2021). OCBIL theory examined: reassessing evolution, ecology and conservation in the world's ancient, climatically buffered and infertile landscapes. *Biol. J. Linn. Soc.* 133, 266–296. doi: 10.1093/biolinnean/blaa213
- Hopper, S. D., Silveira, F. A. O., and Fiedler, P. L. (2016). Biodiversity hotspots and Ocbil theory. *Plant Soil* 403, 167–216. doi: 10.1007/s11104-015-2764-2
- Ibanez, T., Birnbaum, P., Gateblé, G., Hequet, V., Isnard, S., Munzinger, J., et al. (2019). Twenty years after Jaffré et al. (1998), is the system of protected areas now adequate in New Caledonia? *Biodivers. Conserv.* 28, 245–254. doi: 10.1007/s10531-018-1659-y
- ICMM. (2005). *Biodiversity Offsets - A Briefing Paper for the Mining Industry*. London: ICMM.
- Isnard, S., L'Huillier, L., Rigault, F., and Jaffré, T. (2016). How did the ultramafic soils shape the flora of the New Caledonian hotspot? *Plant Soil* 403, 53–76. doi: 10.1007/s11104-016-2910-5
- Itescu, Y. (2019). Are islandlike systems biologically similar to islands? A review of the evidence. *Ecography* 42, 1298–1314. doi: 10.1111/ecog.03951
- Ititiaty, Y., Brescia, F., Bordez, L., Gensous, S., McCoy, S., and Fogliani, B. (2020). Implications of seed traits for restoration of ultramafic plant taxa from the Goro plateau in the subtropical hotspot of New Caledonia. *Restor. Ecol.* 28, 1505–1516. doi: 10.1111/rec.13248
- IUCN (2012). *Catégories et Critères de la Liste Rouge de l'UICN*. Gland: UICN.
- Jaffré, T., Bouchet, P., and Veillon, J.-M. (1998). Threatened plants of New Caledonia: is the system of protected areas adequate? *Biodivers. Conserv.* 7, 109–135.
- James, S. A., Soltis, P. S., Belbin, L., Chapman, A. D., Nelson, G., Paul, D. L., et al. (2018). Herbarium data: global biodiversity and societal botanical needs for novel research. *Appl. Plant Sci.* 6:e1024. doi: 10.1002/aps.3.1024
- Justeau-Allaire, D., Vieilledent, G., Rinck, N., Vismara, P., Lorca, X., and Birnbaum, P. (2021). Constrained optimization of landscape indices in conservation planning to support ecological restoration in New Caledonia. *J. Appl. Ecol.* 58, 744–754. doi: 10.1111/1365-2664.13803
- Karger, D. N., Cornad, O., Böhner, J., Kawohl, T., Krefl, H., Soria-Auza, R. W., et al. (2017). Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4:170122. doi: 10.1038/sdata.2017.122
- Kier, G., Krefl, H., Lee, T. M., Jetz, W., Ibsch, P. L., Nowicki, C., et al. (2009). A global assessment of endemism and species richness across island and mainland regions. *PNAS* 106, 9322–9327. doi: 10.1073/pnas.0810306106
- L'Huillier, L., Jaffré, T., and Wulff, A. (2010). *Mines et Environnement en Nouvelle-Calédonie: les Milieux sur Substrats Ultramafiques et Leur Restauration*. Païta: IAC.
- Lannuzel, G., Balmot, J., Dubos, N., Thibault, M., and Fogliani, B. (2021). High-resolution topographic variables accurately predict the distribution of rare plant species for conservation area selection in a narrow-endemism hotspot in New Caledonia. *Biodivers. Conserv.* 30, 963–990. doi: 10.1007/s10531-021-02126-6
- Laporte, C. (1903). *Carte de la Nouvelle Calédonie Dressée Pour l'Union Agricole Calédonienne*. Paris: A. Challamel.
- Laporte, C. (1939). *Carte de la Nouvelle-Calédonie*. Paris: Société d'éditions géographiques, maritimes et coloniales.
- Le Bras, G., Pignal, M., Jeanson, M. L., Muller, S., Aupic, C., Carré, C., et al. (2017). The French Muséum national d'histoire naturelle plant herbarium collection dataset. *Sci. Data* 4:170016. doi: 10.1038/sdata.2017.16
- Losfeld, G., L'Huillier, L., Fogliani, B., Jaffré, T., and Grison, C. (2015). Mining in New Caledonia: environmental stakes and restoration opportunities. *Environ. Sci. Pollut. Res.* 22, 5592–5607. doi: 10.1007/s11356-014-3358-x
- Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Alban, J., et al. (2015). Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases. *Glob. Ecol. Biogeogr.* 24, 973–984. doi: 10.1111/geb.12326
- May, J., Hobbs, R. J., and Valentine, L. E. (2017). Are offsets effective? An evaluation of recent environmental offsets in Western Australia. *Biol. Conserv.* 206, 249–257. doi: 10.1016/j.biocon.2016.11.038
- McGann, T. D. (2002). How Insular are Ecological “Islands”? An Example from the Granite Outcrops of the New England Batholith of Australia. *Proc. R. Soc. Qld.* 110, 1–13.
- Meunier, J.-Y., and Tessereau, F. (2017). Le professeur André Guillaumin et la mission franco-suisse de botanique en Nouvelle-Calédonie (1950–51). *Bull. de la Soc. d'Études Historiques de Nouvelle-Calédonie* 193, 56–77.
- Meyer, C., Weigelt, P., and Krefl, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* 19, 992–1006. doi: 10.1111/ele.12624
- Meyer, S., Birnbaum, P., Bruy, D., Cazé, H., Garnier, D., Gateblé, G., et al. (2021). “The New Caledonia plants RL: bringing botanists together for the conservation of the archipelago's crown jewel,” in *Reference Module in Earth Systems and Environmental Sciences* (Amsterdam: Elsevier). doi: 10.1016/B978-0-12-821139-7.00171-9
- Miller, J. A., Georgiev, T., Stoev, P., Sautter, G., and Penev, L. (2015). Corrected data re-harvested: curating literature in the era of networked biodiversity informatics. *Biodivers. Data J.* 3:e4552. doi: 10.3897/BDJ.3.e4552
- Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., and Gascon, C. (2011). “Global Biodiversity Conservation: The Critical Role of Hotspots,” in *Biodiversity Hotspots*, eds F. Zachos, J. Habel (Berlin: Springer), 3–22. doi: 10.1007/978-3-642-20992-5\_1
- Mod, H. K., Scherrer, D., Luoto, M., and Guisan, A. (2016). What we use is not what we know: environmental predictors in plant distribution models. *J. Veg. Sci.* 27, 1308–1322. doi: 10.1111/jvs.12444
- Morat, P. (2010). Les botanistes récolteurs en Nouvelle-Calédonie de 1774 à 2005. *Adansonia* 32, 159–216. doi: 10.5252/a2010n2a1
- Munzinger, J., Morat, P., Jaffré, T., Gateblé, G., Pillon, Y., Rouhan, G., et al. (2022). *FLORICAL: Checklist of the vascular indigenous flora of New Caledonia*. Available Online at: <http://publish.plantnet-project.org/florical> (accessed Jan 27, 2021).
- Nelson, G., and Ellis, S. (2018). The history and impact of digitization and digital data mobilization on biodiversity research. *Philos. Trans. R. Soc. Biol. Sci.* 374:20170391. doi: 10.1098/rstb.2017.0391
- Panter, C. T., Clegg, R. L., Moat, J., Bachman, S. P., Klitgard, B. B., and White, R. L. (2020). To clean or not to clean: cleaning open-source data improves extinction risk assessments for threatened plant species. *Conserv. Sci. Pract.* 2:e311. doi: 10.1111/csp2.311
- Pecchi, M., Marchi, M., Burton, V., Gianetti, F., Moriondo, M., Bernetti, I., et al. (2019). Species distribution modelling to support forest management. A literature review. *Ecol. Model.* 411:108817. doi: 10.1016/j.ecolmodel.2019.108817
- Phaka, F. M., Vanhove, M. P. M., du Preez, L. H., and Hugé, J. (2022). Reviewing taxonomic bias in a megadiverse country: primary biodiversity data, cultural salience, and scientific interest of South African animals. *Environ. Rev.* 30, 39–49. doi: 10.1139/er-2020-0092
- Pillon, Y., Jaffré, T., Birnbaum, P., Bruy, D., Cluzel, D., Ducouso, M., et al. (2021). Infertile landscapes on an old oceanic island: the biodiversity hotspot of New Caledonia. *Biol. J. Linn. Soc.* 133, 317–341. doi: 10.1093/biolinnean/blaa146
- Rabinowitz, D. (1981). “Seven forms of rarity,” in *The Biological Aspects of Rare Plant Conservation*, ed J. Synge (Hoboken: John Wiley), 205–217.
- Rajakaruna, N. (2018). Lessons on evolution from the study of edaphic specialization. *Bot. Rev.* 84, 39–78. doi: 10.1007/s12229-017-9193-2
- Smith, B. E., Johnston, M. K., and Lücking, R. (2016). From genbank to GBIF: phylogeny-based predictive niche modelling tests accuracy of taxonomic identifications in large occurrence data repositories. *PLoS One* 11:e0151232. doi: 10.1371/journal.pone.0151232
- Tomlinson, S., Lewandrowski, W., Elliott, C. P., Miller, B. P., and Turner, S. R. (2020). High-resolution distribution modelling of a threatened short-range endemic plant informed by edaphic factors. *Ecol. Evol.* 10, 763–777. doi: 10.1002/ece3.5933
- Triebel, D., Hagedorn, G., and Rambold, G. (2012). An appraisal of megascience platforms for biodiversity information. *Mycoskeys* 5, 45–63. doi: 10.3897/mycokeys.5.4302
- Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., and Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* 7:9132. doi: 10.1038/s41598-017-09084-6
- Wulff, A., Hollingsworth, P. M., Ahrends, A., Jaffré, T., Veillon, J.-M., L'Huillier, L., et al. (2013). Conservation priorities in a biodiversity hotspot: analysis of narrow endemic plant species in New Caledonia. *PLoS One* 8:e73371. doi: 10.1371/journal.pone.0073371
- Zizka, A., Carvalho, F. A., Calvente, A., Baez-Lizarazo, M. R., Cabral, A., Coelho, J. F. R., et al. (2020). No one-size-fits-all solution to clean GBIF. *PeerJ* 8:e9916. doi: 10.7717/peerj.9916