

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

We gathered a list of vascular plant species names by downloading species collections available for our study area from the Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/occurrence/download/0064983-160910150852091>), Consortium of Pacific Herbaria (CPH, <https://www.re3data.org/repository/r3d100012011>), Australasian Virtual Herbarium (AVH, <http://avh.ala.org.au>) and the Kew Herbarium Catalogue (<http://apps.kew.org/herbcat/>). We also downloaded type specimens from the Harvard University Herbaria (https://kiki.huh.harvard.edu/databases/specimen_index.html), the Natural History Museum (<https://data.nhm.ac.uk>), New York Botanical Garden (<http://sweetgum.nybg.org/science/collections>), Royal Botanic Garden Edinburgh (<https://data.rbge.org.uk/search/herbarium>), Smithsonian National Museum of Natural History (<https://collections.nmnh.si.edu/search/botany>), Missouri Botanical Garden (<http://www.tropicos.org>) and the Muséum national d'Histoire naturelle (<https://science.mnhn.fr/institution/mnhn/search>).

Data analysis

The list of taxonomic names obtained from public repositories was submitted to the Taxonomic Names Resolution Service (<http://tnrs.iplantcollaborative.org>) to resolve names. An independent contrast against a list of 13,073 accepted species names contained in Plants of the World Online for the 'New Guinea' locality helped confirm the robustness of the expert review. Other calculations and analyses were performed with R version 3.5.2 and map artwork was designed using QGIS v 2.18.9.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data and materials availability: The data that support the findings of this study are available as Supplementary Tables 1-4, and Extended Data Tables 1-4.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We quantified the total number of native vascular plant species in the New Guinea region by combining a data-driven approach and expert knowledge. After gathering a list of names from public repositories, we resolved the names of 23,381 taxa using two contrasting approaches: 1. A 'Big data' approach typically used in macroecological studies, i.e., using the Taxonomic Name Resolution Service software (TNRS: (<http://tnrs.iplantcollaborative.org>)). 2. An expert approach, based on the knowledge of 99 taxonomic, floristic and monographic experts of the flora of New Guinea. Experts reviewed the list of original names in their respective families of expertise and verified whether the original list of names was correctly resolved by TNRS, and included additional information about taxonomy (basionym name, basionym year), geographic range (native, endemic, distribution) and life form (tree, herb, shrub, epiphyte, palm, etc.). When experts considered that a name that was accepted by TNRS was not correct, they wrote the correct name. Similarly, when experts considered a species note to be native, they were asked to write an explanation (e.g. geographic error, taxonomic misidentification). Finally, experts also included names that had been missed from the original list. To assess the discrepancy between the TNRS and expert verification, we compared the total number of accepted species in both lists for 254 plant families using a Wilcoxon-signed rank test.

Additionally, we performed an independent contrast against a list of 13,073 accepted species names contained in Plants of the World Online for the 'New Guinea' locality (POWO; <http://www.plantsoftheworldonline.org>, accessed December 21, 2019). POWO was launched in 2017 with an initial focus on tropical Africa, but aims to become a single point of access for authoritative information on all plant species by 2020. Accordingly, for the names in the POWO list that were missing in our checklist, experts assessed if they were incorrect (i.e., synonyms) and/or not native to the study area, and which names were correct and native. The former represent false presences in POWO; the latter represent species that experts overlooked and which were subsequently included in the final checklist.

Research sample

We focused our analysis on all the native vascular plant species occurring in the New Guinea region, which we delimit as the area spanning a latitudinal range of -0.08 to -10.66 S and a longitudinal range of 129.42 to 150.21 E.

Sampling strategy

No statistical method was used for sample size calculation. The sample size in our case reflects the data that is available from public or institutional repositories for our study area of New Guinea.

Data collection

Data collection was based on natural history collections available for the study area in:

1. Public repositories, including the Global Biodiversity Information Facility (<http://www.gbif.org>), Consortium of Pacific Herbaria (<http://www.pacificherbaria.org>), Australasian Virtual Herbarium (<http://avh.chah.org.au>), Kew Herbarium Catalogue (<http://apps.kew.org/herbcat/>), Harvard University Herbaria (https://kiki.huh.harvard.edu/databases/specimen_index.html), the Natural History Museum (<https://data.nhm.ac.uk>), New York Botanical Garden (<http://sweetgum.nybg.org/science/collections>), Royal Botanic Garden Edinburgh (<https://data.rbge.org.uk/search/herbarium>), Smithsonian National Museum of Natural History (<https://collections.nmnh.si.edu/search/botany>), Missouri Botanical Garden (<http://www.tropicos.org>) and the Muséum national d'Histoire naturelle (<https://science.mnhn.fr/institution/mnhn/search>).
2. Institutional repositories of the Naturalis Biodiversity Center, Royal Botanic Gardens Kew and University of Papua New Guinea.
3. Databases curated by taxonomists for the families Orchidaceae, Arecaceae, Araliaceae and Cyatheaceae.

Timing and spatial scale

We used all collections available through time for the study area.

Data exclusions

We pre-established different filters to restrict our analyses on vascular plants and clean scientific names: we excluded fungi, lichens, algae, bryophytes and marine species (e.g., sea grasses); doubtful species identifications (e.g., 'cf.', 'sp. nov.', 'aff.', 'sp.') were classified to generic level; Misspelled genera were manually corrected and doubtful cases excluded; we removed all known hybrids from the analyses.

Reproducibility

As this is not an experimental study, replication was not conducted. All data used in our analysis is made available as Supplementary Tables. Comparison of names among datasets was done using Excel and the R computer programming language and the analysis code is available on request from the corresponding author.

Randomization

Our study was not experimental but based on taxonomic knowledge and on published literature of the vascular plants of New Guinea. Thus, no randomization was required.

Blinding

Investigators were not blinded during data acquisition or analysis because our study is not experimental.

Did the study involve field work?

☐ Yes☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging