nature research

Corresponding author(s): Samuel F. Brockington

☐ Initial submission ☒ Revised version ☐ Final submission

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

## Experimental design

1. Sample size

Describe how sample size was determined.

The samples sizes were essentially determined by available data sources. We used BGCI 'GardenSearch' (www.bgci.org/garden_search.php) database (accessed 2016-01-01) for the location of botanic gardens. For the presence and absence of species from gardens we used BGCI 'PlantSearch' (www.bgci.org/plant_search.php) (accessed 2016-01-01). For threatened plants we used a pre-release version of BGCI's ThreatSearch (https://www.bgci.org/threat_search.php) (accessed 2016-01-01). At the point of access, and still currently, these data sources represent the largest sample sizes of data available to analyse the global state of ex-situ conservation by the world's botanic garden network. While in this scenario, more data is always welcome, these datasets are sufficient to reveal the fundamental patterns driving ex-situ plant conservation across the global botanic garden network, and to estimate baseline statistics with respect to conservation achievement.

2. Data exclusions

Describe any data exclusions.

Data sources were subject to extensive cleaning which led to data exclusion athe start of the process. We were interested in biological species only with accepted taxon names. However the data sources were derived from uploads from many institutions world wide, who use a variety of taxonomic systems and grow a range of horticultural varieties and cultivars, not just biological species. For all datasets, records were filtered to remove assessments of taxa that were not land plants e.g fungal, algal, and animal taxa. Undescribed taxa were ignored for these analyses e.g. "Asparagus sp. nov. A" . We discarded 'orphan' BGCI plant records that were not currently associated with any gardens in the network (e.g. historical records of dead plants that are no longer held in a garden). We also discarded records of horticultural taxa such as cultivars, as we were interested in true biological species. We computationally-normalised the taxonomy of records using the R package Taxonstand version 1.8, so that all taxa match an accepted or unresolved taxon listed by The Plant List v1.1 . Raw input species names that could not be automatically matched to a species name listed at The Plant List v1.1 were manually resolved to the correct species name. By matching to TPLv1.1 in a minority of cases we were back-converting names into older ones for the sake of consistency. BGCI records were de-duplicated using the R package stringdist using Damerau-Levenshtein distance so that there was only one record for each unique taxon, as gardens around the world can apply different names to the same taxon. After normalisation to The Plant List (TPL) some taxa were demoted from species rank in the original assessment to subspecies rank. For consistency and comparability only species-level taxa were retained for analysis, subspecies taxa were discarded. GBIF georeferenced location data was also cleaned to filter-out corrupt data indicated by coordinates that did not match the country stated on the record, or that had coordinates in marine areas. We also refined location data to include only species with at least five georeferenced occurrences, whose latitudinal range is either temperate or tropical.

3. Replication

Describe whether the experimental findings were

Each analysis was repeated multiple times, to check internal consistency.

reliably reproduced.

Biogeographic analyses which used GBIF location data were repeated using a number of cut-offs (>5 georeferenced location, and >10 georeferenced location) to confirm that the findings held. We ultimately reported findings at the >5 georeferenced locations

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Randomisation was not relevant in our analyses because all analyses depended on knowing the location of species within botanic gardens, and around the world. We did not perform any modeling estimates, that required random sampling.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding was not relevant in our analyses because all analyses depended on our knowing and interpreting the location of species within botanic gardens, and around the world.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ A statement indicating how many times each experiment was replicated |
| ☐ | ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☐ | ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted |
| ☐ | ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☒ | ☐ Clearly defined error bars |

*See the web collection on statistics for biologists for further resources and guidance.*

## Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Various analyses were implemented in the software package 'R', with individual packages and the version used cited in the methdology, and repeated here: 'Taxonstand' v1.8; 'stringdist' v0.9.4.4; 'rgbif' version 0.9.7; 'chloroplethr' 3.6.1; 'fmsb' v0.6.1 . The software programme FigTree v1.4.3 was used to visualise and draw the phylogenetic tree. Prism V7.0c was used to draw all graphs. Illustrator 21.0.0 was used to assemble figures.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique material were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used. — No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used. — No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination. — No eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. — No commonly misidentified cell lines were used.

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study. — No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants. — No human research participants were used.