

Research coverage of three plant families under the Global Biodiversity Framework

39,052 accepted species in Orchidaceae, Apocynaceae and Lecythidaceae, assessed against Targets 4, 20 and 21 of the Kunming-Montreal Global Biodiversity Framework.

COMPASS ID, October 2026. Prepared for the seventeenth meeting of the Conference of the Parties to the Convention on Biological Diversity, Yerevan, 19 to 30 October 2026. Version 1.0, released 16 October 2026. Every figure in this document is produced by a script in the COMPASS ID repository and checked by the published-figures audit before release.

Summary

This is the first instalment of a repeatable assessment of which accepted vascular plant species have indexed research, where that research was carried out, and by whom. The denominator is fixed for the whole series: Version 16 of Kew's World Checklist of Vascular Plants (WCVF), 365,813 accepted species. Report 1 covers three families, 39,052 accepted species, 10.7 percent of the checklist, and 63,782 indexed papers linked to them.

- 29,018 of the 39,052 species (74.3 percent) have no paper in the indexed literature whose title or abstract names the species or one of its checklist synonyms.
- Of 1,574 species assessed as threatened on the IUCN Red List, 1,082 (68.7 percent) have no indexed paper. Of 334 Data Deficient species, 279 (83.5 percent) have none. Of 1,283 Least Concern species, 617 (48.1 percent) have none. Only 3,375 of the 39,052 species (8.6 percent) carry any Red List assessment.
- Research is not done where the species are. Of species with known native range countries, 14.3 percent of orchids, 17.5 percent of Apocynaceae and 28.5 percent of Lecythidaceae have at least one located study inside a country they are native to. Papua New Guinea holds 3,356 species of these families and 39 located studies; of the 16 of those studies with recorded author institutions, 13 have no institution in the country.
- The indexed literature almost never engages the Convention's instruments by name: 14 of 63,782 papers mention the CBD, the Framework, the Aichi Targets, the Global Strategy for Plant Conservation, a national strategy, or the Nagoya or Cartagena Protocols.
- Content can be mapped onto the Framework's targets, but only some targets reliably. A blind check of a documented tag-to-target crosswalk showed that research serving Target 4 (species recovery) is identified with 98 percent agreement, restoration (Target 2) and invasive species (Target 6) with 75 percent, sustainable use (Target 9) with 72 percent, and most other targets at 50 percent or below. Target 4 research is 5.0 percent of orchid papers, 0.7 percent of Apocynaceae papers and 0.8 percent of Lecythidaceae papers.

1. Scope and method

Census

A species is in the census if WCVF Version 16 (extracted 4 June 2026, DOI 10.34885/egs6-cp24) lists it as an accepted species of the family. Names the checklist treats as synonyms are kept and their papers are counted for the accepted species. Orchidaceae 31,892 accepted species; Apocynaceae 6,759; Lecythidaceae 401. Range countries come from the WCVF distribution table (TDWG level-3 areas mapped to ISO countries; native and historical-native areas as the native range, introduced areas held separately, doubtful records excluded).

Papers

Every accepted name and synonym was searched in OpenAlex titles and abstracts (harvest completed 1 to 3 October 2026). A paper is linked to a species when the name appears as a whole term. Link precision was measured by an adversarial role judge on this corpus: 84 percent of sampled orchid links are papers that study the species (95 percent interval 78 to 89, 180 links from 130 species); a further 12 percent name the species without studying it.

Tags

Each paper with an abstract was read against a controlled codebook of 46 shared and 22 family-specific dimensions. A tag survives only if its value is in the codebook, its evidence is a quote found in the abstract, and an independent judge model confirms it against the codebook definition. On the finished orchid corpus a second, blind independent model confirmed 85.5 percent of 200 sampled tags (95 percent interval 80 to 90); on the platform's earlier published corpus the same check gave 79.8 percent.

Places

Study sites are read from each abstract, resolved through GeoNames and OpenStreetMap, then hardened: sites outside the species' known countries, places merely cited, and range statements mistaken for sites are removed by a second pass. Author affiliation countries come from OpenAlex. A paper is "located" when it carries at least one hardened study site.

What the measure is

"No indexed paper" means no paper indexed by OpenAlex matched the species or a synonym. It is not a claim that nothing is known. Literature not indexed by OpenAlex, grey literature, herbarium records and local-language journals are outside it, and the size of that exclusion is partly measurable (section 4).

2. Research coverage by family and by threat status (Target 4)

CENSUS, INDEXED PAPERS AND IUCN RED LIST COVERAGE, THREE FAMILIES

MEASURE	ORCHIDACEAE	APOCYNACEAE	LECYTHIDACEAE	THREE FAMILIES
Accepted species (WCVP v16)	31,892	6,759	401	39,052
Indexed papers linked	26,501	33,900	3,588	63,782
Species with no indexed paper	24,503 (76.8%)	4,292 (63.5%)	223 (55.6%)	29,018 (74.3%)
Species with a Red List assessment	2,237 (7.0%)	874 (12.9%)	264 (65.8%)	3,375 (8.6%)
Threatened species (CR, EN, VU)	1,127	346	101	1,574
Threatened species with no indexed paper	778 (69.0%)	227 (65.6%)	77 (76.2%)	1,082 (68.7%)
Data Deficient species with no indexed paper	232 of 278	35 of 42	12 of 14	279 of 334 (83.5%)
Least Concern species with no indexed paper	418 of 710	142 of 440	57 of 133	617 of 1,283 (48.1%)

Three patterns hold in every family.

1. Research effort does not rise with threat status. Threatened species are no better covered than the family average, and Data Deficient species, the ones whose status cannot be assessed for lack of information, are the least covered of all.
2. Assessment coverage is the first gap. Over 91 percent of these species have never been assessed for the Red List, so the threatened figures above describe a known fraction, not the family.
3. Research concentrates on a few economically useful species. Six cultivated orchid genera account for 40.6 percent of orchid papers; in the Apocynaceae, phytochemistry and pharmacology dominate; in the Lecythidaceae, the Brazil nut.

Target 4 of the Framework asks for urgent management action to halt extinctions and recover species. Research that serves it directly, meaning ex-situ conservation, reintroduction and propagation, species monitoring and Red List assessment, is 1,323 orchid papers (5.0 percent of the corpus), 248 Apocynaceae papers (0.7 percent) and 30 Lecythidaceae papers (0.8 percent). Section 4 describes how these counts were produced and checked.

3. Where research happens relative to where species occur (Target 20)

Target 20 concerns capacity, technology transfer and scientific cooperation. The indexed literature records two things that bear on it: where the fieldwork was done, and where the authors' institutions are.

For every country, the table lists the species of the three families native to it, the share with no indexed paper, the share with at least one located study inside the country, the number of located studies inside the country, and the share of those studies with recorded author institutions that include no institution in the country.

NATIVE SPECIES, RESEARCH COVERAGE AND AUTHOR INSTITUTIONS, SELECTED COUNTRIES

COUNTRY	NATIVE SPECIES (3 FAMILIES)	NO INDEXED PAPER	SPECIES WITH A STUDY SITE IN THE COUNTRY	LOCATED STUDIES IN THE COUNTRY	STUDIES WITH NO DOMESTIC AUTHOR INSTITUTION
Indonesia	6,793	81.7%	6.5%	755	8.1% of 484
Colombia	5,068	77.6%	6.2%	303	31.0% of 216
Ecuador	4,724	80.4%	5.4%	191	36.8% of 136
Brazil	3,892	64.3%	18.9%	1,700	4.2% of 1,037
Papua New Guinea	3,356	88.2%	1.2%	39	81.3% of 16
Peru	2,918	73.6%	6.1%	164	31.1% of 119
Malaysia	2,582	65.3%	5.5%	137	28.1% of 64
Venezuela	2,263	66.9%	3.7%	55	56.4% of 39
China	2,040	29.5%	31.3%	1,210	5.8% of 830
Mexico	1,849	57.9%	18.8%	439	9.1% of 317
India	1,781	27.5%	37.6%	1,278	3.0%
Australia	1,720	69.5%	15.3%	299	6.2%
Madagascar	1,310	78.1%	10.2%	109	78.5%
South Africa	1,193	57.8%	12.2%	176	1.5%
United States	413	30.3%	38.5%	527	3.2%

Full table for 231 countries: [data/country-table.json](#).

Three observations, each stated at the scope of these three families, not of national floras.

1. The countries with the most species have the least research per species. Indonesia, Colombia, Ecuador and Papua New Guinea together hold 19,941 native species of these families; between 78 and 88 percent have no indexed paper, and between 1 and 7 percent have ever had a located study inside the country.
2. Where fieldwork does happen in the most diverse countries, it is often done from elsewhere. In Papua New Guinea 13 of 16 located studies with recorded institutions have no domestic author; in Madagascar 79 percent; in Venezuela 56 percent; in Colombia, Ecuador and Peru between 31 and 37 percent. In Brazil, China, India, South Africa and the United States the share is below 6 percent.
3. Capacity and coverage move together. South Africa and Madagascar hold comparable numbers of these species (1,193 and 1,310). South Africa has 12 percent of its species studied at home and 2 percent of in-country studies without a domestic author; Madagascar has 10 percent and 79 percent. These are three families, not national floras; the full-flora comparison is the subject of a later instalment.

4. What the indexed literature can and cannot show (Target 21)

Explicit engagement with the Convention

A tag fires for a named instrument only when the abstract engages it. Across 63,782 papers, 14 name the CBD, the Kunming-Montreal Framework, the Aichi Targets, the Global Strategy for Plant Conservation, a national biodiversity strategy, or the Nagoya or Cartagena Protocols: 8 orchid papers, 5 Apocynaceae papers, 1 Lecythidaceae paper. Plant research on these families is conducted almost entirely without reference to the policy framework it is expected to inform.

Mapping content to targets

Because papers do not name targets, we tested a documented crosswalk from the evidence-grounded tags to the 23 targets (phase1/ref/gbf_crosswalk_v2.json). Each target lists the tag values whose codebook definition is the target's subject. A paper counts under a target when it carries at least one such confirmed tag. The crosswalk was checked blind: an independent model, given the target label, the paper and the triggering tag, judged whether the paper's research content contributes to what the target aims at.

BLIND AGREEMENT ON THE TAG-TO-TARGET CROSSWALK, AND PAPERS COUNTED UNDER EACH TARGET

TARGET	AGREEMENT IN THE BLIND CHECK	ORCHIDACEAE PAPERS	APOCYNACEAE PAPERS	LECYNTHIDACEAE PAPERS
4 Species recovery	59 of 60 (98%)	1,323 (5.0%)	248 (0.7%)	30 (0.8%)
2 Restoration	45 of 60 (75%)	73	196	61
6 Invasive species	45 of 60 (75%)	82	208	9
9 Sustainable use	43 of 60 (72%)	383	562	432 (12.0%)
Any target, first-version crosswalk	93 of 194 (48%)			

The first version of the crosswalk agreed with the independent reviewer on only 48 percent of 194 sampled pairs (95 percent interval 41 to 55). The failures were systematic: a tag describing where or what a paper studies was being read as what the paper is for. A study set in a city is not research on urban biodiversity; a measurement of metal content in a medicinal plant is not research on pollution reduction; documenting who uses a plant is not research on participation and rights. The revised crosswalk retains those mappings only in a contributory band, and this report quotes as measured only Target 4, with Targets 2, 6 and 9 shown with their agreement rates. Targets 16, 18 and 23 cannot be measured from research content at all. The validation records are published with the crosswalk.

Indexing and language

OpenAlex coverage of local-language journals is incomplete, and the gap is not uniform. Among Apocynaceae papers with a recorded language, 15.5 percent are not in English (Portuguese 1,632, Spanish 889, Indonesian 706, German 502, French 454). Among Lecythydaceae papers, 40.0 percent are not in English, 833 of them Portuguese. Part of the apparent gap in Lusophone, Hispanophone and Francophone countries is therefore an indexing gap, and the country table should be read with that in mind; Madagascar's Francophone literature is a specific case. Language metadata for the orchid corpus is still being back-filled and is not reported here.

What this measure is not

A count of indexed papers is not a measure of knowledge. Herbarium specimens, floras, grey literature and the knowledge of local people are outside it. The measure is of the published, indexed record that decision makers and researchers can actually find, which is the record Target 21 concerns.

5. What the data supports

1. **Assessment before research.** For these families, the Red List has reached 8.6 percent of species. The 279 Data Deficient species with no indexed paper and the 29,018 unassessed, unstudied species are where the first evidence is missing, and where new field data would change a status rather than refine one.

2. **Capacity where the species are.** In the four most species-rich countries of these families, under 7 percent of species have ever had a study inside the country, and in Papua New Guinea and Madagascar most in-country studies are led from abroad. Target 20 investment that places research capacity in those countries would address the largest measurable gap in this record.
 3. **Make the Framework legible in research.** Fourteen of 63,782 papers engage the Convention's instruments. A content-to-target mapping can be measured and audited for a few targets, as shown here; for most it cannot. If the Parties want research to be reported against the targets, the targets need indicators that research can be mapped to, not the reverse.
 4. **Extend the denominator.** This report covers 10.7 percent of accepted vascular plants. The same measure, census, harvest and location, without the content tagging, can be extended to the whole checklist within the series, and the 2030 edition can report change rather than a snapshot.
-

Appendix A. Reproducibility

All figures are produced by the scripts below from the published COMPASS ID database and are re-checked by `phase1/audit_published_figures.js` before release.

- Census and roll-up: `phase1/wcvp_reconcile.js`, `phase1/wcvp_load_absent.js`, `phase1/rollup_from_accepted.js`; reference file `phase1/ref/wcvp_v16_<family>_names.csv`.
 - Range countries: `phase1/wcvp_range.js` with `phase1/ref/tdwg_wgsrpd_level3.txt` and `level4.txt`.
 - Target counts: `phase1/gbf_target_counts.js`; crosswalk `phase1/ref/gbf_crosswalk_v2.json`; outputs `data/gbf-<family>.json`.
 - Crosswalk validation: `phase1/gbf_crosswalk_validate.js`; records `data/gbf-crosswalk-validation.json` (v1, seed 20261004), `data/gbf-crosswalk-validation-v2.json` (seed 20261005), `data/gbf-crosswalk-validation-v2-deep.json` (Targets 2, 4, 6, 9 at 60 pairs each, seed 20261006).
 - Country table: `phase1/gbf_country_table.js`; output `data/country-table.json`.
 - Tag precision: `phase1/rejudge_gate.js`; record `audit/rejudge/2026-10-02-orchidaceae-n200.json`. Link precision: `phase1/run.js validate-links`; record `audit/2026-10-04/validate-links-flora-orchidaceae.json`.
 - Atlases with the full per-family figures: `orchidaceae.html`, `apocynaceae.html`, `brazil-nut.html`.
-

Appendix B. Definitions

- **Accepted species:** a name WCVF Version 16 lists with status Accepted at species rank in the family; natural hybrids included, artificial hybrids and unplaced names excluded.
- **No indexed paper:** no OpenAlex-indexed work whose title or abstract contains the accepted binomial or a WCVF synonym as a whole term, after roll-up of synonym links to the accepted species.
- **Located study:** a paper with at least one study-site placement that survived hardening (sources `geoparse+geonames`, `geoparse+osm` and refinements, or coordinates reported in the text).
- **Native range country:** a country containing a TDWG level-3 area the WCVF distribution table lists for the species as native or historically native, excluding doubtful records; introduced areas are held separately and not used here.
- **Domestic author institution:** at least one OpenAlex-recorded author institution located in the same country as the study site.
- **Threatened:** IUCN categories CR, EN and VU from the Red List reference extract (assessment years 1996 to 2025).

How to cite

COMPASS ID, Inc. (2026). Research coverage of three plant families under the Global Biodiversity Framework. Research coverage of the world's vascular plants, 2026 to 2030, Report 1, version 1.0, released 16 October 2026. www.compassid.org/reports/research-coverage-vascular-plants-2026-report-1.html

REVISIONS

- Figures revised after the Orchidaceae re-harvest of seven genera (10 October 2026).
- Study-site and author-institution figures revised after the places republish of 12 and 13 October 2026.

Version 1.0 released, 16 October 2026.