# CactEcoDB: Trait, spatial, environmental, phylogenetic and diversification data for the cactus family

Jamie B. Thompson[1*], Catherine Martinez[1,2], Jorge Avaria-Llautureo[1], Santiago Ramírez-Barahona[3], Gerardo Manzanarez-Villasana[3], Alastair Culham[1], Andrew Gdaniec[1], Chris Venditti[1], Georgia Keeling[4], Nicholas K. Priest[4*]

[1] School of Biological Sciences, University of Reading, Reading, UK
[2] School of Archaeology, Geography and Environmental Science, University of Reading, Reading, UK
[3] Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México (UNAM), Circuito Exterior s/n, Ciudad de México, Mexico
[4] The Milner Centre for Evolution, Department of Life Sciences, University of Bath, Bath, UK
* Corresponding authors

## Abstract

Integrated datasets linking traits, spatial distributions, environmental variables and phylogenies are essential for comparative research, but remain limited for many plant taxa, including those which are most threatened. Cactaceae are a morphologically and ecologically diverse succulent family that are iconic components of ecosystems across the Americas, and face high extinction risk. To support future comparative research, we present CactEcoDB (The Cactus Ecological Database), an Open Access dataset of curated spatial, ecological, trait, phylogenetic, and diversification data for over 1,000 cactus species. CactEcoDB includes species-level trait data, geographic occurrence records, environmental variables, range size estimates, speciation rates, and the largest time-calibrated phylogeny of cacti to date. By integrating these diverse data in a single platform, CactEcoDB enables analyses of trait evolution, trait–environment interactions, biogeography, climatic niche evolution, and macroevolutionary dynamics in one of the most celebrated and threatened plant families.

## Background & Summary

Recent research highlights the importance of comparative methods for our understanding of the origins and ecological drivers of global biodiversity in large families of plants [1–12]. This type of comparative research relies on comprehensive datasets that adequately capture the complexity of trait evolution, environmental variation, and lineage diversification [13–35]. There is a growing appreciation of how this complexity, especially ecological interactions, coevolutionary dynamics, and environmental heterogeneity, shapes macroevolutionary and macroecological trajectories [9,10,12,36–41]. As global change accelerates and data availability expands unevenly across lineages, so does the need for integrative, high-quality datasets for threatened and unique groups

[25,26,42–50]. However, assembling such datasets remains difficult and time-consuming, often requiring expert curation and synthesis across disparate data types and sources.

The cactus family (Cactaceae) is among the most morphologically and ecologically distinctive plant lineages [51–57]. Comprising approximately 1,850 species, cacti are found in an extraordinary range of habitats, from hyper-arid deserts and tropical dry forests to high-elevation grasslands and volcanic islands [58]. Centers of species richness are found in Mexico, the Andes, and Eastern Brazil [10,51,58], but their distribution extends from southern Canada to Patagonia, with one species probably occurring naturally outside of the Americas (*Rhipsalis baccifera*) [59]. Cacti are well known for their highly specialised morphological and physiological adaptations to water-limited environments, including stem succulence, heavily modified leaves, thick cuticles, and crassulacean acid metabolism (CAM) photosynthesis [10,12,51,54]. These traits make them a classic example of adaptive evolution in plants, and in particular, a model system for understanding evolutionary responses to extreme environmental conditions.

Comparative research in the Cactaceae has bearing on diverse fields including plant physiology [60–62], biogeography [51,63–65], ecology [66–68], evolutionary dynamics [10,12,51,55], and conservation biology [53,69–73]. The growing clarity of phylogenetic relationships within the family [52] has facilitated tests of the evolution of key traits such as growth forms [60], pollinators [51,56], and chromosome count [74]. The power to detect drivers of diversification is greatly aided by the remarkably wide distribution and range of growth forms found in diverse environments across gradients including aridity, temperature, and elevation [51,52,55,74].

Beyond their ecomorphological adaptations, cacti hold exceptional cultural and economic significance [75–77]. Archaeological evidence shows that species such as *Opuntia ficus-indica* were consumed more than 9,000 years ago in Mesoamerica [78,79], where they continue to serve as important sources of food, forage, and dye [80–82]. Today, cacti are cultivated globally for ornamental purposes, and their remarkably varied forms, ranging from tiny globose species to massive columnars [60], make them among the most collected and traded groups in horticulture [83] along with orchids. They are unusual in horticulture in that the bulk of trade is of natural species, not highly bred and modified cultivars. The deep cultural history, combined with their horticultural desirability, has led to intense pressure on wild populations through unsustainable collection and illegal trade [84,85]. As a result, Cactaceae is one of the most threatened plant families. A 2015 assessment found that 31% of cactus species are threatened with extinction, with habitat loss, climate change, and overexploitation cited as primary drivers [69]. More recent projections suggest that 60-90% may experience future range contractions under near-future climate change scenarios [70].

The integration of phylogenetic, trait, and spatial data at the species level in cacti has remained fragmented, limiting the ability of researchers to address broad ecological and evolutionary questions, or to support conservation assessments grounded in comprehensive ecological data [71,86]. To fill this gap, we present CactEcoDB, a curated dataset that integrates spatial, ecological, trait, and phylogenetic information for over 1,000 species of Cactaceae. This is a greatly updated and expanded version of previous datasets [10]. CactEcoDB includes biotic trait data (e.g., growth form, plant height, pollinator), expert-curated geographic ranges, numerous important environmental variables, and the largest time-calibrated molecular phylogeny for the group. As compared with the largest previous dataset [10], CactEcoDB expands and improves the available data: i) it increases in species sampling for plant height by 27% and spatial variables by ~6%; ii) it expands the number of classified growth forms from two to nine; iii) it replaces highly incomplete GBIF data with expert ranges reported by the IUCN [53,69–73,87], which were manually verified against independent data, especially the POWO [88]; and iv) it provides a general-purpose resource for ecological, evolutionary, biogeographic, and conservation-focussed research. Furthermore, we have substantially expanded and updated several of the biotic traits with higher resolution and species coverage than previously available [10]. By bringing together large-scale data covering all aspects of cactus diversity in an Open Access resource, CactEcoDB enables new opportunities for research on trait-environment relationships, climatic niche evolution, and conservation, in one of the most unique and threatened plant families. Given the gradual increase of trait, spatial and phylogenetic data generally, we welcome the submission of additional data, which will improve future research into cacti.

## Methods

The methods presented here include a brief description of methods used to generate previous datasets [10], and our new efforts designed to improve and expand sampling of spatial distributions, environmental data, and traits.

### Phylogenetic reconstruction

The largest available time-calibrated phylogeny of Cactaceae was reconstructed using the supermatrix approach [10], to capitalise on existing data and maximise species-level sampling (Figure 1). For this, 18 published plastid and nuclear loci were assembled from GenBank, by identifying and clustering orthologous sequences with the OneTwoTree pipeline [89], and aligned them using MAFFT [90] with quality inspection undertaken in SeaView [91]. Partial sequences were merged with full-length sequences for each species where homologous sequences clustered into a fragmented and full cluster, retaining the longest version. Outgroup sequences from Anacampserotaceae,

Portulacaceae, and Talinaceae were added using MAFFT's "--add" function [92]. After alignment, poorly aligned regions were trimmed using trimAl with the "gappyout" setting [93], and concatenated all loci into a supermatrix using AMAS [94]. A Maximum Likelihood (ML) phylogeny was reconstructed with RAxML v8 [95], applying a GTR+G model partitioned by locus. Topological support was estimated with automatic bootstrapping [96] and the final tree was time-calibrated using Penalized Likelihood with treePL [97]. Stem and crown ages for the family were implemented as uniform constraints, which were derived from a relaxed-clock reconstruction of angiosperm divergence times [5], in the absence of a useful fossil record of Cactaceae [55]. Cross-validation was used to optimize the smoothing parameter across 100 priming runs, and the final dating was based on the smoothing value with the lowest score.

**Diversification rate estimation**

To estimate species-specific speciation rates and identify lineage rate shifts, we used Bayesian Analysis of Macroevolutionary Mixtures (BAMM) [98] (Figure 1). Prior to analysis, outgroups were pruned and incomplete sampling was accounted for by providing BAMM with genus-level sampling fractions. Four Markov Chain Monte Carlo (MCMC) chains were run for 50 million generations with BAMM, sampling every 5,000 generations and discarding the first 10% as burn-in. Priors were set using the BAMMtools package in R [99], and a conservative prior allowing for a single rate shift was chosen to minimise overfitting. Convergence was assessed using the coda R package, ensuring effective sample sizes (ESS) of >200 for all key parameters. Following this, posterior distributions were summarized with BAMMtools, the best shift configuration was estimated, and mean speciation rates at the tips were extracted for each species. As discussed in [10], we provide the entire event data file from BAMM to enable analysis of lineage diversification dynamics, with the caveat that exact divergence timings in cacti are uncertain due to the absence of within-group fossils that would augment specific lineage timings.

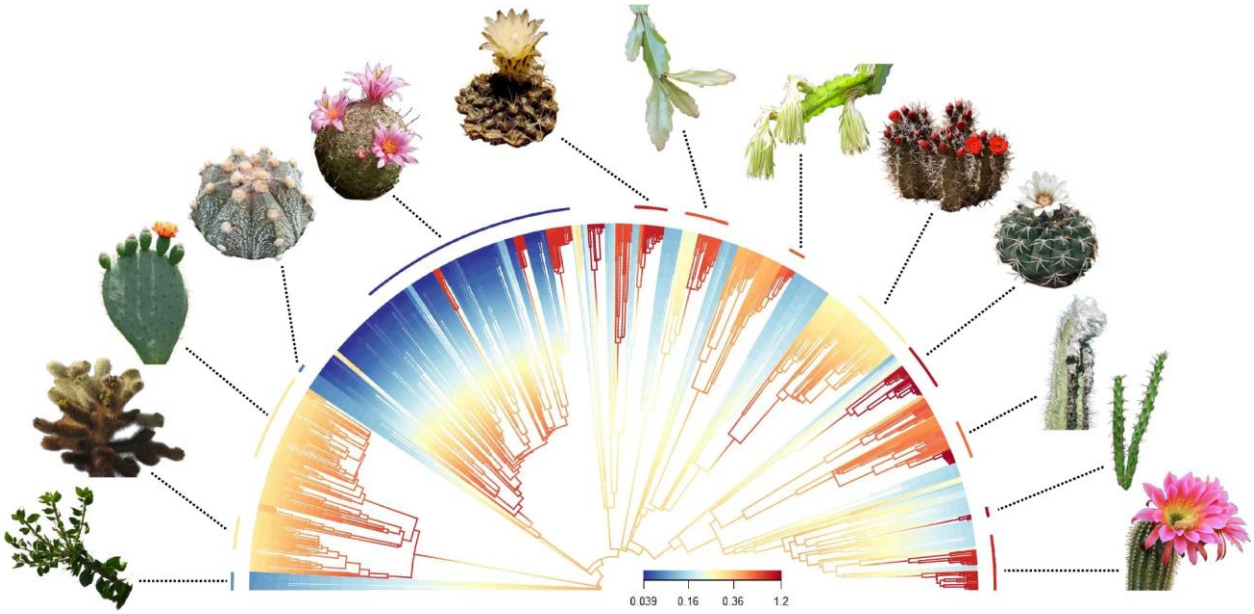


**Figure 1: The phylogeny, showing estimated diversification rate variation.** Phylogenetic branches are coloured according to speciation rates estimated with BAMM [98], and vary 32-fold. Arc segments of median speciation rate for thirteen morphologically varied cactus genera are indicated. This figure and figure legend is reproduced with permission from [10]. Cactus images are used under Creative Commons with modifications allowed. From left to right: images 1, 3, 8, 11, 12, and 13 used photos taken by Amante Darmanin, Forest & Kim Starr, John Tann, Renee Grayson, and Wendy Cutler, which are licensed under a Creative Commons Attribution 2.0 License (https://creativecommons.org/licenses/by/2.0/). Image 2 used a photo marked as being in the Public Domain (https://creativecommons.org/publicdomain/mark/1.0/). Images 4 and 10 used photos taken by Leonora Enking and Lyubo Gadzhev, which are licensed under a Creative Commons Attribution-ShareAlike 2.0 License (https://creativecommons.org/licenses/by-sa/2.0/). Images 5, 7 and 9 used photos marked as being in the Public Domain using the CC0 1.0 Universal Public Domain Dedication (https://creativecommons.org/publicdomain/zero/1.0/). Image 6 used a photo taken by Christer Johansson, which is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/deed.en).

**Trait data compilation**

We compiled species-level trait data from a combination of authoritative textbooks [100], primary taxonomic literature, specimen descriptions in online floras, and prior datasets [10,51,60]; these data, vary in their coverage across cactus species and several are not standardised (Figure 2). Key traits that we updated, expanded, and standardised in our new dataset, include plant height, growth form, pollination syndrome, and chromosome count.

We standardised traits previously recorded with different units, e.g. for height. Where minimum and maximum plant heights were reported, we recorded maximum values to reduce biases from juvenile, diseased, or partial specimens [101]. Descriptions such as “ground level” were scored as 0 cm. We expanded and updated all records for plant maximum size to improve species sampling by 201 species, from 750 to 951.

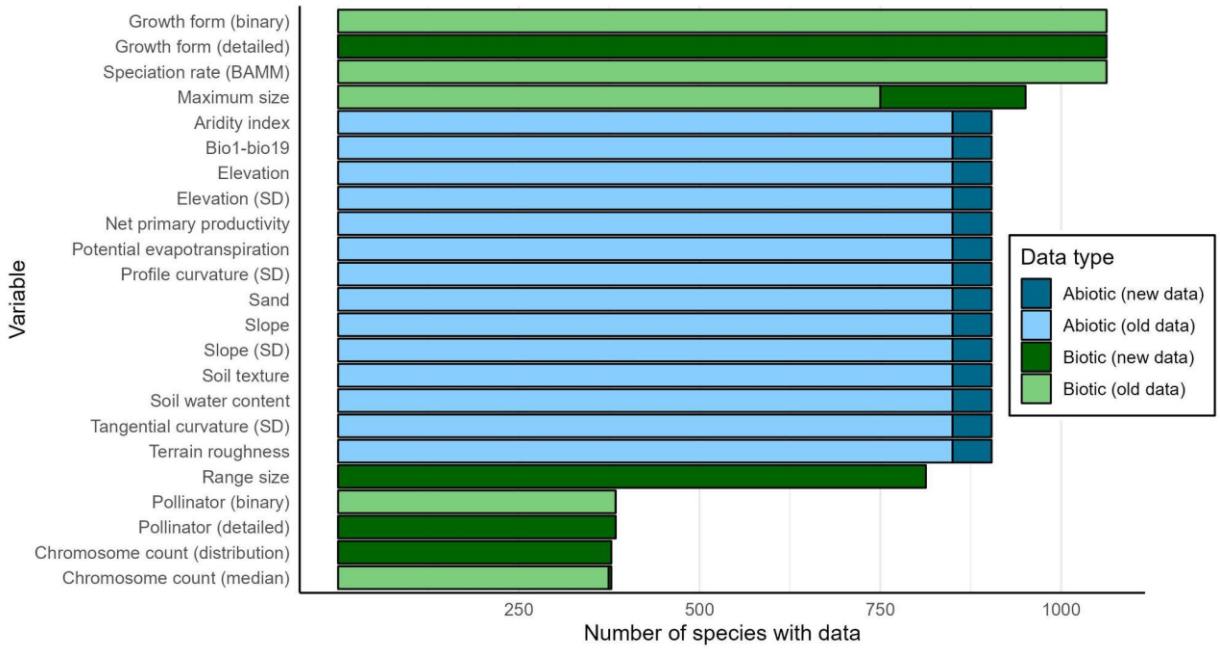


**Figure 2: The distribution of species coverage by variable.** Speciation rate, growth form and plant size (maximum height or length) have highest coverage (1,063, 1,063 and 951 species, respectively), and the spatial environmental variables sampled 904 species, which is determined by occurrence availability. We only calculated range size for those species in IUCN, given the occurrence scarcity in GBIF for the remaining species. Pollinator and chromosome count have lower coverage with 384 and 378 species, respectively. We compare the species coverage for variables in CactEcoDB against those in the largest previous dataset [10].

For growth forms, we provide data on the binary scoring system developed by Hernández-Hernández et al. [51] (see [10]), which included two main categories grouping six subcategories: compact (globose caespitose, globose solitary, and barrel-shaped) and arborescent (columnar, shrubby, and arborescent). We also devised a more refined system using nine growth form categories: arborescent, barrel, columnar, cushion forming, epiphytic, geophytic, globose caespitose, globose solitary, and shrubby based on morphological descriptions primarily drawn from [100] (Figure 3). For context, a previous growth form system used in (Hernández-Hernández et al., 2011) scored species as arborescent, shrubby, columnar, globose solitary, globose caespitose and

barrel. Briefly, arborescent cacti are those with a main trunk that branches above the base, typically with a tree-like structure; shrubby cacti lack a dominant trunk and instead exhibit basitonic branching from the base; columnar species are unbranched with elongated stems, including tall or prostrate cylindrical forms; globose solitary and globose caespitose refer to spherical-stemmed species under 0.5 m tall, growing singly or in clumps, respectively; barrel cacti have spherical stems exceeding 0.5 m in height; cushion-forming species grow in dense mats without a clearly assignable alternative form; geophytic species grow at or below ground level but are not cushion-forming; and epiphytic species are those growing on other plant hosts. Epiphytes were scored according to EpiList 1.0 [102], a vetted compilation of epiphytes. We hypothesise that this system of nine growth forms adequately explains growth form variation in cacti.

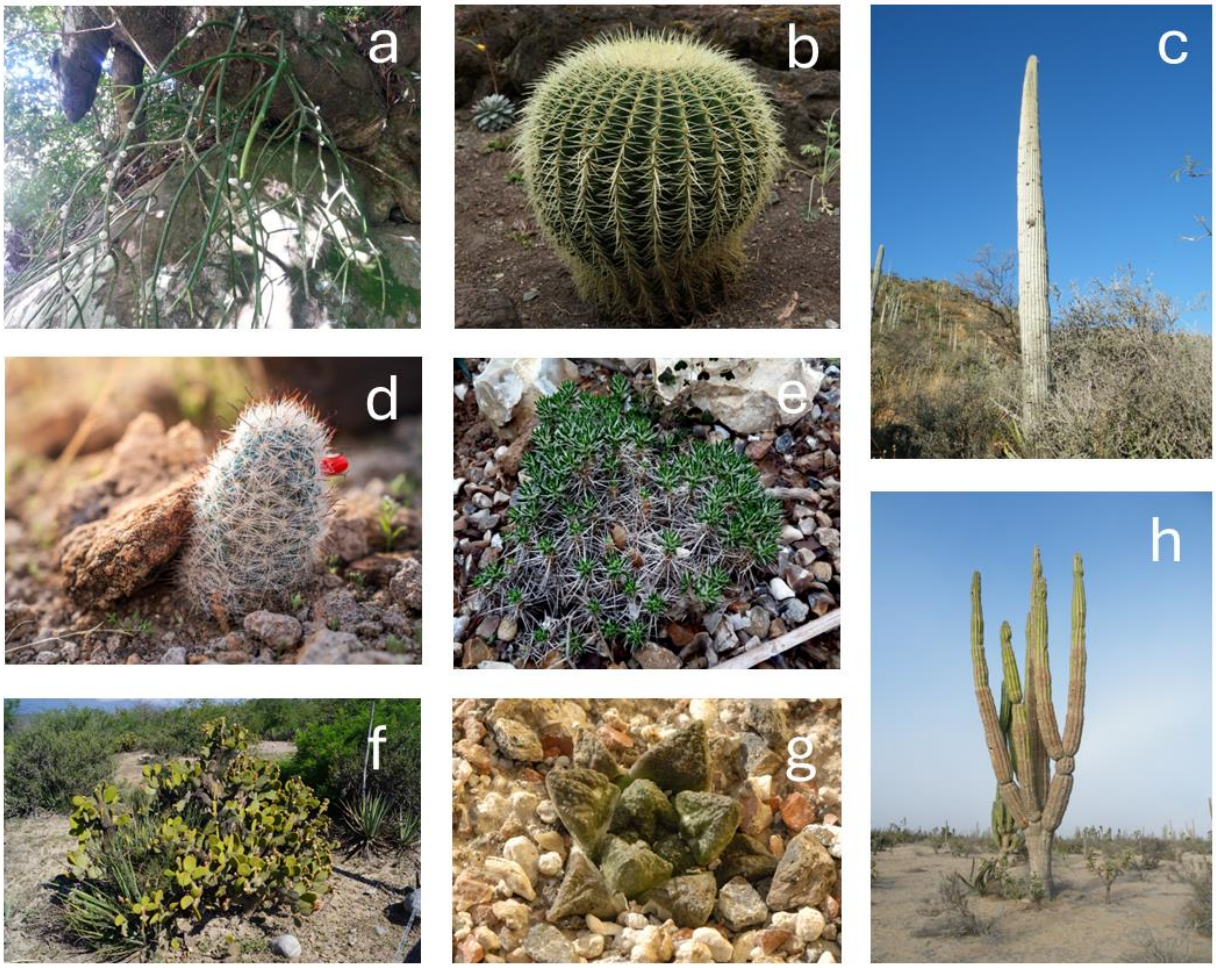


**Figure 3: Examples of each growth form in CactEcoDB.** (a) *Rhipsalis baccifera*, an epiphytic cactus. (b) *Echinocactus grusonii*, a barrel cactus. (c) *Cephalocereus columna-trajani*, a columnar cactus. (d) *Mammillaria tetrancistra*, a globose solitary species. Note that we have not shown a globose caespitose species, which appears as

clusters of globose solitary species. (e) *Maihuenia poeppigii*, a cushion-forming species. (f) *Opuntia microdasys*, a shrubby species. (g) *Ariocarpus fissuratus*, a geophyte. (h) *Pachycereus pringlei*, an arborescent species. Cactus images are used under Creative Commons licenses with modifications allowed. Image (a) used a photo taken by Maria Vorontsova, which is licensed under the Creative Commons CC0 1.0 Universal Public Domain Dedication (https://creativecommons.org/publicdomain/zero/1.0/). Images (b) and (g) used photos taken by Dr. Hans-Günter Wagner, which are licensed under a Creative Commons Attribution-ShareAlike 2.0 License (https://creativecommons.org/licenses/by-sa/2.0/). Images (c) and (h) used photos taken by Amante Darmanin, which are licensed under a Creative Commons Attribution 2.0 License (https://creativecommons.org/licenses/by/2.0/). Image (d) used a photo taken by Jesse Pluim (BLM), which is marked as being in the Public Domain using the Public Domain Mark 1.0 (https://creativecommons.org/publicdomain/mark/1.0/). Image (e) used a photo taken by Laurent Houmeau, and image (f) used a photo taken by Sergio Niebla; both are licensed under a Creative Commons Attribution-ShareAlike 2.0 License (https://creativecommons.org/licenses/by-sa/2.0/).

We scored pollination syndromes according to different pollinators divided into: ants, bats, bees, birds, moths, and unspecified insects (Figure 4). This new classification of pollination syndromes substantially improves the binary scoring system developed by [10,51], which included only two categories (bee pollinated (ancestral) and derived pollinators). Thus, we here provide pollination data at higher resolution than previously available. There are several examples of the limitations of using a binary classification system. Some species had derived pollinators, but also had “other insects” as recorded pollinators, some of which are bees. In these cases, species were incompletely scored as solely having ‘derived pollinators’. We validated traits through cross-checking with different sources wherever possible, and continuous traits were visually checked for outliers. Finally, we captured data on chromosome counts from the Chromosome Count Database [103], while retaining all recorded counts (Figure 5). This allowed us to correct the previous median count for several species that had a median of zero, despite having non-zero chromosome counts.

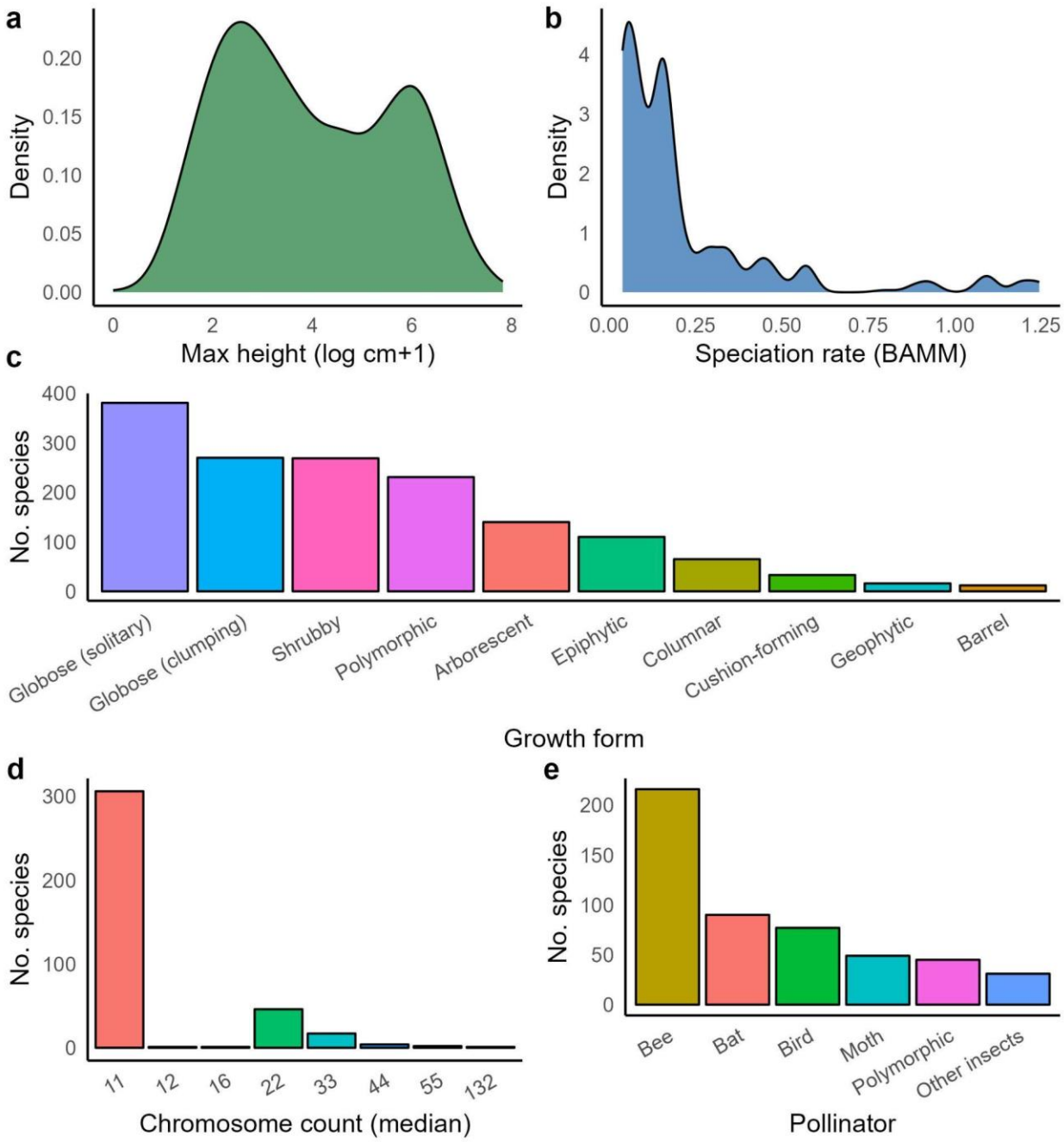


**Figure 4: The distribution of plant height, speciation rate, growth form, chromosome count, and pollinator.** (a) Cactus size (height or length) is bimodal, which is shaped by underlying growth form variation [10]. (b) Tip speciation rate is mostly relatively slow, with fewer relatively faster species. As a whole, cacti exhibit some of the fastest speciation rates in all plants [4]. (c) The most common growth form is globose solitary, followed by globose caespitose and shrubby. The least common growth forms are cushion forming, geophytic and barrel. 231 species are polymorphic, which mostly consists of globose solitary and globose caespitose (160 species), shrubby and

arborescent (52 species), and shrubby and columnar (11 species). (d) Most species have the base number of 11 chromosomes, and most chromosome count variation is generated by polyploidy [74], reflected in the distribution (e.g. 22, 33, 44, 55). (e) Most species are pollinated by derived pollinators (bats, birds, moths) rather than the ancestral pollinators (bees) [51].

**Distribution data**

Previous distribution range datasets used data downloaded and cleaned from the Global Biodiversity Information Facility (GBIF). However, we have greatly improved species and distributional sampling with a new strategy (Figure 5). For CactEcoDB, we have primarily used expert range polygons available in the IUCN Red List of Threatened Species [69,104], with some occurrences newly-acquired and curated from GBIF to maximise species sampling, following [105]. This approach increases species sampling by 54 species (~6%) and provides more accurate estimates of geographic range sizes, which allows us to fully characterise the environments of the entire ranges of species. Contrary to previous datasets, our new data reduce geographic bias, especially for species with poor GBIF records. To compile the distributional data, we downloaded IUCN ranges and manually checked them against the independent data sources available in POWO [88], by which we identified that a few IUCN ranges had sampled introduced populations. We corrected these range maps accordingly and then we sampled coordinates randomly from within each range in relationship to the range size, following [105]. Specifically, we used a binning function to assign sample sizes based on range area (in km²): species with ranges <10 km² were sampled 20 times; 10–100 km², 50 times; 100–1,000 km², 150 times; 1,000–10,000 km², 300 times; 10,000–100,000 km², 500 times; 100,000–1,000,000 km², 700 times; 1,000,000–10,000,000 km², 900 times; and >10,000,000 km², 1,200 times. This sampling approach ensured adequate environmental representation for both narrow and wide-ranging species, while avoiding excessive data volume or downstream computational demands, particularly for the most widespread taxa. For species not covered by IUCN, we curated the most up-to-date GBIF occurrences following [106,107]. We queried all records under “Cactaceae”, considering only “Preserved Specimens”. We removed species present in the IUCN range data, corrected species taxonomy to match POWO, and removed suspect records based on several overlapping criteria. We defined suspect occurrence records as those matching at least one of the following criteria: i) records with missing or badly formatted coordinates; ii) records associated with biodiversity institutions; iii) records with equal or zero latitude and longitude; iv) records within a 10km radius of a country’s capital; v) records within 5km of a country’s political centroid; vi) records in the sea [108]. We also remove occurrence records that were outside the range described by POWO for each species [88].

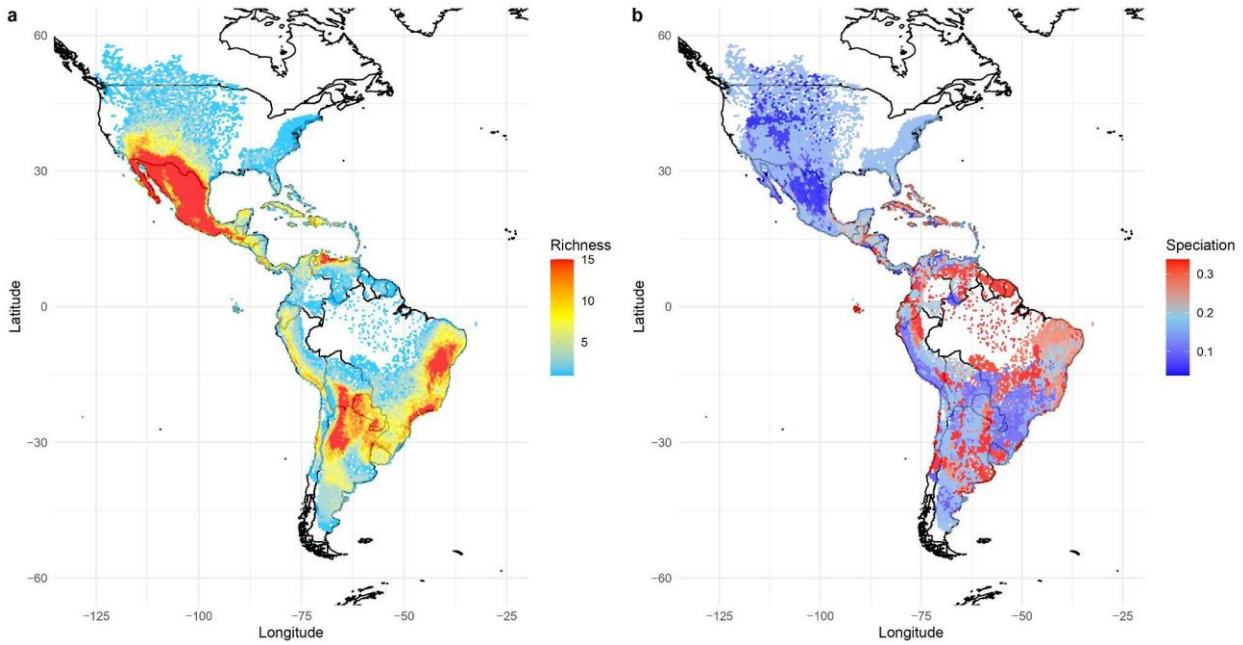


**Figure 5: The spatial distribution of cactus richness and tip speciation rates.** (a) Species richness is plotted per grid cell (~50*50 miles), and shows Mexican, Eastern Brazilian and Andean richness hotspots. (b) Median tip speciation rate is plotted per grid cell, showing a relative decoupling between areas of highest species richness and high speciation rate. This figure was presented in the supplementary materials of [10], but with much lower resolution and different underlying spatial data.

**Environmental data and verifying against GBIF data**

For each of the newly generated geographic coordinates from IUCN ranges or GBIF data, we extracted the commonly-used 19 bioclimatic variables from CHELSA [109], along with aridity index, potential evapotranspiration, and net primary productivity. We extracted elevation and six measures of topographic heterogeneity from [110], including slope, roughness, standard deviation of slope, and profile curvature. Soil variables (sand content, water availability, and texture) were obtained from SoilGrids [111]. We retained the median value per species, since they are less likely to be influenced by outliers. We confirmed that our new approach retrieves valid environmental data compared to previous GBIF-occurrence only approaches , by estimating linear regressions and Pearson correlation coefficients between the variables estimated by the new procedure and those from [10]. We found that comparisons between all 31 environmental variables are significant at $p < 0.05$, with a median slope and correlation coefficient of 0.89 and 0.88, respectively. Furthermore, 25 comparisons recovered a slope and correlation coefficient of > 0.75. Our new approach also retrieved valid data for soil texture, which is discrete and has strong agreement (72.60%) with data retrieved from GBIF occurrences [10].

By sampling multiple coordinates for each species across their full geographic range, CactEcoDB captures intraspecific environmental variation. This is especially valuable when analysing trait–environment relationships, because it accounts for environmental heterogeneity within species ranges, instead of relying on a point estimate. Incorporating intraspecific variance can lead to more realistic ecological and evolutionary inferences [112–115].

**Data Records**

All data files described below are available in the GitHub repository (https://github.com/jamie-thompson/CactEcoDB) and will be archived in an Open Access repository (e.g., Zenodo or Figshare) upon publication. The dataset comprises cleaned species-level trait, spatial, environmental, phylogenetic, and diversification rate data for up to 1,063 species of Cactaceae (Figures 1, 2 and 5). Each file is formatted for ease of integration into ecological and evolutionary analyses.

**1. CactEcoDB_v1.csv**

Format: Spreadsheet (.csv)

Description:

This file contains species-level data for up to 1,063 species of Cactaceae, integrating trait, phylogenetic, spatial, and environmental information. Specifically, it includes:

- Mean tip-specific speciation rates estimated using BAMM (lambda)
- Morphological and ecological trait data, including maximum plant height (cm), growth form (binary and refined), pollination syndrome, and epiphytism
- Summary statistics (typically medians) of environmental variables based on georeferenced occurrence data
- Geographic range size estimates ($km^2$)

Fields:

- Diversification
    - lambda: Mean species-specific speciation rate estimated using BAMM [98]

- Morphological and ecological traits
    - growth_form_binary: Coarse growth form classification (e.g. columnar vs. globose group), following [10,51]
    - growth_form_detailed: Detailed growth form description based on literature
    - max_size_height_or_length: Maximum reported plant height or stem length (cm), building on [10]
    - pollinator_binary: Pollination syndrome grouped into ancestral (bee) vs. derived (bat, bird, moth), according to [10,51]

    - pollinator_detailed: Detailed pollination syndromes (e.g. bee, bat, bird, ant)
    - chromosome_count_median: Median chromosome count (n)
    - chromosome_count_distribution: Summary information or distribution notes on chromosome counts (n) sourced from the Chromosome Counts Database [103]
    - range_size: Geographic range size estimates ($km^2$)

- Environmental variables (CHELSA bioclimatic variables [109], SoilGrids [111], and Global Multi-resolution Terrain Elevation Data (GMTED) [110,116])
    - Bio1–bio19: Standard bioclimatic variables, including temperature and precipitation metrics, extracted from CHELSA [109]. Bio1 to bio11 are temperature-related variables, and bio12-bio19 are precipitation-related. For example:
        - bio1: Annual mean temperature (°C)
        - bio12: Annual precipitation (kg $m^{-2}$ $year^{-1}$)
        - (See CHELSA documentation for full descriptions, especially with regards to scales and offsets used in some variables [109])
    - Aridity and productivity
        - aridity_index: Aridity index (mean annual precipitation/potential evapotranspiration)
        - pet: Potential evapotranspiration (kg $m^{-2}$ $year^{-1}$)
        - npp: Net primary productivity (g C $m^{-2}$ $yr^{-1}$)
    - Elevation and topography extracted from GMTED [110,116]
        - elevation: Median elevation across occurrence records (m)
        - slope: Median slope of the terrain (°)
        - sd_slope: Median standard deviation (sd) of slope
        - sd_elev: Median standard deviation of elevation
        - roughness: Median terrain roughness (largest inter-cell difference of a focal cell and its 8 surrounding cells, m)
        - prof_curve: Median standard deviation of profile curvature (sd of radians per meter)
        - tang_curve: Median standard deviation of tangential curvature (sd of radians per meter)
    - Soil Variables extracted from SoilGrids [111]
        - soil_sand: Median proportion of sand particles (> 0.05/0.063 mm) in the fine earth fraction
        - soil_water: Median derived available soil water capacity (volumetric fraction)
        - texture: Soil texture class, classified into twelve categories (e.g. sand, silt, clay, loam, silty clay)

Notes

- Speciation rates (lambda) represent mean species-specific estimates from the entire posterior sample estimated by BAMM. The full BAMM posterior event data is available separately in CactEcoDB_BAMM_eventdata.txt, and can be analyzed using the R package BAMMtools [99].
- Plant height values are not statistical species means from population samples, but are derived from maximum reported sizes in taxonomic or ecological literature, as is common across plant research generally [10,101,117–120].
- Pollinator data was rarely specific to individual species. Therefore, we classified them into major taxonomic groups described in [10,51].

**2. CactEcoDB_coords_envs.csv**
Format: csv
Description: Curated distribution coordinates with extracted environmental values for every coordinate
Fields: species, decimal latitude, decimal longitude, environmental variables (e.g. bio1-bio19, elevation, prof_curve, texture)
Notes: Most of these coordinates are randomly sampled from expert IUCN ranges, some of them are curated GBIF occurrences.

**3. CactEcoDB_phylogeny.nwk**
Format: Newick
Description: Time-calibrated phylogeny for 1,063 Cactaceae species, inferred using a supermatrix approach and Penalized Likelihood dating.
Notes: Outgroups are pruned and tips are matched to species names in the trait and spatial datasets to facilitate research.

**4. CactEcoDB_BAMM_eventdata.txt**
Format: txt
Description: The posterior sample of rate shift configurations (eventdata) estimated by BAMM.
Fields: specific to BAMM [98], use the R package BAMMtools [99] to access and analyse this.

**5. CactEcoDB_variable_descriptions.txt**
Format: txt
Description: Definitions, units, and data sources for all variables in the dataset.
Notes: Intended to support reproducibility and proper variable interpretation.

## Technical validation

Trait data were checked for plausibility and consistency. Values that were extreme or contradictory, e.g. height outliers by growth forms, were flagged and verified against original sources when available. Generally, only minimum and maximum plant height values were available. We retained maximum values, as is standard in plant research [101,117]. Traits (e.g. pollination syndrome, growth form) classification were updated and cross-validated using multiple independent sources, where possible, and most followed consistent definitions derived from previous macroevolutionary studies [51,60]. Environmental variables were extracted from high-resolution raster datasets (e.g. CHELSA [109], SoilGrids [111] at each coordinate, and species-level summaries were calculated as medians to minimise the influence of outliers following [10]. Critically, we statistically compared values of extracted spatial data between previous methods that employed GBIF-derived data [10] and our current method, and found very strong agreement, demonstrating that our new approach of sampling from expert IUCN ranges is appropriate.

The phylogenetic tree was constructed using 18 multiple sequence alignments of commonly sequenced nucleotide loci available in GenBank. They were manually inspected for alignment quality and coverage. Poorly aligned regions were trimmed [93], and Maximum Likelihood tree reconstruction was conducted with partitioned models and bootstrapping [95,96]. Time calibration used cross-validated Penalized Likelihood [97] with secondary calibrations based on estimated angiosperm divergence times [5], in the absence of a useful fossil record of Cactaceae [55].

Diversification rate estimates generated with BAMM used genus-level sampling fractions to correct for imbalanced taxonomic sampling, and a conservative prior of one single rate shift [98]. Convergence was assessed using standard MCMC diagnostics, and effective sample sizes exceeded recommended thresholds. We do not recommend using alternative metrics which do not account for incomplete sampling, such as the DR statistic [35,121].

## Usage Notes

If CactEcoDB is used in publications, both this present paper and that of Thompson et al. [10] should be cited appropriately. Furthermore, the Chromosome Count Database [103] must be cited if chromosome counts are used.

As is typical of large-scale biodiversity datasets, CactEcoDB reflects uneven data availability across species and data types [47,122]. Trait coverage is extensive across the family, but completeness varies by trait and specific group. For plant height, values were derived from descriptive sources that reported minimum and maximum height. In

these cases, maximum values were retained to reduce bias from juvenile, incomplete, or diseased specimens. However, these records often lack associated metadata (e.g., number of individuals measured, habitat context), and users should treat them as approximations of mature phenotypes rather than statistically derived species means. The same can be true for chromosome count, which can vary intraspecifically.

Phylogenetic coverage is broad but incomplete. The time-calibrated phylogeny includes 1,063 species and represents all major lineages of Cactaceae, but remains sparse for some poorly sequenced clades. To account for this, sampling fractions were used to correct for incomplete sampling, where possible, in the diversification analyses. We provide only one topology, which does not account for temporal and topological uncertainty. However, we note that in the absence of useful fossils [55], the timings of cactus evolution will likely always remain uncertain. Additionally, the relationships and internal timings are broadly in agreement with other reconstructions, which seem to be converging on a crown age of ~27-37 Mya [5,51,55]. We hypothesise that the divergence estimates in the phylogeny are relatively correct in their branch lengths, albeit not absolutely correct to geological time.

In summary, CactEcoDB offers a comprehensive resource for large-scale ecological and evolutionary analyses, but as with any aggregated dataset, its contents should be interpreted in light of the inherent variability in data completeness and precision. Where greater taxonomic or spatial resolution is required, users may find it helpful to consult original sources or species-level references. As stated before, we welcome the submission of new data to update and augment CactEcoDB. We intend to update the data regularly and reconstruct new phylogenies.

**Code availability**
Does not apply.

**Funding**
This work was funded by a Roger and Sue Whorrod OBE PhD studentship awarded to J.B.T and N.K.P, and a British Cactus and Succulent Society research grant to J.B.T and C.M.

**Acknowledgments**
We acknowledge Dr Marilyn Vásquez-Cruz and Dr Tania Hernández-Hernández for contributions to generating the spatial dataset that was used to compare our newly generated data. We acknowledge Michiel Pillet for discussion on the complexities of cactus growth forms.

**References:**


1. Pérez-Escobar, O. A. *et al.* The origin and speciation of orchids. *New Phytol* **242**, 700–716 (2024).
2. Sauquet, H. & Magallón, S. Key questions and challenges in angiosperm macroevolution. *New Phytol* **219**, 1170–1187 (2018).
3. Panero, J. L. & Crozier, B. S. Macroevolutionary dynamics in the early diversification of Asteraceae. *Mol Phylogenet Evol* **99**, 116–132 (2016).
4. Magallón, S., Sánchez-Reyes, L. L. & Gómez-Acevedo, S. L. Thirty clues to the exceptional diversification of flowering plants. *Ann Bot* **123**, 491–503 (2019).
5. Ramírez-Barahona, S., Sauquet, H. & Magallón, S. The delayed and geographically heterogeneous diversification of flowering plant families. *Nat Ecol Evol* **4**, 1232–1238 (2020).
6. Cabral, A. *et al.* Seed-dispersing vertebrates and the abiotic environment shape functional diversity of the pantropical Annonaceae. *New Phytol* **246**, 2263–2279 (2025).
7. Onstein, R. E. *et al.* Frugivory-related traits promote speciation of tropical palms. *Nat Ecol Evol* **1**, 1903–1911 (2017).
8. Onstein, R.E. et al. 2014. Diversification rate shifts in the Cape Floristic Region: The right traits in the right place at the right time. *Perspect. Pl. Ecol. Evol. Syst.* **16**(6), 331-340 (2014).
9. Givnish, T. J. *et al.* Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc Biol Sci* **282**, (2015).
10. Thompson, J. B., Hernández-Hernández, T., Keeling, G., Vásquez-Cruz, M. &

Priest, N. K. Identifying the multiple drivers of cactus diversification. *Nat Commun* **15**, 7282 (2024).

11. Thompson, J. B., Davis, K. E., Dodd, H. O., Wills, M. A. & Priest, N. K. Speciation across the Earth driven by global cooling in terrestrial orchids. *Proc Natl Acad Sci U S A* **120**, e2102408120 (2023).
12. Thompson, J., Ramírez-Barahona, S., Priest, N. & Hernández-Hernández, T. Did succulents diversify in response to aridity? Evolutionary analyses of major succulent lineages around the world. *bioRxiv* 2023.05.23.541957 (2023) doi:10.1101/2023.05.23.541957.
13. Kattge, J. *et al.* TRY plant trait database - enhanced coverage and open access. *Glob Chang Biol* **26**, 119–188 (2020).
14. Tavşanoğlu, Ç. & Pausas, J. G. A functional trait database for Mediterranean Basin plants. *Sci Data* **5**, 180135 (2018).
15. Falster, D. *et al.* AusTraits, a curated plant trait database for the Australian flora. *Sci Data* **8**, 254 (2021).
16. Wang, H. *et al.* The China plant trait database version 2. *Sci Data* **9**, 769 (2022).
17. Weigelt, P., König, C. & Kreft, H. GIFT – A Global Inventory of Floras and Traits for macroecology and biogeography. *J. Biogeogr.* **47**, 16–43 (2020).
18. Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol* **17**, e3000494 (2019).
19. Payne, A. R. D., Mannion, P. D., Lloyd, G. T. & Davis, K. E. Decoupling speciation and extinction reveals both abiotic and biotic drivers shaped 250 million years of

diversity in crocodile-line archosaurs. *Nat Ecol Evol* **8**, 121–132 (2024).

20. Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *Am J Bot* **105**, 302–314 (2018).
21. Portik, D. M., Streicher, J. W. & Wiens, J. J. Frog phylogeny: A time-calibrated, species-level tree based on hundreds of loci and 5,242 species. *Mol Phylogenet Evol* **188**, 107907 (2023).
22. Waller, J. T., Willink, B., Tschol, M. & Svensson, E. I. The odonate phenotypic database, a new open data resource for comparative studies of an old insect order. *Sci Data* **6**, 316 (2019).
23. Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C. & Costa, G. C. AmphiBIO, a global database for amphibian ecological traits. *Sci Data* **4**, 1–7 (2017).
24. Pottier, P. *et al.* A comprehensive database of amphibian heat tolerance. *Sci Data* **9**, 600 (2022).
25. Wang, J. *et al.* ChelonianTraits: a comprehensive trait database of global turtles and tortoises. *Sci Data* **12**, 1–7 (2025).
26. Wei, P. *et al.* CaecilianTraits, an individual level trait database of Caecilians worldwide. *Sci Data* **12**, 1–7 (2025).
27. Chaudhary, V. B. *et al.* TraitAM, a global spore trait database for arbuscular mycorrhizal fungi. *Sci Data* **12**, 1–10 (2025).
28. Gómez-Gras, D. *et al.* The Octocoral Trait Database: a global database of trait information for octocoral species. *Sci Data* **12**, 1–32 (2025).
29. Báez, S. *et al.* FunAndes – A functional trait database of Andean plants. *Sci Data*

**9**, 1–9 (2022).

30. Herberstein, M. E. *et al.* AnimalTraits - a curated animal trait database for body mass, metabolic rate and brain size. *Sci Data* **9**, 1–11 (2022).
31. Kissling, W. D. *et al.* PalmTraits 1.0, a species-level functional trait database of palms worldwide. *Sci Data* **6**, 1–13 (2019).
32. Jetz, W. & Pyron, R. A. The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nat Ecol Evol* **2**, 850–858 (2018).
33. Tonini, J.F.R. *et al.* Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biol. Conserv.* **204**, 23–31 (2016).
34. Stein, R. W. *et al.* Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nat Ecol Evol* **2**, 288–298 (2018).
35. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
36. Givnish, T. J. *et al.* Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Mol Phylogenet Evol* **71**, 55–78 (2014).
37. Siqueira, A. C., Morais, R. A., Bellwood, D. R. & Cowman, P. F. Trophic innovations fuel reef fish diversification. *Nat Commun* **11**, 2669 (2020).
38. Bouchenak-Khelladi, Y., Onstein, R. E., Xing, Y., Schwery, O. & Peter Linder, H. On the complexity of triggering evolutionary radiations. *New Phytol* **207**, 313–326 (2015).
39. Padilla-Morales, B. *et al.* Fast species diversification among dragonflies (Anisoptera: Odonata: Insecta) inhabiting lentic environments regardless of wing

pigmentation. *Ecological Entomology* **47**, 314–322 (2022).

40. Hembry, D.H. and Weber, M.G. Ecological interactions and macroevolution: a new field with old roots. *Annual Review of Ecology, Evolution, and Systematics* ***51***(1), 215-243 (2020).
41. Swiston, S. K. & Landis, M. J. Testing Relationships between Multiple Regional Features and Biogeographic Processes of Speciation, Extinction, and Dispersal. *Syst Biol* **74**, 282–300 (2025).
42. Reginato, M. *et al.* MelastomaTRAITs 1.0: A database of functional traits in Melastomataceae, a large pantropical angiosperm family. *Ecology* **105**, e4308 (2024).
43. Etard, A., Morrill, S. & Newbold, T. Global gaps in trait data for terrestrial vertebrates. *Glob. Ecol. Biogeogr.* **29**, 2143–2158 (2020).
44. Grenié, M. *et al.* Gaps in global alien plant trait data and how to fill them. *EcoEvoRxiv* (2024).
45. Iversen, C.M. and McCormack, M.L. Filling gaps in our understanding of belowground plant traits across the world: an introduction to a Virtual Issue. *New Phytol*, ***231***(6), 2097-2103 (2021).
46. Cutts, V. *et al.* Scientific floras can be reliable sources for some trait data in a system with poor coverage in global trait databases. *Journal of Vegetation Science* **32**, e12996 (2021).
47. Díaz, S. *et al.* The global spectrum of plant form and function: enhanced species-level trait dataset. *Sci Data* **9**, 1–18 (2022).
48. Cordeiro, A. L. *et al.* TropiRoot 1.0: Database of tropical root characteristics across

environments. *Ecology* **106**, e70074 (2025).

49. Guerrero-Ramírez, N. R. *et al.* Global root traits (GRooT) database. *Glob. Ecol. Biogeogr.* **30**, 25–37 (2021).
50. Cornwell, W.K., Pearse, W.D., Dalrymple, R.L. and Zanne, A.E. What we (don't) know about global plant diversity. *Ecography*, ***42***(11), 1819-1831 (2019).
51. Hernández-Hernández, T., Brown, J. W., Schlumpberger, B. O., Eguiarte, L. E. & Magallón, S. Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytol* **202**, 1382–1397 (2014).
52. Guerrero, P. C., Majure, L. C., Cornejo-Romero, A. & Hernández-Hernández, T. Phylogenetic Relationships and Evolutionary Trends in the Cactus Family. *J Hered* **110**, 4–21 (2019).
53. Hultine, K. R. *et al.* Global change impacts on cacti (Cactaceae): current threats, challenges and conservation solutions. *Ann Bot* **132**, 671–683 (2023).
54. Gibson, A. C. & Nobel, P. S. *The Cactus Primer*. (Harvard University Press, 1986).
55. Arakaki, M. *et al.* Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad Sci U S A* **108**, 8379–8384 (2011).
56. Mandujano, M., Carrillo-Angeles, I., Martínez-Peralta, C. & Golubov, J. Reproductive Biology of Cactaceae. *Desert Plants* 197–230 (2010).
57. Godinez-Alvarez, H., Valverde, T. and Ortega-Baes, P. Demographic trends in the Cactaceae. *Bot. Rev.*, ***69***(2), 173-201 (2003).
58. Barthlott, W. & Hunt, D. R. Cactaceae. in *Flowering Plants · Dicotyledons* 161–197 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1993).

59. Calvente, A., Zappi, D.C., Forest, F. and Lohmann, L.G. Molecular phylogeny, evolution, and biogeography of South American epiphytic cacti. *Int. J. Plant Sci.*, ***172***(7), 902-914 (2011).

60. Hernández-Hernández, T. *et al.* Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *Am J Bot* **98**, 44–61 (2011).

61. Ventura-Aguilar, R. I., Bosquez-Molina, E., Bautista-Baños, S. & Rivera-Cabrera, F. Cactus stem (*Opuntia ficus-indica* Mill): anatomy, physiology and chemical composition with emphasis on its biofunctional properties. *J Sci Food Agric* **97**, 5065–5073 (2017).

62. Edwards, E. J. & Donoghue, M. J. *Pereskia* and the origin of the cactus life-form. *Am Nat* **167**, 777–793 (2006).

63. Amaral, D. T. *et al.* Tracking the xeric biomes of South America: The spatiotemporal diversification of Mandacaru cactus. *J. Biogeogr.* **48**, 3085–3103 (2021).

64. Majure, L. C. *et al.* Pleistocene aridification underlies the evolutionary history of the Caribbean endemic, insular, giant *Consolea* (Opuntioideae). *Am J Bot* **108**, 200–215 (2021).

65. Majure, L. C., Barrios, D., Díaz, E., Bacci, L. F. & Piñeyro, Y. E. Phylogenomics of the Caribbean melocacti: Cryptic species and multiple invasions. *TAXON* **71**, 993–1012 (2022).

66. Fleming, T. H. & Valiente-Banuet, A. *Columnar Cacti and Their Mutualists: Evolution, Ecology, and Conservation*. (University of Arizona Press, 2002).

67. Novoa, A., Le Roux, J. J., Robertson, M. P., Wilson, J. R. U. & Richardson, D. M.

Introduced and invasive cactus species: a global review. *AoB Plants* **7**, (2014).

68. Aranda-Pineda, J. A., Valverde, P. L., Búrquez, A., Golubov, J. & Mandujano, M. C. Effect of brousse tigrée on the dynamics of nurse–protégé interactions of a cactus in the Chihuahuan Desert. *Plant Species Biol.* **36**, 450–462 (2021).
69. Goettsch, B. *et al.* High proportion of cactus species threatened with extinction. *Nat Plants* **1**, 15142 (2015).
70. Pillet, M. *et al.* Elevated extinction risk of cacti under climate change. *Nat Plants* **8**, 366–372 (2022).
71. Goettsch, B., Durán, A. P. & Gaston, K. J. Global gap analysis of cactus species and priority sites for their conservation. *Conserv Biol* **33**, 369–376 (2019).
72. Spatial patterns of evolutionary diversity in Cactaceae show low ecological representation within protected areas. *Biol. Conserv.* **273**, 109677 (2022).
73. Hultine, K. R. *et al.* The Role of Botanical Gardens in the Conservation of Cactaceae. *BioScience* **66**, 1057–1065 (2016).
74. Castro, J. P. *et al.* Karyotype characterization and evolution of chromosome number in Cactaceae with special emphasis on subfamily Cactoideae. *Acta Bot. Bras.* **34**, 135–148 (2019).
75. Yetman, D. *The Great Cacti: Ethnobotany & Biogeography*. (University of Arizona Press, 2007).
76. Pérez-Negrón, E., Dávila, P. & Casas, A. Use of columnar cacti in the Tehuacán Valley, Mexico: perspectives for sustainable management of non-timber forest products. *J Ethnobiol Ethnomed* **10**, 79 (2014).
77. Ortiz-Martínez, L. E., Manzanero-Medina, G. I., Golubov, J., Vásquez-Dávila, M. A.

& Mandujano, M. C. Nopal de Monte: Cacti Named and Used by a Mixtec Community in Mountainous Oaxaca. *Ethnobotany of the Mountain Regions of Mexico* 899–908 (2023).

78. Patrick Griffith, M. The origins of an important cactus crop, *Opuntia ficus-indica* (Cactaceae): new molecular evidence. *Am. J. Bot.* **91**, 1915–1921 (2004).
79. Kiesling, R., 1998. Origin, domestication and distribution of *Opuntia ficus-indica*. *Journal of the Professional Association for Cactus Development*, **3**, 50-59 (1998).
80. Food and Agriculture Organization of the United Nations. *Cactus (Opuntia Spp.) as Forage*. (Food & Agriculture Org., 2001).
81. Food and Agriculture Organization of the United Nations. *Crop Ecology, Cultivation and Uses of Cactus Pear*. (Food & Agriculture Org., 2018).
82. Donkin, R.A. Spanish red: an ethnogeographical study of cochineal and the *Opuntia* cactus. *Trans. Am. Philos. Soc*, ***67***(5), 1-84 (1977).
83. Margulies, J. D., Moorman, F. R., Goettsch, B., Axmacher, J. C. & Hinsley, A. Prevalence and perspectives of illegal trade in cacti and succulent plants in the collector community. *Conserv Biol* **37**, e14030 (2023).
84. Novoa, A., Le Roux, J. J., Richardson, D. M. & Wilson, J. R. U. Level of environmental threat posed by horticultural trade in Cactaceae. *Conserv Biol* **31**, 1066–1075 (2017).
85. Sajeva, M. *et al. CITES and Cacti: A User's Guide*. (Royal Botanic Gardens Kew, 2012).
86. Franco, F. F. *et al.* Evolutionary Genetics of Cacti: Research Biases, Advances and Prospects. *Genes,* **13**, (2022).

87. Yesson, C. *et al.* How Global Is the Global Biodiversity Information Facility? *PLOS ONE* **2**, e1124 (2007).

88. Brown, M. J. M. *et al.* rWCVP: a companion R package for the World Checklist of Vascular Plants. *New Phytol* **240**, 1355–1365 (2023).

89. Drori, M. *et al.* OneTwoTree: An online tool for phylogeny reconstruction. *Mol Ecol Resour* **18**, 1492–1499 (2018).

90. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).

91. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**, 221–224 (2010).

92. Katoh, K. & Frith, M. C. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinform,* **28**, 3144–3146 (2012).

93. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinform* **25**, 1972–1973 (2009).

94. Borowiec, M. L. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4**, e1660 (2016).

95. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinform*, **30**, 1312–1313 (2014).

96. Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**, 758–771 (2008).

97. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized

likelihood for large phylogenies. *Bioinform,* **28**, 2689–2690 (2012).

98. Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One* **9**, e89543 (2014).
99. Rabosky, D. L. *et al.* BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* **5**, 701–707 (2014).
100. Anderson, E. F. *The Cactus Family*. (Timber Press (OR), 2001).
101. Moles, A. T. *et al.* Global patterns in plant height. *J. Ecol.* **97**, 923–932 (2009).
102. Zotz, G., Weigelt, P., Kessler, M., Kreft, H. & Taylor, A. EpiList 1.0: a global checklist of vascular epiphytes. *Ecology* **102**, e03326 (2021).
103. Rice, A. *et al.* The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers. *New Phytol* **206**, 19–26 (2015).
104. IUCN, *The IUCN Red List of Threatened Species*. Version 2022-2. https://www.iucnredlist.org. Accessed on [10/02/2025]. (2022).
105. Avaria-Llautureo, J. *et al.* The radiation and geographic expansion of primates through diverse climates. *EcoEvoRxiv* (2024).
106. Carta, A., Peruzzi, L. & Ramírez-Barahona, S. A global phylogenetic regionalization of vascular plants reveals a deep split between Gondwanan and Laurasian biotas. *New Phytol* **233**, 1494–1504 (2022).
107. Ramírez-Barahona, S. *et al.* Upslope plant species shifts in Mesoamerican cloud forests driven by climate and land use change. *Science* **387**, 1058–1063 (2025).
108. Zizka, A. *et al.* CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol. Evol.***10**, 744–751 (2019).
109. Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface

areas. *Sci Data* **4**, 170122 (2017).

110. Amatulli, G. *et al.* A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Sci Data* **5**, 180040 (2018).

111. Poggio, L. *et al.* SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL* **7**, 217–240 (2021).

112. Fisher-Reid, M. C. & Wiens, J. J. Is geographic variation within species related to macroevolutionary patterns between species? *J Evol Biol* **28**, 1502–1515 (2015).

113. Avaria-Llautureo, J., Rivadeneira, M. M., Venditti, C. & Luna-Jorquera, G. Seabird range contraction and dispersal under climate change. *EcoEvoRxiv* (2025).

114. Mitchell, N., Carlson, J. E. & Holsinger, K. E. Correlated evolution between climate and suites of traits along a fast-slow continuum in the radiation of *Protea*. *Ecol Evol* **8**, 1853–1866 (2018).

115. Nakagawa, S. & De Villemereuil, P. A General Method for Simultaneously Accounting for Phylogenetic and Species Sampling Uncertainty via Rubin's Rules in Comparative Analysis. *Syst Biol* **68**, 632–641 (2019).

116. Danielson, J. J. & Gesch, D. B. Global multi-resolution terrain elevation data 2010. *US Department of the Interior and US Geological Survey* (2011), Report 2011–1073, Page 34.

117. Boucher, F. C., Quatela, A.-S., Ellis, A. G. & Anthony Verboom, G. Diversification rate vs. diversification density: Decoupled consequences of plant height for diversification of Alooideae in time and space. *PLOS ONE* **15**, e0233597 (2020).

118. Mao, L. *et al.* The geographic and climatic distribution of plant height diversity for 19,000 angiosperms in China. *Biodivers. Conserv.*, **29**, 487–502 (2019).

119. Liu, H. *et al.* Hydraulic traits are coordinated with maximum plant height at the global scale. *Sci Adv* **5**, eaav1332 (2019).

120. Mao, L., Chen, S., Zhang, J. & Zhou, G. Altitudinal patterns of maximum plant height on the Tibetan Plateau. *J Plant Ecol* **11**, 85–91 (2016).

121. Title, P. O. & Rabosky, D. L. Tip rates, phylogenies and diversification: What are we estimating, and how good are the estimates? *Methods Ecol. Evo.,* **10**, 821–834 (2019).

122. Díaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167–171 (2015).