Traits related to pollination and mating result in diverse reproductive strategies in angiosperms

Andrew J. Helmstetter[a*], Marcos Méndez[b], Jürg Schönenberger[c], Concetta Burgarella[d], Bruce Anderson[e], Maria von Balthazar[c], Sylvain Billiard[f], Hugo de Boer[g], Johanne Cros[h], Pierre-Alexandre Delecroix[i], Mathilde Dufay[j], John Pannell[k],Dessislava Savova Bianchi[k], Daniel Schoen[l], Mario Vallejo-Marin[m], Rosana Zenil-Ferguson[n], Hervé Sauquet[op], Sylvain Glémin[im*], Jos Käfer[q*]

[a]Fondation pour la Recherche sur la Biodiversité - Centre for the Synthesis and Analysis of Biodiversity, 34000 Montpellier, France
[b]Area of Biodiversity and Conservation, Universidad Rey Juan Carlos, 28933 Móstoles (Madrid), Spain
[c]Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, 1030 Vienna, Austria
[d]Department of Organismal Biology, University of Uppsala, Uppsala, 75236, Sweden
[e]Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa
[f]University Lille, CNRS, UMR 8198 – Evo-Eco-Paleo, 59000 Lille, France
[g]Natural History Museum, University of Oslo, 0318 Oslo, Norway
[h]Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR 5558, Villeurbanne, France
[i]CNRS, Ecosystèmes Biodiversité Evolution (Université de Rennes), 35000 Rennes, France
[j]CEFE, University Montpellier, CNRS, University Paul Valéry Montpellier 3, Ephe, IRD, Montpellier, France
[k]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland
[l]Department of Biology, McGill University, Montreal, QC, Canada
[m]Department of Ecology and Genetics, Uppsala University, Uppsala, 752 36, Sweden
[n]Department of Biology, University of Kentucky, 101 T.H. Morgan Building, Lexington, KY 40506 USA
[o]National Herbarium of New South Wales, Royal Botanic Gardens and Domain Trust, Sydney, New South Wales, 2000, Australia
[p]Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia
[q]ISEM, Univ Montpellier, CNRS, IRD, Montpellier, France

[*] **Correspondence:** Andrew J. Helmstetter, FRB-CESAB Institut Bouisson Bertrand, 5, rue de l'École de médecine - 34000 Montpellier. Email: andrew.j.helmstetter@gmail.com
Sylvain Glémin, CNRS, Ecosystèmes Biodiversité Evolution (Université de Rennes), 35000 Rennes, France. Email: sylvain.glemin@univ-rennes1.fr
Jos Käfer, ISEM, Univ Montpellier, CNRS, IRD, Montpellier, France. Email: jos.kafer@cnrs.fr

Authorship statement: SG and JK initiated the project; BA, SB, CB, HdB, MD, SG, JK, MM, SPO, JP, DR, HS, DS, JS, MVM, RZF conceived the project; MvB, HS and JS provided data management; MvB, CB, JC, PAD, SG, JK, MM, DSB, HS and JS entered data; SG, AH, JK and HS performed analyses; AH visualized results; SG, AH, JK drafted the first version of the manuscript; all authors contributed to and approved of the final version.
Data accessibility statement:. To be provided

Abstract

Reproductive and floral traits present an enormous diversity in angiosperms (flowering plants) and are associated with differences in species diversity. These traits also interact in non-trivial ways but their interdependence in the influence of species' evolutionary success has only rarely been taken into account. Here we characterize the strategies plants have evolved to achieve pollination and reproduction using 21 floral and life-history traits from an original and representative set of 360 species sampled in 170 families across the angiosperms. We found that, while outcrossing rates *per se* are associated with a plant size / growth form axis, pollination-related traits such as flower gender and floral reward represent an almost equally important axis of variation. This correlation with pollination clearly sets unisexuality (monoecy and dioecy) apart as a separate outcrossing strategy that likely results from more ecological selective pressures than the avoidance of selfing and inbreeding alone. Species are not evenly distributed across trait space and we identified

three main reproductive strategies corresponding to combinations of traits that repeatedly evolved together (herbaceous bisexual, woody bisexual and woody unisexual species). We argue that pollination-related traits, which have largely been overlooked in studies of plant functional ecology, allow the integration of ecological and evolutionary timescales and provide a new perspective in the study of mating and sexual system evolution and ecosystem functioning.

Introduction

Angiosperms (flowering plants) are by far the most species-rich group of plants today. They also present an exceptional diversity of modes of reproduction, with variation both in sex distribution within and among individuals (sexual systems: from hermaphroditism to dioecy; Renner, 2014) and in mating patterns (mating systems: from strict outcrossing to predominant selfing; Igic and Kohn, 2006). However, this diversity is not evenly distributed and most species are hermaphroditic and outcrossing (Barrett, 2002).

The benefits of outcrossing are thought to be partly responsible for the differences in evolutionary success between species (Goldberg et al., 2010; Grossenbacher et al., 2015), but why are there so many ways to avoid selfing, and what drives their evolution? Genetic self-incompatibility occurs in many species and has been maintained over long evolutionary timescales despite frequent transitions to self-compatibility (Igic & Kohn, 2006; Goldberg et al., 2010). On the other hand, dioecy, i.e. the separation of female and male individuals that results in obligate outcrossing, arose frequently but is found in only a small minority of species (Renner, 2014). How dioecy evolves and whether it influences species' evolutionary success, is not clear (Goldberg et al., 2017; Käfer et al., 2017).

A possible reason for these observations is that traits do not evolve independently from one another, but rather concertedly. First, specific combinations of traits can occur through natural selection. This is the case for the joint evolution of lifespan and mating system that leads to either long-lived outcrossers or short-lived selfers (Scofield & Schultz, 2006; Lesaffre & Billiard, 2020), which is a well-known association in plants (Barrett et al., 1996). Alternatively, after a first trait has evolved, selection pressures could favour a second one. For instance, specialised pollinators that more efficiently deliver pollen would lead to higher outcrossing rates, but also put species at the risk of pollinator failure. Species could thus evolve other mechanisms to ensure pollination, as is probably the case with cleistogamy, the occurrence of closed and obligate self-pollinated flowers, which is often associated with bilaterally symmetric (zygomorphic) flowers adapted to specialized pollinators (Joly & Schoen, 2021).

Second, traits can affect speciation and extinction in a complex way and thus influence lineage selection. This process is thought to cause the association of dioecy with certain traits, such as wind pollination or fleshy fruits (Vamosi & Vamosi, 2004), leading to higher extinction rates when dioecy evolves in lineages that do not have these traits. Indeed, taken in isolation, most traits likely have contrasting effects on diversification (i.e., the difference of speciation and extinction; Helmstetter et al., 2022; Anderson et al., 2023). The net outcome thus depends on the specific combination of traits a species possesses, as well as its environment.

Although several correlates of mating and sexual systems have been well characterized, as illustrated above, they usually explain a rather low proportion of the observed diversity. This is mainly because such studies have generally examined correlations between only one kind of system (e.g., monoecy vs dioecy, outcrossing vs selfing) and one or a few life-history/ecological traits (Renner & Ricklefs, 1995; Barrett et al., 1996). An integrated analysis of mating and sexual systems and their associated traits is needed to account for interactions between traits.

Thanks to large-scale databases, plant ecologists have already characterized the main axes of

trait variation and identified trait combinations forming ”strategies” (Díaz et al., 2016; Salguero-Gómez et al., 2016). These traits are predicted to influence plant survival according to the levels of stress, competition and disturbance following seminal work by Grime (1974) and can be interpreted in terms of ecological processes, such as the colonization-exploitation (r-K) continuum, the leaf economic spectrum (Díaz et al., 2016), or the fast-slow continuum (Salguero-Gómez, 2017). Apart from vegetative traits (e.g. size, leaf area and mass), seed characteristics (e.g. seed mass and number) are the only traits related to reproduction.

Traits associated with pollination, mating and sexual systems, which play a role in the reproductive processes that happen before seed production, have largely been ignored in the description of plant ecological strategies (cf Munoz et al., 2016; E-Vojtkó et al., 2020). Pre-zygotic processes are nevertheless a key parameter of plants' life cycles and evolution. They are crucial for the reproductive success of a plant and they influence the way individuals exchange genetic material, and thus have many demographic and genetic consequences. On short timescales, pollination directly influences an individual's fitness through the number of seeds produced (e.g., depending on pollen limitation; Ashman et al., 2004). Second, pollination can influence seed quality (if affected by inbreeding depression; Crnokrak and Roff, 1999), and could thus affect plants' abilities to compete and cope with stress (Craig & Mertz, 1994; Cheptou et al., 2000; Petrone Mendoza et al., 2018). On longer timescales, pollination and reproduction affect genetic diversity and the adaptive capacity, and thus most likely play an important role in long-term population and species survival (Burgarella & Glémin, 2017; Anderson et al., 2023). Furthermore, through the cost of reproduction (Obeso, 2002), allocation to flower and seed production depends on environmental conditions, such as the availability of and competition for resources, and these could affect female and male functions differently. Such ecological parameters could influence the evolution of gender plasticity or dimorphism (Delph, 2003; Tonnabel et al., 2017). Last, because reproduction affects gene transmission, sexual and mating systems can be selected even if they do not directly affect survival and fecundity, as in the well-known example of selfing rate modifiers that can evolve without immediately affecting the mean fitness of a population (Fisher, 1941).

Our goal was two-fold: first, by providing an integrated analysis of co-variation of plant reproductive traits, we aimed to get a better understanding of the processes involved in mating and sexual system evolution. In particular, we asked what sets sexually dimorphic species apart from other outcrossing species. Second, we aimed to bring pollination-related traits together with life history and ecological traits for a more synthetic understanding of plant form and function. To this end, we compiled a data set 21 traits related to or associated with reproduction for 360 species sampled from across the angiosperm tree of life, and determine the major axes of variation. Compared to previous studies based on pre-existing databases (e.g. Kattge et al., 2020) or datasets (e.g. from community surveys as in E-Vojtkó et al., 2022; Lanuza et al., 2023), we designed our sampling to maximize coverage of major taxa in angiosperms, including all families containing at least 100 species, and almost all orders. Beyond the questions we addressed, we thus also provide a unique and original dataset of reproductive diversity in angiosperms.

Materials & Methods

Data collection

We collated trait data for angiosperm species using the PROTEUS collaborative database (Sauquet, 2019). Mating system traits are the least well documented in the literature, so we first selected a list of 207 species for which information about the mating system (e.g., outcrossing rates, self-(in)compatibility) was available in the literature, and that already had some floral trait

data in PROTEUS. Next, we added dioecious species that already had some data in PROTEUS or belonged to completely dioecious families, as the first list contained only hermaphroditic species. Finally, we adjusted the species sampling by adding at least one species from each family with more than 100 species, and adding more species for the most species-rich families (e.g., Asteraceae, Orchidaceae) choosing species in the main clades of the family to best represent its diversity. This lead to an initial set of 363 species.

We selected a list of traits based on prior knowledge of how reproduction-related and associated traits influence evolutionary success (Helmstetter et al., 2022; Anderson et al., 2023). These traits were selected primarily to encompass the main aspects of angiosperm reproduction and include those related to mating system, sexual system, floral morphology (flower sex, ovary position, flower colour, flower size, flower symmetry) as well as dispersal distance/mode and pollination mode. We also included several vegetative traits related to growth form and lifespan, which can also be related to reproductive traits. For each of these 22 traits, detailed scoring instructions were provided; for traits already in PROTEUS, we used the instructions from Sauquet et al. (2017) and Schönenberger et al. (2020), while for newly added traits, we compiled instructions based on already available guidelines (Perez-Harguindeguy et al., 2013; Cardoso et al., 2018). These instructions are available as Supplementary File XXX . Seed mass was added outside PROTEUS as the average value for each of the species according to the Seed Information Database (Society for Ecological Restoration et al., 2023).

To compare our results with classical functional trait spaces, we also analysed a data set of six plant traits (leaf area, leaf mass per area, leaf nitrogen per mass, diaspore mass, stem specific density and plant height) from > 45,000 species (Díaz et al., 2016; Díaz et al., 2022). This data set was filtered to leave only those species that had measurements for at least four traits (n = 7968) to ensure distances could be calculated between all species pairs while increasing computational feasibility and accuracy of distance calculations.

Trait encoding

To facilitate downstream analyses, we modified the initial trait encoding as recorded initially to obtain 21 traits that could be used in the analyses (Table S1). For categorical traits, we reduced the number of states to 2 to 4 for comparisons to be informative. In some cases we split an initially complex trait into multiple different ones for easier interpretation of results (e.g., habit was recoded into woodiness, climbing and aquatic). For quantitative traits, if several values were available for a species/trait combination (either several measurements or indication of minimum and maximum values), we took the mean of these. The outcrossing rate was transformed into a categorical trait using three bins: selfing (< 0.2), mixed ($0.2 - 0.8$) and outcrossing (> 0.8), which allowed us to combine species with a proper estimate of outcrossing rate with species for which this simple qualitative classification was available ("Phenotypic mating system" trait in PROTEUS).

We used two different ways of encoding the categorical traits. Either we had one variable per trait with as many values as there are states in the trait, plus separate values for cases in which the trait was polymorphic for a species (we refer to this encoding as the "original" data set). Alternatively, categorical traits were encoded using a one-hot approach, where each category of a trait is treated as a distinct binary variable (e.g., woody vs non-woody; Harris and Harris, 2013). While this may introduce some redundancy (e.g., most species that are herbaceous are not woody and vice-versa), it is an alternative way of dealing with polymorphic states while keeping the relations between values. For example, in the original encoding, a species that can be both woody and herbaceous is assigned to a separate category with no explicit relation to other woody and herbaceous species, while in the one-hot encoding, such a species is similar to both herbaceous

and woody species.

Filtering, transformations and missing data imputation

To limit the impact of missing data on our analyses, traits were removed from the original and one-hot data sets if they contained more than 50% missing data in the original data set. Likewise, species were removed if they contained more than 50% missing data. We log-transformed quantitative traits to conform better to normality expectations, except for fusion of ovaries that is coded as a proportion. We also scaled and centered these variables to limit biases when building the trait spaces.

After filtering, many of the traits still contained missing data. To assess the effect of this missing data on distances between species and to facilitate some downstream analyses we conducted imputation with *missForest* (Stekhoven & Bühlmann, 2012). We followed the approach outlined in Debastiani et al. (2021). Briefly, a pairwise phylogenetic distance matrix (see below for how the phylogenetic tree was generated) containing all species was decomposed to a set of eigenvectors using the function *PVRdecomp* from the 'PVR' R package. The first 10 eigenvectors were added to the trait data as additional, complete traits to conduct imputation of missing data with *missForest*. This was done separately for the original and one-hot data sets.

Trait spaces

To build trait spaces we first calculated pairwise distance matrices among species using Gower's distance (Gower, 1971) with the function *daisy* from R package 'cluster' (Maechler et al., 2022). Gower's distance was used because it can deal with missing data and mixed data types (e.g. categorical and continuous). We then performed principal coordinate analysis (PCoA) on the resultant distance matrix with the *pcoa* function of the R package 'ape' (Paradis & Schliep, 2019) to generate a set of orthogonal eigenvectors and their associated eigenvalues. For the one-hot data, we used the *wcmdscale* function in the R package 'vegan' for the PCoA, and fitted the individual traits on the resulting trait space with the function *envfit*. We also built an additional trait space using the traits in the Díaz et al. (2022) data set. To compare how this functional trait space differs from our own we extracted the 161 species common to both datasets and re-built the two corresponding trait spaces.

We quantified the "quality" of dimensionality reduction of the resultant trait spaces using the method outlined in Mouillot et al. (2021). Briefly, the difference between the initial distance matrix and the distance matrix after dimensionality reduction is examined using PCoA. High-quality trait spaces are those in which a reduced number of PCoA dimensions accurately represents initial distances among species. Quality is quantified using the area under curve (AUC) metric for each number of retained PCoA axes. This approach provides an indication for how many axes are sufficient to summarize the variation in the initial dataset.

PCoA is a linear dimension reduction approach that does not account for more complex non-linear patterns. We thus also used an alternative dimensional reduction approach, Uniform Manifold Approximation and Projection (UMAP, McInnes et al., 2020), to visualise non-linear patterns in our data. UMAP is based on manifold learning techniques and allows the user to assess patterns at local and global scales simultaneously, depending on the size of the neighbourhood ('n_neighbours') chosen.

Clustering

We performed three different clustering approaches in order to account for variation in output by different algorithms. First, we conducted clustering using a hierarchical approach with the *hclust* function of the 'stats' R package and the ward.D2 algorithm (Murtagh & Legendre, 2014). This approach uses a dissimilarity matrix and progressively groups all species into clusters, starting with a cluster per species and joining the most similar clusters until a single cluster is reached. Second, we clustered species using k-prototypes (Huang, 1998) implemented in the 'clustMixType' package (Szepannek, 2018). This method is similar to k-means but allows for mixed data types. Euclidean distances are used between quantitative traits and the number of matching categories is used for qualitative traits. Third, we used partition around medoids (PAM) (Kaufman & Rousseeuw, 1990) as implemented in the 'cluster' R package (Maechler et al., 2022). This method is based on determining a set of medoids (points) that represent the structure of the data. Each species is assigned to a medoid, with the goal of minimizing the sum of the dissimilarities of each species to their closest medoid. The Gower's distance matrix was used as input for this approach.

We used different approaches to determine the number of clusters (k) depending on the clustering method. For hierarchical clustering, values of k were calculated based on changes in average silhouette width. For the remaining two methods (PAM and k-prototypes), we calculated 'robust' groups of species - those that remained clustered together as values of k were increased from $k = 2$ to $k = 7$. Groups were only considered if they contained 10 or more species to facilitate downstream interpretation. For the main clusters (or robust groups) we calculated functional diversity indices using the R package mFD (Magneville et al., 2022) to compare their characteristics.

Phylogenetic signal, transition rates and ancestral state estimation

We built a phylogenetic tree among our species using V.PhyloMaker2 (Jin & Qian, 2022). We used the default 'GBOTB.extended.TPL' tree that is derived from a large phylogenetic tree of all seed plants (Smith & Brown, 2018) and built the tree using the default approach described as 'scenario 3' (Jin & Qian, 2022). Prior to building the tree we standardized genus and species epithets using U.Taxonstand (Zhang & Qian, 2023) and a database derived from the World Checklist of Vascular Plants (Govaerts et al., 2021) found at https://github.com/nameMatch/Database. Using this tree we calculated the level of phylogenetic signal for each one-hot encoded trait to examine how well phylogeny, or the relatedness of species, explains the distribution of trait values and states. For categorical traits we used the δ statistics (Borges et al., 2019), a phylogenetic analog of Shannon entropy. Significance was tested by randomly shuffling states 100 times, recalculating δ and recording the percentage that were lower than the observed value for δ. For continuous traits we used the *phylosig* function in the 'phytools' package (Revell, 2012) to calculate λ (Pagel, 1999) and 'K' (Blomberg et al., 2003) with 1000 simulations to calculate p-values.

We then examined how transition rates varied among the different categorical traits in our data set. Transition rates and ancestral states at nodes in the phylogenetic tree were calculated using corHMM (Boyko & Beaulieu, 2021). We built a model of trait evolution for each one-hot encoded trait independently. We estimated a transition rate (per million years) of change between the two states using an equal rates ('ER') model with one rate category in which the transition rate is the same in each direction, thus providing a single rate per binary trait that can be compared to others. We also used the clusters and robust groups generated by our different clustering approaches as states and built models of trait evolution in which transition rates were free to vary (all-rates-different or 'ARD').

Simulated data sets

To determine how phylogeny influences trait space for our set of species and traits we simulated trait data using the phylogenetic tree of our species with models of trait evolution. For continuous traits we fitted Ornstein-Uhlenbeck (OU) models using the *fitContinuous* function in the R package 'geiger' (Pennell et al., 2014) to estimate OU model parameters and root state values. For categorical traits we fitted continuous-time Markov models ('Mk') to generate transition rate matrices and ancestral likelihoods for the root state. We then used the estimated parameters and the phylogenetic tree to simulate new datasets with *rTraitCont* from the 'ape' R package (Paradis et al., 2004) and *sim.history* from 'phytools' (Revell, 2012). Traits were simulated independently and then combined into a single simulated dataset, from which we calculated distance matrices and ran PCoAs, as above.

Results

After recoding and filtering, our final data set consisted of 21 traits (Table S1) for 360 species and 12.5% missing data, down from 25% initially (Fig. S1). With 170 out of 416 families, and 61 out of 64 orders (Fig. 1, Table S2; The Angiosperm Phylogeny Group et al., 2016), our dataset included a broad range of angiosperm diversity. Imputing data or using one-hot encoding only slightly changed Gower's distances between species, showing that missing data and encoding have relatively little impact on the following results ($R^2 = 0.90$ and 0.83, respectively for the comparison of distance matrices, Fig. S2). In the following we did not use the imputed data unless explicitly stated.

Exploring reproductive trait space

From the distance matrix, we conducted a PCoA on our data sets to summarise and visualize variation (Figs. 2 and 3). The first two PCoA axes explained 30 to 34% of the variation, depending on the data set used. These first two axes were correlated with a combination of general life-history and floral traits. The life-history traits are woody vs. herbaceous, lifespan and plant height, and they were distributed along a diagonal in this 2D trait space. Many pollination-related traits were more or less orthogonal to this diagonal: flower sex, floral reward, biotic vs. abiotic pollination, showiness and plant sexual system (Fig. 3). The third PCoA axis was mainly associated with dispersal (mode and distance), and the fourth with ovary position and flower symmetry (Fig. S3). By calculating the correlation of the traits to the PCoA axes (Table S3) and to one another (Fig. S4), we confirmed that there are two groups of traits, one containing vegetative traits, mating and dispersal, and another containing flower anatomy, pollination and sexual system.

We identified three main density peaks in the trait space formed by the first two PCoA axes. The two most dense peaks corresponded to herbaceous species with bisexual flowers and woody species with bisexual flowers (Fig. 2), while a less dense peak corresponded to woody species with unisexual flowers. Species outside of these high density areas (e.g., at the top of Figure 2) tended to be abiotically pollinated herbaceous species.

Comparison of reproductive and vegetative trait spaces

We compared this trait space to the one obtained by Díaz et al. (2016) with six traits classically used in plant functional ecology (Díaz et al., 2022). First, we found that the species in our dataset are scattered throughout the mainly vegetative trait space of Díaz et al. (2016), as shown in Fig. 4a.

There is almost no correlation between the Gower's distances obtained by using the six functional ecological traits and the full trait set we gathered in this study (Fig. S5).

Then, using the approach of Mouillot et al. (2021), we calculated statistics allowing comparison of our trait space to others. The AUC criterion, which indicates how well the first n axes summarize the total variation in the data set, indicated that to get a good representation of our trait space, we must keep more dimensions than for the data set derived from Díaz et al. (2022) as shown in Fig. 4b. This is equally the case when comparing our results to those of the other trait spaces analyzed by Mouillot et al. (2021). This means that, in our data set, traits are weakly redundant and most of them contribute small but significant amounts of variation that cannot be reduced to variation in other, more structurally important traits. We did not see notable differences in the trait space quality analyses between our different data sets (original, imputed, one-hot; Fig. S6).

Describing major reproductive strategies

UMAP (Fig. 5), which uses an alternative decomposition approach that allows the visualisation of non-linear and local patterns, demonstrated a clear separation of different groups across the trait space, where three major groups gradually separated in the first two axes as the size of the neighbourhood (knn) decreased (Fig. S7). These groups corresponded to the woody bisexual, herbaceous bisexual, and unisexual species groups identified using standard PCoA, the difference being that herbaceous unisexual species were associated more closely to the woody unisexual ones in the UMAP trait space.

The number of major reproductive strategies ranged from three to five depending on the clustering method used and evaluation criteria. To tackle this subjectivity issue, we examined how cluster membership changed as values of k were changed using Sankey plots (Figs. S8 and S9). We then identified which species stayed consistently together despite a changing k, which we term 'robust groups'. Taking the results from the original data set as examples, we found that the number of robust groups was eight for k-prototypes (Fig. S10) and seven for PAM (Fig. 5). The differences between the two methods were rather subtle, mainly the number of species (146 and 46 respectively) that did not belong to a robust group. These species tended to fall between two robust groups (Fig. 5). The other two methods were hierarchical, i.e., species did not move between major groups as k increased.

The overarching pattern in our clustering tended to reflect the first two axes of our trait space decompositions, that is, three different combinations of flower sex (or sexual system) and woodiness. These were (1) species with unisexual flowers (i.e. dioecious and monoecious, regardless of woodiness), (2) woody species with bisexual flowers and (3) herbaceous species with bisexual flowers, as exemplified by hierarchical clustering on the original data set with three clusters (Figs. S11 and S12). These three groups were present to at least some extent in the results of each of our clustering approaches. Those approaches that yielded more complex results tended to split these three major groups by other traits including dispersal distance/mode, pollination mode, mating system and floral symmetry (e.g. Figs. 5, S10 and S13).

Functional diversity indices for the three clusters from our hierarchical clustering (Table S4) revealed that species with unisexual flowers (cluster 2) were furthest from the centre of the trait space while woody species with bisexual flowers (cluster 3) were the closest (functional dispersion, FDis). Likewise, species in cluster 3 were also the most tightly packed in the trait space (functional mean nearest neighbour distance, FNND, and functional mean pairwise distance, FMPD) while species with unisexual flowers (cluster 2) and the herbaceous species with bisexual flowers (cluster 1) were generally further apart. Functional divergence (FDiv) was highest for cluster 1 indicating that species were more often at the extremes (the edge of the cluster), and that there is still a lot

of diversity in this cluster. The unisexual cluster had the highest values of functional originality (FOri) and functional specialisation (FSpe), which means its species were among the most distinct in the global species pool. Species in cluster 3 (woody, bisexual) had the lowest values of these indices. All in all, these indices confirm the results inferred by visual inspection of the PCoA trait space (Fig. 2) with respect to the density of the groups of species with bisexual flowers, while the species with unisexual flowers were more widely scattered.

Trait space and phylogeny

Almost half of the qualitative traits (17/35 one-hot traits) showed significant levels of phylogenetic signal, though the majority of these were weak (Table S5). Those qualitative traits with the strongest phylogenetic signal were biotic/abiotic pollination, lack of floral reward and aquatic habit. All quantitative traits had significant phylogenetic signal, with values of λ ranging from 0.74 for flower size to 1.08 for fusion of ovaries (Table S6).

From the fitted transition matrices, we extracted the transition rates for each traits. The traits that had the highest rate were biotic/abiotic dispersal mode and outcrossing mating system (Table S7). Some of the least common traits, aquatic and climbing habit, rare types of symmetry, had the lowest rates as they evolve infrequently.

We then repeated the ancestral state reconstruction with the original dataset and allowing all rates to be different (ARD models) to produce transition rate matrices that were used to simulate new data sets where all traits evolved independently. As expected, distance matrices from these simulated data sets were not correlated to the real data sets and maximum distances were greater when using real data (Fig. S2). When we ran PCoA on simulated data sets we found that the first three to four axes explained substantially less variation than any of the real data sets (Fig. S14). Thus, at least the first three PCoA axes from the analysis of our original data set are due to co-evolution of traits rather than neutral phylogenetic co-occurrence.

When reconstructed on the phylogeny with an ARD model, the three clusters from our hierarchical analysis showed many transitions across angiosperms (Fig. 1). There were only few clades for which all species belonged to the same cluster, and the largest among them did not contain more than 10 species. Due to the high number of missing species and the sampling strategy (maximizing the representation of angiosperm diversity) the estimated transition rates and inferred ancestral states should not be directly interpreted. Interestingly however, the ancestor of the angiosperms was inferred to be a woody species with bisexual flowers, which is in accordance with other studies (Zanne et al., 2014; Sauquet et al., 2017).

Discussion

Reproductive strategies beyond plant economics

We here characterize the reproductive strategies of angiosperms, using an extensive sample representative of their phylogeny, and 21 traits related to or associated with reproduction. This analysis reveals that the inclusion of reproductive traits adds an important dimension to the angiosperm trait space.

Classical work has identified several plant strategies based on plant size, investment in biomass and reproductive output (Grime, 1974; Díaz et al., 2016; Salguero-Gómez, 2017). These ecological strategies have been considered to arise from trade-offs in optimization of resource use and dealing with stress, given the physiological and morphological constraints to plant growth. Optimization and economics rationales have also been applied to plant organs to define the "leaf economic

spectrum", where leaves are defined as "costly" vs. "cheap" based on the ratio of leaf area and mass or nitrogen and carbon content (Wright et al., 2004; Díaz et al., 2016) or the "flower economic spectrum" organized along two axes, flower production costs and flower longevity (Roddy et al., 2021). In these theories, reproduction has been summarized largely by the production of seeds in varying quantities and qualities.

We argue that other aspects of reproduction, that can be affected by different levels of selection, should be taken into account. This renders framing trait correlations in terms of trade-offs more complex, as advantages and disadvantages of traits occur at different scales (i.e. demographic and evolutionary, Anderson et al., 2023). A substantial amount of co-variation among plant traits is related precisely to the way pollen is transferred, and this key part of plant reproduction has been previously overlooked (cf E-Vojtkó et al., 2020). Here we show that flower sex, sexual system, pollination mode, floral reward and to a lesser extent flower size, were all associated with an axis of variation that is almost orthogonal to the well-documented plant and seed size axis. Although the term "reproductive strategies" has been used before to refer to seed traits only (Salguero-Gómez et al., 2016), we propose to extend its meaning to include flower, pollination and mating traits as well. These reproductive strategies thus have more dimensions than only a "fast-slow" axis (Salguero-Gómez, 2017).

The correlates of sexual and mating systems

An important aspect of reproduction is outcrossing, and plants have evolved several ways to achieve this. We found that the variation in selfing and outcrossing *per se* is mostly associated with plant size and life span, as has been widely acknowledged. Dioecy has sometimes been viewed as an easier way to achieve outcrossing than (genetic) self-incompatibility, the latter being a complex trait that evolves infrequently (Barrett, 2013). However, we found that sexual system is clearly associated with a number of other pollination-related traits, and not primarily with plant size and lifespan. Despite the fact that we considered all dioecious species as outcrossing, sexual system is not correlated with mating system (Fig. 3 and S4). This indicates that flower unisexuality evolves under pressures that are not primarily related to the avoidance of selfing, although this has frequently been considered to be the case (e.g. Charlesworth & Charlesworth, 1987). Instead, flower unisexuality could evolve through ecological interactions with pollinators, or the absence thereof (cf Bawa, 1980; Charlesworth, 1993). Many studies on the evolution of dioecy consider monoecy and flower cosexuality as similar strategies, as the latter two do not exclude self-fertilization *a priori*. . In the light of our results it seems more appropriate to focus on flower unisexuality, combining monoecy and dioecy, as a distinct strategy. Indeed, transitions between monoecy and dioecy seem to occur rather frequently (Renner, 2014; Goldberg et al., 2017; Käfer et al., 2017), which is plausible if both occur in otherwise similar species.

By contrast, the variation from selfing to outcrossing is not consistently linked to floral traits. Indeed, selfing is known to occur primarily in annual species, while the vast majority of trees reproduce through outcrossing. While the so-called "selfing syndrome" might be readily observed among closely related species, with a reduction in corolla, stamen or pistil size (Sicard & Lenhard, 2011), our sampling does not allow the detection of such covariation at the angiosperm scale. A transition to predominant selfing can arise in very different pollination contexts, e.g., wind-pollinated grasses (Burgarella et al., 2023), small-flowered herbs with generalist pollinators (Sicard et al., 2011), or in groups where specialist pollination syndromes have evolved (Rose & Sytsma, 2021). Furthermore, selfing is not expected to persist at long evolutionary timescales, as it is often associated with higher lineage turnover rates (Goldberg et al., 2010). The scope for co-evolution of multiple traits is thus rather limited. However, we expect variation in outcrossing rates to be

linked with variation in flower traits at smaller scales, e.g., on the genus or family level.

The major reproductive strategies of flowering plants

Species are not equally distributed in the reproductive trait space (Fig. 2). The largest density of species occurs in the bisexual/monomorphic part of the trait space, with two density peaks also observed by Díaz et al. (2016) based on vegetative traits, i.e., herbaceous and woody species. The species in these two density peaks are clearly biotically pollinated and have mostly brightly coloured flowers, which corresponds to the majority of angiosperms. A less dense group can be seen in the PCoA analysis, corresponding to trees with unisexual flowers, often wind-pollinated, and containing many dioecious species. These species occupy a relatively large part of the trait space, despite this group being less species-rich. Wind-pollinated herbs occupy an even larger part of the trait space, at even lower densities. These species often (but not uniquely) have unisexual flowers and many aquatic species are found among them. They represent rather unique strategies that did not evolve frequently (and do not consistently cluster together; Fig. S13).

The three distinct peaks of species density in the trait space correspond to three different reproductive strategies. Nevertheless, both within the peaks as between them, much variation remains. First, our results indicate that other floral characteristics, e.g., flower symmetry, explain part of the variation within the strategies. This variation, which is secondary at the angiosperm level, will likely become more important as one focuses on smaller clades or specific floras, as some of the variation is reduced (only trees or herbs, all animal- or wind-pollinated, etc.). This is partly observed on the non-linear projection where sub-clusters appear within at least two of the three main clusters (Figure 5). Second, species falling outside these density peaks could either harbour trait combinations that are beneficial in specific conditions, or represent trait combinations that do not lead to long-term evolutionary success (e.g. Olivieri et al., 2016).

Dispersal mode and distance were the major traits associated with the third PCoA axis, thus mostly independent from variation in woodiness/plant size and pollination/flower sex. This does not, however, imply that these traits evolve without constraints from the traits associated with the first two axes, but rather that the relation is a complex one. As can be seen in Fig. 3, biotic dispersal is associated with larger height and longer lifespan, and long-distance dispersal to flower unisexuality and sexual dimorphism. However, these traits explain only a small fraction of the variation in dispersal mode and distance, which seem to be further correlated to traits not present in our data.

We demonstrated that the strength of trait association we observed is much stronger than under a null model in which traits would evolve along the phylogeny but without interaction. On small evolutionary timescales, traits might simply be correlated due to shared ancestry ("phylogenetic inertia"). However, the evolutionary timescale of this study (crown node of angiosperms 140-270 Ma; Sauquet et al., 2022) is such that all traits included here have had many transitions. If traits that have been shown to be labile remain associated at large evolutionary timescales, the most likely explanation is that their association is selectively favoured.

Integrating floral and mating traits in plant functional diversity

It is increasingly recognized that floral traits should be considered when characterizing plant functional diversity. Recently, two studies have attempted to do so. First, E-Vojtkó et al. (2022) studied two datasets of European species (central Europe and Alps), and found that floral trait variation is largely independent of variation in vegetative traits, a conclusion similar to ours. They did not seem to find any clusters or density peaks in their trait spaces, which might be due to the

exclusion of trees and grasses in their analyses, and acknowledge that variation is reduced at the European scale they studied.

Second, Lanuza et al. (2023) analyzed plant reproductive strategies in the light of their interactions with pollinators. Their data, collected from studies on plant-pollinator interactions, contain no wind-pollinated species and only few woody species, but included measurements of flower number and the size of flower organs. The patterns they find seem to be mainly driven by these measurements, which further limits the comparison with our results. The most important axis of variation they found was characterized by the negative correlation between flower size and flower number whereas flower size did not have a large contribution to the major axes of variation in our results. The differences with the results from Lanuza et al. (2023) could thus be explained by differences in sampling, notably the small proportion of woody species in their study. Considering the world flora, almost half of the species are woody (FitzJohn et al., 2014), a proportion closer to the one in our data set.

A challenge for the inclusion of floral traits in large-scale evolutionary and ecological studies seems to be their lack of availability in databases. The largest plant database today, TRY, contains limited data about flowers as compared to vegetative and seed traits (Kattge et al., 2020). Many studies and floras contain such data, but collecting them requires botanical knowledge in order to correctly interpret the terminology. Although floras are increasingly available online, they still represent a small fraction of known plant species, with tropical floras typically being underrepresented (cf Römer et al., 2023). We argue that publicly available data with well-described trait standards are the most convenient way to ensure that datasets can be combined and extended.

Future studies will be able to refine the reproductive strategies described here and add potentially important traits that we have not been able to include. Such traits might be related to total flower display, synchronization and duration of flowering, or flower longevity, allowing more explicit comparisons with other studies (e.g., Roddy et al., 2021; Lanuza et al., 2023). In particular, it will be useful to study to what extent the variation in traits that are difficult to measure or observe can be summarized by traits that are more readily available. Ideally, it would be possible to describe the reproductive strategies with just a few traits, analogous to the Leaf-Height-Seed framework in plant ecology (Westoby, 1998). Given the diversity of pollination syndromes and flower morphologies in angiosperms it is likely that a similar framework requires more than three traits. Such a description of the reproductive functional diversity could then be used to improve our understanding of ecosystem functioning, as has been the case for vegetative functional diversity (e.g. Lamanna et al., 2014; Schuldt et al., 2019). Furthermore, reproductive strategies integrating many traits are useful for evolutionary studies in which it is challenging to study traits in isolation (cf O'Meara et al., 2016; Helmstetter et al., 2022; Anderson et al., 2023).

Acknowledgements

This research is supported by the Fondation pour la Recherche sur la Biodiversité (FRB) through the CESAB project 'DiveRS'. We thank the DiveRS members Sally Otto and Denis Roze for constructive discussions and feedback on the manuscript. We thank Nicolas Casajus, Nicolas Loiseau and Marion Chartier for advice on data analyses, Solenn Sarton for contributing data, and Maud Calmet for the organisation of the workshops of the DiveRS group. We thank the University of Vienna for hosting the PROTEUS database and eFLOWER server.

Figures

Figure 1: Phylogenetic tree representing the 360 species used in our analyses. Tips are labelled with circles indicating which group they were assigned to in our hierarchical clustering analysis. Broadly, the three major groups, or strategies, are summarised as: (1) Monomorphic herbaceous species (2) Dimorphic species (both woody and herbaceous) and (3) Monomorphic woody species. Fig. S12 shows the frequency of qualitative states per trait for species in each cluster. An associated principal coordinate analysis (PCoA) can be found in Figure S11. Pie charts at nodes depict the results of an ancestral state estimation analysis using an all-rates-different (ARD) model. Pies are coloured based on the ancestral state probability of each node. Some well-known orders that are present in our set of species are highlighted around the outside of the tree alongside representative silhouettes (which are not necessarily of species in our data set).

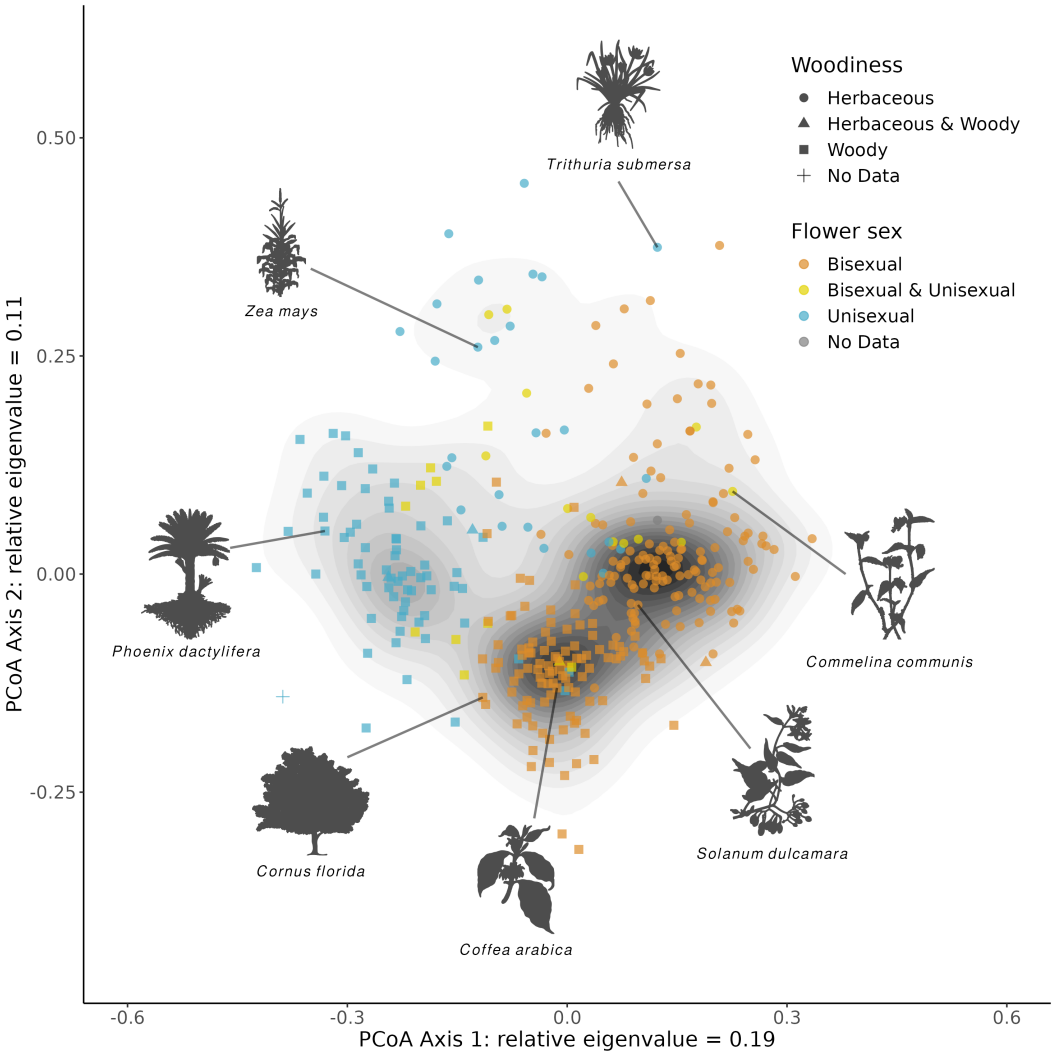

Figure 2: Location of 360 angiosperm species on the first two axes of our principal coordinate analysis (PCoA) on the original data set. These two axes represent approximately 30% of the variation in the data set. Points are coloured by the flower sex trait and shapes represent the woodiness trait. Polygons are used to show the density of points (= species), with darker sections indicating higher density. Silhouettes of various species found in our data set are shown with lines linking them to their position in the trait space.

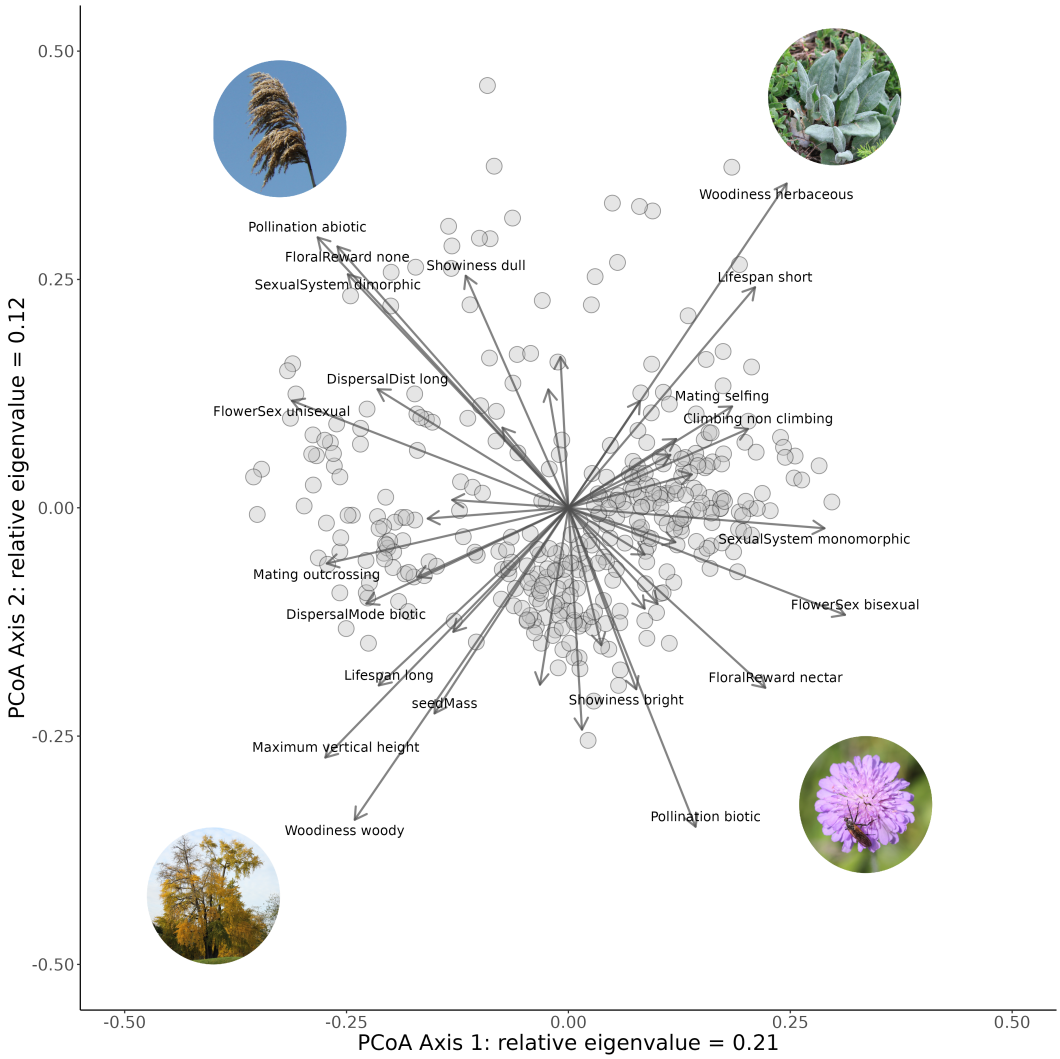

Figure 3: Principal coordinate analysis performed on the one-hot encoded data set. The first two axes represent approximately 33% of variation. Using one-hot encoding allowed us to plot arrows indicating how different states and traits affect the trait space, as well as which traits are acting in a similar manner. The length of the arrows indicates the strength of the effect of the state and labels were added to arrows representing those states with large effects. Images representing aspects of the trait space can be found in the four corners of the plot.

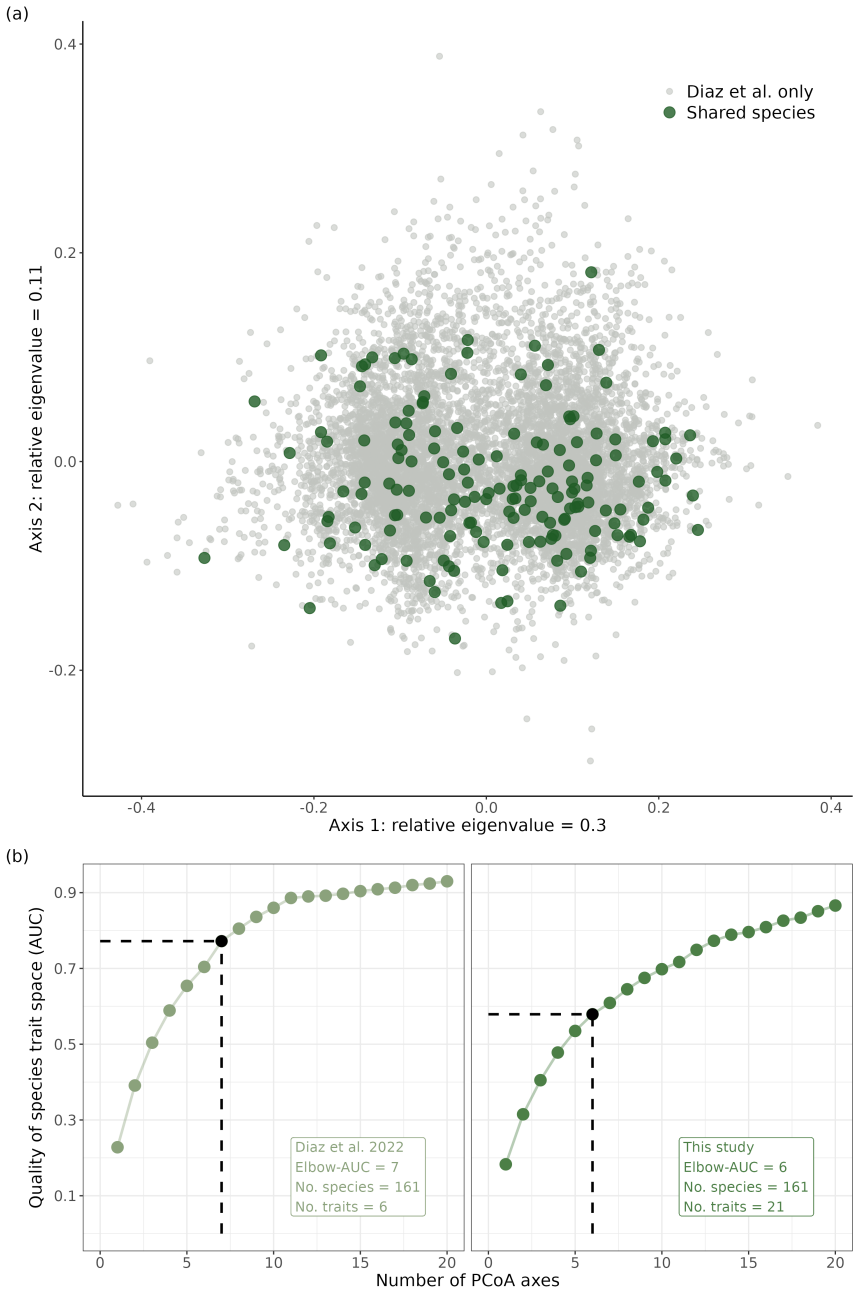

Figure 4: Comparison of the species sampled in the current study and the obtained trait space with the data and trait space of Díaz et al. (2022). Panel (a) shows a scatterplot of a subset of approximately 8,000 species from Díaz et al. (2022) on the first two axes of a principal coordinate analysis (PCoA). Species that overlap with those in this study are highlighted in green. Panel (b) shows two line plots (left: Díaz et al., 2022, right: this study) depicting how trait space quality (measured using AUC) changes as PCoA axes are added. Dashed lines show the number of axes to be kept, as determined by the elbow method, when optimising the trade-off between trait space quality and operationality. See Mouillot et al., 2021 for further details.

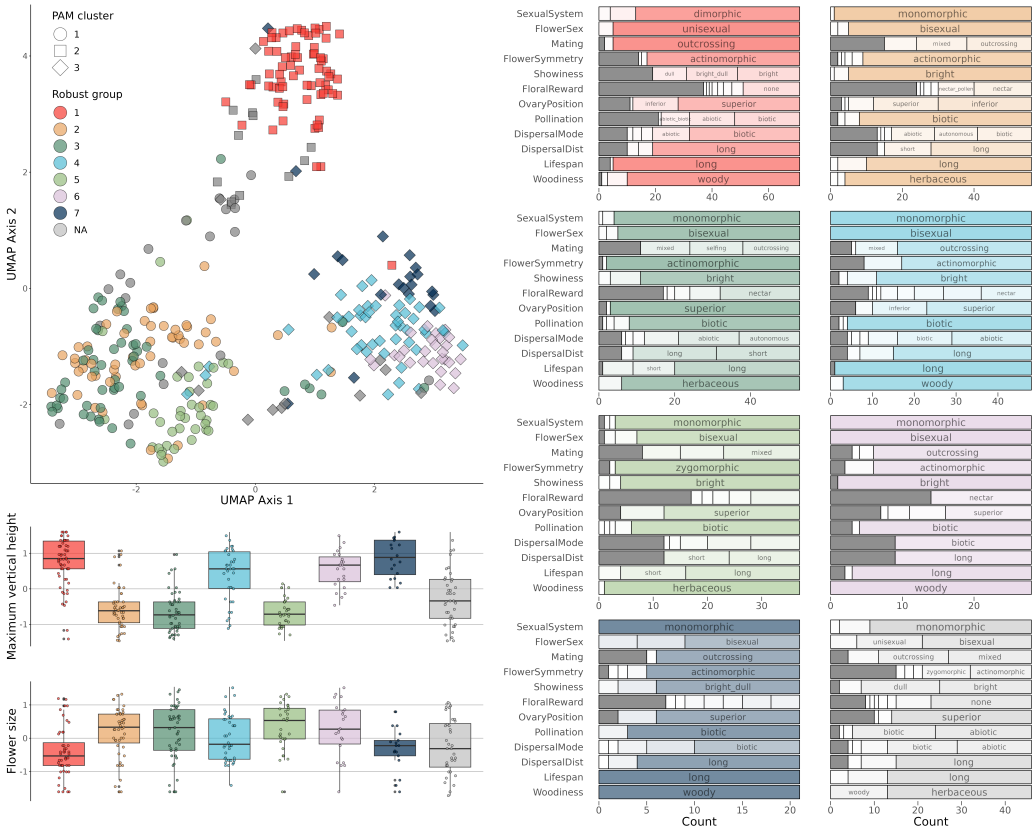

Figure 5: Characterizing robust groups identified using the Partitioning Around Mediods (PAM) clustering approach. Seven different robust groups were defined, and a final group was made with those species that did not fall into a robust group. The scatterplot in the top left shows the distribution of species along the first two axes of a UMAP (uniform manifold approximation) analysis with a neighbourhood size of 25. Points are coloured by robust group, and their shape indicates their PAM cluster assignment ($k = 3$). Boxplots in the bottom left show the distribution (after log transformation and scaling) of values for two traits per cluster: maximum vertical height and flower size. Values of species within each cluster are overlain and outliers were removed for display purposes. The stacked barplots on the right side show the frequency of states for 12 traits for each of the eight groups. Colours match those in the other panels and bars are split into sections depending on the frequency of each state. Sections representing those states with high frequencies are labelled and the dark grey sections correspond to missing data.

References

Anderson, B., Pannell, J., Billiard, S., Burgarella, C., Boer, H. d., et al. (2023). Opposing effects of plant traits on diversification. *iScience*, *26*(4). https://doi.org/10.1016/j.isci.2023.106362

Ashman, T.-L., Knight, T. M., Steets, J. A., Amarasekare, P., Burd, M., et al. (2004). Pollen limitation of plant reproduction: Ecological and evolutionary causes and consequences. *Ecology*, *85*(9), 2408–2421. https://doi.org/10.1890/03-8024

Barrett, S. C. H. (2002). The evolution of plant sexual diversity. *Nature Reviews Genetics*, *3*(4), 274–284. https://doi.org/10.1038/nrg776

Barrett, S. C. H. (2013). The evolution of plant reproductive systems: How often are transitions irreversible? *Proceedings of the Royal Society B: Biological Sciences*, *280*(1765), 20130913. https://doi.org/10.1098/rspb.2013.0913

Barrett, S. C. H., Harder, L. D., & Worley, A. C. (1996). The comparative biology of pollination and mating in flowering plants. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1345), 1271–1280. https://doi.org/10.1098/rstb.1996.0110

Bawa, K. S. (1980). Evolution of dioecy in flowering plants. *Ann. Rev. Ecol. Syst.*, *11*, 15–39. https://doi.org/10.1146/annurev.es.11.110180.000311

Blomberg, S. P., Garland JR., T., & Ives, A. R. (2003). Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution*, *57*(4), 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

Borges, R., Machado, J. P., Gomes, C., Rocha, A. P., & Antunes, A. (2019). Measuring phylogenetic signal between categorical traits and phylogenies. *Bioinformatics*, *35*(11), 1862–1869. https://doi.org/10.1093/bioinformatics/bty800

Boyko, J. D., & Beaulieu, J. M. (2021). Generalized hidden Markov models for phylogenetic comparative datasets. *Methods in Ecology and Evolution*, *12*(3), 468–478. https://doi.org/10.1111/2041-210X.13534

Burgarella, C., Brémaud, M.-F., Hirschheydt, G. V., Viader, V., Ardisson, M., et al. (2023). Mating systems and recombination landscape strongly shape genetic diversity and selection in wheat relatives, 2023.03.16.532584. https://doi.org/10.1101/2023.03.16.532584

Burgarella, C., & Glémin, S. (2017). Population genetics and genome evolution of selfing species. In J. Wiley & L. Sons (Eds.), *Els* (1st ed., pp. 1–8). Wiley. https://doi.org/10.1002/9780470015902.a0026804

Cardoso, J. C. F., Viana, M. L., Matias, R., Furtado, M. T., Caetano, A. P. D., et al. (2018). Towards a unified terminology for angiosperm reproductive systems. *Acta Bot. Bras.*, *32*(3), 329–348. https://doi.org/10.1590/0102-33062018abb0124

Charlesworth, D. (1993). Why are unisexual flowers associated with wind pollination and unspecialized pollinators? *The American Naturalist*, *141*(3), 481–490.

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Ann. Rev. Ecol. Syst.*, *18*, 237–268.

Cheptou, Imbert, Lepart, & Escarre. (2000). Effects of competition on lifetime estimates of inbreeding depression in the outcrossing plant Crepis sancta (Asteraceae). *Journal of Evolutionary Biology*, *13*(3), 522–531. https://doi.org/10.1046/j.1420-9101.2000.00175.x

Craig, D. M., & Mertz, D. B. (1994). Inbreeding effects on competition inTribolium. *Researches on Population Ecology*, *36*(2), 251–254. https://doi.org/10.1007/BF02514941

Crnokrak, P., & Roff, D. A. (1999). Inbreeding depression in the wild. *Heredity*, *83*(33), 260–270. https://doi.org/10.1038/sj.hdy.6885530

Debastiani, V. J., Bastazini, V. A. G., & Pillar, V. D. (2021). Using phylogenetic information to impute missing functional trait values in ecological databases. *Ecological Informatics*, *63*, 101315. https://doi.org/10.1016/j.ecoinf.2021.101315

Delph, L. F. (2003). Sexual dimorphism in gender plasticity and its consequences for breeding system evolution. *Evolution & Development*, *5*(1), 34–39. https://doi.org/10.1046/j.1525-142X.2003.03006.x

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., et al. (2016). The global spectrum of plant form and function. *Nature*, *529*(7585), 167–171. https://doi.org/10.1038/nature16489

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., et al. (2022). The global spectrum of plant form and function: Enhanced species-level trait dataset. *Scientific Data*, *9*(1), 755. https://doi.org/10.1038/s41597-022-01774-9

E-Vojtkó, A., de Bello, F., Durka, W., Kühn, I., & Götzenberger, L. (2020). The neglected importance of floral traits in trait-based plant community assembly. *Journal of Vegetation Science*, *31*(4), 529–539. https://doi.org/10.1111/jvs.12877

E-Vojtkó, A., Junker, R. R., de Bello, F., & Götzenberger, L. (2022). Floral and reproductive traits are an independent dimension within the plant economic spectrum of temperate central Europe. *New Phytologist*, *236*(5), 1964–1975. https://doi.org/10.1111/nph.18386

Fisher, R. A. (1941). Average excess and average effect of a gene substitution. *Annals of Eugenics*, *11*(1), 53–63. https://doi.org/10.1111/j.1469-1809.1941.tb02272.x

FitzJohn, R. G., Pennell, M. W., Zanne, A. E., Stevens, P. F., Tank, D. C., & Cornwell, W. K. (2014). How much of the world is woody? *Journal of Ecology*, *102*(5), 1266–1272. https://doi.org/10.1111/1365-2745.12260

Goldberg, E. E., Kohn, J. R., Lande, R., Robertson, K. A., Smith, S. A., & Igić, B. (2010). Species Selection Maintains Self-Incompatibility. *Science*, *330*(6003), 493–495. https://doi.org/10.1126/science.1194513

Goldberg, E. E., Otto, S. P., Vamosi, J. C., Mayrose, I., Sabath, N., et al. (2017). Macroevolutionary synthesis of flowering plant sexual systems. *Evolution*, *71*(4), 898–912. https://doi.org/10.1111/evo.13181

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., & Paton, A. (2021). The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data*, *8*(1), 215. https://doi.org/10.1038/s41597-021-00997-6

Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, *27*(4), 857–871. https://doi.org/10.2307/2528823

Grime, J. P. (1974). Vegetation classification by reference to strategies. *Nature*, *250*(5461), 26–31. https://doi.org/10.1038/250026a0

Grossenbacher, D., Briscoe Runquist, R., Goldberg, E. E., & Brandvain, Y. (2015). Geographic range size is predicted by plant mating system. *Ecology Letters*, *18*(7), 706–713. https://doi.org/10.1111/ele.12449

Harris, D. M., & Harris, S. L. (2013). 1 - from zero to one. In D. M. Harris & S. L. Harris (Eds.), *Digital design and computer architecture (second edition)* (pp. 2–52). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-394424-5.00001-X

Helmstetter, A. J., Zenil-Ferguson, R., Sauquet, H., Otto, S. P., Méndez, M., et al. (2022). Trait-dependent diversification in angiosperms: Patterns, models and data. *bioRxiv*, 2022.05.18.490882. https://doi.org/10.1101/2022.05.18.490882

Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, *2*(3), 283–304. https://doi.org/10.1023/A:1009769707641

Igic, B., & Kohn, J. R. (2006). The Distribution of Plant Mating Systems: Study Bias Against Obligately Outcrossing Species. *Evolution*, *60*(5), 1098–1103. https://doi.org/10.1111/j.0014-3820.2006.tb01186.x

Jin, Y., & Qian, H. (2022). V.PhyloMaker2: An updated and enlarged R package that can generate very large phylogenies for vascular plants. *Plant Diversity*, *44*(4), 335–339. https://doi.org/10.1016/j.pld.2022.05.005

Joly, S., & Schoen, D. J. (2021). Repeated evolution of a reproductive polyphenism in plants is strongly associated with bilateral flower symmetry. *Current Biology*, *31*(7), 1515–1520.e3. https://doi.org/10.1016/j.cub.2021.01.009

Käfer, J., Marais, G. A. B., & Pannell, J. R. (2017). On the rarity of dioecy in flowering plants. *Molecular Ecology*, *26*(5), 1225–1241. https://doi.org/10.1111/mec.14020

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., et al. (2020). TRY plant trait database – enhanced coverage and open access. *Global Change Biology*, *26*(1), 119–188. https://doi.org/10.1111/gcb.14904

Kaufman, L., & Rousseeuw, P. (1990). Partitioning Around Medoids (Program PAM). In *Finding Groups in Data* (pp. 68–125). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470316801.ch2

Lamanna, C., Blonder, B., Violle, C., Kraft, N. J. B., Sandel, B., et al. (2014). Functional trait space and the latitudinal diversity gradient. *Proceedings of the National Academy of Sciences*, *111*(38), 13745–13750. https://doi.org/10.1073/pnas.1317722111

Lanuza, J. B., Rader, R., Stavert, J., Kendall, L. K., Saunders, M. E., & Bartomeus, I. (2023). Covariation among reproductive traits in flowering plants shapes their interactions with pollinators. *Funct. Ecol.*, *0*, 1–13. https://doi.org/10.1111/1365-2435.14340

Lesaffre, T., & Billiard, S. (2020). The joint evolution of lifespan and self-fertilization. *Journal of Evolutionary Biology*, *33*(1), 41–56. https://doi.org/10.1111/jeb.13543

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2022). *Cluster: Cluster Analysis Basics and Extensions*.

Magneville, C., Loiseau, N., Albouy, C., Casajus, N., Claverie, T., et al. (2022). mFD: An R package to compute and illustrate the multiple facets of functional diversity. *Ecography*, *2022*(1). https://doi.org/10.1111/ecog.05904

McInnes, L., Healy, J., & Melville, J. (2020). Umap: Uniform manifold approximation and projection for dimension reduction. (arXiv:1802.03426). https://doi.org/10.48550/arXiv.1802.03426

Mouillot, D., Loiseau, N., Grenié, M., Algar, A. C., Allegra, M., et al. (2021). The dimensionality and structure of species trait spaces. *Ecology Letters*, *24*(9), 1988–2009. https://doi.org/10.1111/ele.13778

Munoz, F., Violle, C., & Cheptou, P.-O. (2016). Csr ecological strategies and plant mating systems: Outcrossing increases with competitiveness but stress-tolerance is related to mixed mating. *Oikos*, *125*(9), 1296–1303. https://doi.org/10.1111/oik.02328

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, *31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

Obeso, J. R. (2002). The costs of reproduction in plants. *New Phytologist*, *155*(3), 321–348. https://doi.org/10.1046/j.1469-8137.2002.00477.x

Olivieri, I., Tonnabel, J., Ronce, O., & Mignot, A. (2016). Why evolution matters for species conservation: Perspectives from three case studies of plant metapopulations. *Evolutionary Applications*, *9*(1), 196–211. https://doi.org/10.1111/eva.12336

O'Meara, B. C., Smith, S. D., Armbruster, W. S., Harder, L. D., Hardy, C. R., et al. (2016). Non-equilibrium dynamics and floral trait interactions shape extant angiosperm diversity.

Proceedings of the Royal Society B: Biological Sciences, *283*(1830), 20152304. https://doi.org/10.1098/rspb.2015.2304

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. https://doi.org/10.1038/44766

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., et al. (2014). Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, *30*(15), 2216–2218. https://doi.org/10.1093/bioinformatics/btu181

Perez-Harguindeguy, N., Diaz, S., Garnier, E., Lavorel, S., Poorter, H., et al. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Aust. J. Bot.*, *61*, 167–234. https://doi.org/10.1071/BT12225_CO

Petrone Mendoza, S., Lascoux, M., & Glémin, S. (2018). Competitive ability of Capsella species with different mating systems and ploidy levels. *Annals of Botany*, *121*(6), 1257–1264. https://doi.org/10.1093/aob/mcy014

Renner, S. S. (2014). The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany*, *101*(10), 1588–1596. https://doi.org/10.3732/ajb.1400196

Renner, S. S., & Ricklefs, R. E. (1995). Dioecy and Its Correlates in the Flowering Plants. *American Journal of Botany*, *82*(5), 596–606. https://doi.org/10.2307/2445418

Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Roddy, A. B., Martínez-Perez, C., Teixido, A. L., Cornelissen, T. G., Olson, M. E., et al. (2021). Towards the flower economics spectrum. *New Phytol.*, *229*(2), 665–672. https://doi.org/10.1111/nph.16823

Römer, G., Dahlgren, J. P., Salguero-Gómez, R., Stott, I. M., & Jones, O. R. (2023). Plant demographic knowledge is biased towards short-term studies of temperate-region herbaceous perennials. *Oikos*, *e10250*. https://doi.org/10.1111/oik.10250

Rose, J. P., & Sytsma, K. J. (2021). Complex interactions underlie the correlated evolution of floral traits and their association with pollinators in a clade with diverse pollination systems. *Evolution*, *75*(6), 1431–1449. https://doi.org/10.1111/evo.14220

Salguero-Gómez, R., Jones, O. R., Jongejans, E., Blomberg, S. P., Hodgson, D. J., et al. (2016). Fast-slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proc. Natl. Acad. Sci. U.S.A.*, *113*(1), 230–235. https://doi.org/10.1073/pnas.1506215112

Salguero-Gómez, R. (2017). Applications of the fast–slow continuum and reproductive strategy framework of plant life histories. *New Phytologist*, *213*(4), 1618–1624. https://doi.org/10.1111/nph.14289

Sauquet, H. (2019). PROTEUS: A database for recording morphological data and fossil calibrations.

Sauquet, H., Ramírez-Barahona, S., & Magallón, S. (2022). What is the age of flowering plants? *Journal of Experimental Botany*, *73*(12), 3840–3853. https://doi.org/10.1093/jxb/erac130

Sauquet, H., von Balthazar, M., Magallón, S., Doyle, J. A., Endress, P. K., et al. (2017). The ancestral flower of angiosperms and its early diversification. *Nature Communications*, *8*(1), 16047. https://doi.org/10.1038/ncomms16047

Schönenberger, J., von Balthazar, M., López Martínez, A., Albert, B., Prieu, C., et al. (2020). Phylogenetic analysis of fossil flowers using an angiosperm-wide data set: Proof-of-concept

and challenges ahead. *American Journal of Botany*, *107*(10), 1433–1448. https://doi.org/10.1002/ajb2.1538

Schuldt, A., Ebeling, A., Kunz, M., Staab, M., Guimarães-Steinicke, C., et al. (2019). Multiple plant diversity components drive consumer communities across ecosystems. *Nat. Commun.*, *10*(1), 1460. https://doi.org/10.1038/s41467-019-09448-8

Scofield, D. G., & Schultz, S. T. (2006). Mitosis, stature and evolution of plant mating systems: Low-Phi and high-Phi plants. *Proc. Biol. Sci.*, *273*(1584), 275–282. https://doi.org/10.1098/rspb.2005.3304

Sicard, A., & Lenhard, M. (2011). The selfing syndrome: A model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany*, *107*(9), 1433–1443. https://doi.org/10.1093/aob/mcr023

Sicard, A., Stacey, N., Hermann, K., Dessoly, J., Neuffer, B., et al. (2011). Genetics, evolution, and adaptive significance of the selfing syndrome in the genus capsella. *The Plant Cell*, *23*(9), 3156–3171. https://doi.org/10.1105/tpc.111.088237

Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, *105*(3), 302–314. https://doi.org/10.1002/ajb2.1019

Society for Ecological Restoration, International Network for Seed Based Restoration, & Royal Botanic Gardens Kew. (2023). *Seed Information Database (SID)*. https://ser-sid.org/

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Szepannek, G. (2018). clustMixType: User-Friendly Clustering of Mixed-Type Data in R. *The R Journal*, *10*(2), 200–208.

The Angiosperm Phylogeny Group, Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., et al. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, *181*(1), 1–20. https://doi.org/10.1111/boj.12385

Tonnabel, J., David, P., & Pannell, J. R. (2017). Sex-specific strategies of resource allocation in response to competition for light in a dioecious plant. *Oecologia*, *185*(4), 675–686. https://doi.org/10.1007/s00442-017-3966-5

Vamosi, J. C., & Vamosi, S. M. (2004). The role of diversification in causing the correlates of dioecy. *Evolution*, *58*(4), 723–731.

Westoby, M. (1998). A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant Soil.*, *199*(2), 213–227. https://doi.org/10.1023/A:1004327224729

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., et al. (2004). The worldwide leaf economics spectrum. *Nature*, *428*(6985), 821–827. https://doi.org/10.1038/nature02403

Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., et al. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, *506*(7486), 89–92. https://doi.org/10.1038/nature12872

Zhang, J., & Qian, H. (2023). U.Taxonstand: An R package for standardizing scientific names of plants and animals. *Plant Diversity*, *45*(1), 1–5. https://doi.org/10.1016/j.pld.2022.09.001