Supplementary materials

Pollination and mating traits underlie diverse reproductive strategies in flowering plants

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

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

[*] **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

Contents

1 Appendix S1: Trait scoring guide

General instructions:

- Don't score if you don't have a reference or an observation for the character
- Prefer data from wild populations, or at least wild genotypes (e.g. plants of wild origin cultivated in botanical gardens or greenhouses). If only data from cultivated crop or ornamental populations are available, please indicate it as a note
- If there are multiple states within the species (polymorphism, subspecies, varieties) code these as separate values / states. There is currently no way to link polymorphism between several traits, so if female flowers are large and blue and male flowers small and green, the only way to indicate this linkage is through comments.
- For rare traits, information is only available for the species possessing them. Please do not score such traits as "absent" yourself if you can't find any information, but only if there is an explicit mention of the absence in the species you're scoring.
- Many traits can be scored from general botanical literature (floras), but some traits (e.g. outcrossing traits, pollination, pollen) require specific literature.

Plant-level traits

1. Habit (D1)

Tree, shrub, liana, herb, vine, aquatic herb, subterranean. Lianas are woody climbers (including "climbing shrubs" etc), vines herbaceous climbers (thus, *Vitis vinifera* is a liana). Trees are usually taller than shrubs, and are characterized by one or several trunks, while shrubs have many ramifications. Aquatic herbs are plants that have their leaves at (e.g. *Nymphaea*) or under the water surface (e.g. *Ceratophyllum*), or are free-floating (e.g. *Eichhornia*).

2. Maximum vertical height (C1)

Maximum vertical height (i.e. not the height at first flowering). Indicate in m. Can be a single value or a range of values.

4. Plant sexual system (D1)

This trait combines flower sex and the distribution of the flowers among individual plants. In bisexuality and monoecy (including gynomonoecy and andromonoecy), plants are hermaphroditic, but the distinction is made at the flower level.

Bisexual: all individuals have perfect (i.e., bisexual) flowers; some authors use cosexuality or hermaphroditism for this state.

Monoecious: all individuals have both male and female flowers

Dioecious: each individual is either completely male or completely female.

Gynomonoecious: all individuals have both female and perfect flowers

Gynodioecious: each individual is either completely female or completely hermaphrodite. Hermaphrodites can be of two kinds, depending on the species (if you have the information, please indicate as a comment). Either each hermaphrodite individual has perfect flowers, or is monoecious (or gyno- or andromonoecious).

Andromonoecious: all individuals have both male and perfect flowers

Androdioecious: each individual is either completely male or completely hermaphrodite. Hermaphrodites can be of two kinds, depending on the species (if you have the information, please indicate as a comment). Either each hermaphrodite individual has perfect flowers, or is monoecious (or gyno- or andromonoecious).

Trimonoecy: all individuals have male, female and perfect flowers.

Polygamous: everything that doesn't fit in the preceding categories. Leave a comment.

5. Life history (D1)

Annual, biennial, perennial monocarpic, perennial polycarpic, perennial (undetermined). If possible, for perennial species, indicate whether it flowers only once and then dies (monocarpic), or whether it can flower multiple times (polycarpic). If the information is not available, choose "perennial (undetermined)". If the plant can have several lifecycles a year, it should be considered annual (leave a comment).

80. Phenotypic mating system (D1)

Score this trait only if the mating system is explicitly characterized and noted in a study. Please do not try to infer selfing or outcrossing yourself from flower showiness! Studies might use a combination of floral characters and/or empirical tests (ex P/O ratio, dichogamy, seedset from bagged and unbagged plants,...) in a comparative way, i.e. several species have been characterized in the same study. Use the categorization given by the author. If they used more than three categories, group them into three:

predominantly selfing: selfing, predominantly selfing, autogamy, obligate autogamy, facultative autogamy

mixed: mixed mating, facultative xenogamy

predominantly outcrossing: outcrossing, predominantly outcrossing, obligate outcrossing, xenogamy, obligate xenogamy

If the study provides information such as P/O ratio, autonomous selfing, outcrossing rate, etc.. but does not categorize or qualify the mating system don't do it yourself. We here also score apomixis, which is not a form of selfing strictly speaking, but is predicted to have some similar evolutionary, ecological and genomic consequences.

83. Self-incompatibility system (genetic) (D1)

The basic scoring is self-compatible and undetermined self-incompatible. In addition if you have information about the genetic mechanism of self incompatibility, that is phenotypic studies of pollen germination on stigmas, genetic crosses, and/or identification of the genes involved in self-incompatibility that make possible the distinction between gametophytic and sporophytic incompatibility, score the type of incompatibility, then score gametophytic self-incompatible or sporophytic self-incompatible. Note that this only applies to genetic self-(in)compatibility. For instance, heterostylous species should not automatically be scored self-incompatible, nor dioecious species.

84. Outcrossing rate (C1)

Value between 0 and 1. Note that outcrossing rate ¿ 1 can sometimes be given (due to statistical issues). In that case note 1. If the selfing rate is reported take 1 – selfing rate. If several populations have been measured in the same study then score all values (as separate entries). The additional table "Quality score" must be filled (not optional) with the method used to estimate outcrossing rate (corresponding to increasing quality)

1. Fis method, basically, outcrossing rate = (1 - Fis)/(1+Fis)
2. Population-based method: RMES (David et al., 2007) or INSTRUCT (Gao et al., 2007)

3. Pedigree-based method: Ritland's method = MLTR (Ritland, 2002) or BORICE (Koelling et al., 2012)

For pedigree-based methods, report the multi-locus outcrossing rate, not the single locus estimate (except if only one locus is used!). Additional equivalent methods exist but they are very rarely used. In addition the number type of genetic markers can be noted in the information table (recommended but not mandatory): allozyme, RAPD, RFLP, microsatellites, RadSeq, SNPs. When estimates are given for several populations, enter one value for each population. When Fis is used to infer outcrossing rate, make sure that individuals have been sampled in a single natural population. Don't consider collections of individuals (frequently noted as accessions) coming from a same "group" but not necessarily a population..

90. Pollination syndrome (D1)

Inferring how flowers are effectively pollinated requires detailed studies, as many flower visitors might not contribute to pollination. Alternatively, pollination can be assumed from inflorescence or floral traits. Do not infer the syndrome yourself, but only used published descriptions. Please use the confidence level to indicate how pollinators were inferred :

1. (I have doubts about this record): inferred from flower/inflorescence traits
2. (I think this is right): observations of flower visitors
3. (I am certain this is right): pollination effectively demonstrated

Multiple pollinators are possible. For biotic pollination, we do not strictly follow phylogenetic categories for biotic pollinators, but rather intuitive categories, which might be paraphyletic. Possible pollination syndromes are:

autonomous pollination occurs when pollen is transferred to stigmas without other vectors. Examples are pollen falling on stigmas or direct contact between anthers and stigmas. Cleistogamous flowers necessarily fall in this category

wind

water (including rain)

bee: Hymenoptera clade Anthophila, which includes honey bees, bumblebees, and solitary bees

beetle: Coleoptera

butterfly: day-active lepidopterans

moth: night-active lepidopterans

fly: Diptera suborder Brachycera (excluding mosquitos).

carrion fly and/or flesh fly: Diptera families Calliphoridae and Sarcophagidae. These are flies whose larvae live in rotting meat and dung. They are attracted by flowers that imitate (through scent, colour, and/or texture) these substances

wasp: Hymenoptera of the Apocrita suborder that are not ants or bees. Includes social and solitary wasps, hornets, parasitic wasps, fig wasps.

insect (other): use if the pollinating insects belong to none of the above categories

insect (general): use if insect pollination is indicated without further precision on the insect taxon.

bird

bat

non-flying mammal

other biotic pollinator

95. Dispersal syndrome (D1)

As the morphological descriptions do not always capture the fruit-associated structures that determine dispersal mode (e.g. *Fragaria*, *Ficus*), we also score the dispersal syndrome as inferred from actual dispersal observations, or interpreted from the fruit characters. The trait states are copied from Perez-Harguindeguy et al. (2016):

unassisted dispersal: the seed or fruit has no obvious aids for longer-distance transport and merely falls passively from the plant.

wind dispersal (anemochory) includes (A) minute dust-like seeds (e.g. *Pyrola*, *Orchidaceae*), (B) seeds with pappus or other long hairs (e.g. willows (*Salix*), poplars (*Populus*), many *Asteraceae*), 'balloons' or comas (trichomes at the end of a seed), (C) flattened fruits or seeds with large 'wings', as seen in many shrubs and trees (e.g. *Acer*, birch (*Betula*), ash (*Fraxinus*), lime (*Tilia*), elm (*Ulmus*), pine (*Pinus*)); spores of ferns and related vascular cryptogams (Pteridophyta) and (D) 'tumbleweeds', where the whole plant or infructescence with ripe seeds is rolled over the ground by wind force, thereby distributing the seeds. The latter strategy is known from arid regions, e.g. *Baptisia lanceolata* in the south-eastern USA and *Anastatica hierochuntica* (rose-of-Jericho) in northern Africa and the Middle East.

internal animal transport (endo-zoochory), e.g. by birds, mammals, bats; many fleshy, often brightly coloured berries, arillate seeds, drupes and big fruits (often brightly coloured), that are evidently eaten by vertebrates and pass through the gut before the seeds enter the soil elsewhere (e.g. holly (*Ilex*), apple (*Malus*)).

external animal transport (exo-zoochory); fruits or seeds that become attached e.g. to animal hairs, feathers, legs, bills, aided by appendages such as hooks, barbs, awns, burs or sticky substances (e.g. burdock (*Arctium*), many grasses).

dispersal by hoarding; brown or green seeds or nuts that are hoarded and buried by mammals or birds. Tough, thick-walled, indehiscent nuts tend to be hoarded by mammals (e.g. hazelnuts (*Corylus*) by squirrels) and rounded, wingless seeds or nuts by birds (e.g. acorns (*Quercus spp.*) by jays).

ant dispersal (myrmecochory); dispersules with elaiosomes (specialised nutritious appendages) that make them attractive for capture, transport and use by ants or related insects.

dispersal by water (hydrochory); dispersules are adapted to prolonged floating on the water surface, aided for instance by corky tissues and low specific gravity (e.g. coconut).

dispersal by launching (ballistichory); restrained seeds that are launched away from the plant by 'explosion' as soon as the seed capsule opens (e.g. *Impatiens*).

bristle contraction; hygroscopic bristles on the dispersule that promote movement with varying humidity.

no seeds or fruits: to be used for species that are fully clonal and don't have dispersules

Flower-level characters

100. Floral structural sex (D1)

From Schönenberger et al. (2020): "Flowers can be either bisexual (hermaphrodite) or unisexual. In our original study (Sauquet et al., 2017), we used a single character to distinguish among the many possible ways to be unisexual, depending on whether sterile organs of the opposite sex (staminodes or carpellodes) are found in flowers of a given sex, whether male and female flowers are found on the same plant (monoecy) or separate plants (dioecy), as well as the various intermediate combinations that exist (e.g., androdioecy, gynomonoecy). We have now simplified and divided this character into two and here capture only floral sex, distinguishing among three states: bisexual, incompletely unisexual (i.e., with pistillode in male flowers and/or staminodes in female flowers), and unisexual. All data from our previous dataset have now been rescored permanently into one of these three states. A second new character, Plant sexual system, resulting from the simplification outlined above, is excluded here because fossil flowers are usually found as individual, dispersed specimens, making it impossible to distinguish among different sexual systems such as monoecy and dioecy."

Added note: incompletely unisexual plant systems such as gynodioecy and androdioecy will typically be scored as polymorphic for floral structural sex. For instance, gynodioecious species with female and bisexual individuals (e.g., many *Dianthus* species, *Thymus*), each with structurally female and bisexual flowers, should be scored as unisexual / bisexual for this character. However, note that this character cannot be directly scored from plant sexual system without more information on floral structure, as gynodioecy and androdioecy can also refer to species with both unisexual and monoecious individuals. For instance, species of *Ficus* referred to as gynodioecious in the literature typically only have structurally unisexual flowers (with female individuals and figs and structurally bisexual but functionally male individuals and figs) and should be scored here as unisexual only. Likewise, plant sexual systems reported in the literature should be interpreted with caution, depending on whether structure or function is intended. For instance, *Amborella* is sometimes described as androdioecious, but is in fact functionally dioecious, having male individuals with structurally male flowers and female individuals with female flowers that are incompletely unisexual in structure (because they have 1-2 staminodes), but entirely unisexual in function. *Amborella* should be scored here as unisexual / incompletely unisexual.

102. Ovary position (D1)

From Schönenberger et al. (2020): "The ovary is the part of the gynoecium where the ovules are produced. The ovary may be located on the receptacle and thus be positioned above the insertion level of the remaining floral organs (i.e., the ovary is superior and the flower is hypogynous). Alternatively, the ovary may be embedded in the receptacle and therefore be located below the insertion level of the remaining floral organs (i.e., the ovary is inferior and the flower is epigynous). Flowers with a hypanthium may either have a superior ovary (perigyny; e.g., many *Rosaceae*) or an inferior ovary (epiperigyny) (Simpson, 2010). It is also possible that the ovary is inferior to a certain degree only, such as half-inferior, if the receptacle is surrounding the ovary to its mid-level. Here we recorded the ovary position either as superior, inferior, or one of the following intermediate states: ¼ inferior or less, half-inferior, ¾ inferior or more."

103. Flower length (C1)

In cm. Should be measured at anthesis, starting from the base of the receptacle or the base of an inferior ovary, excluding the pedicel.

104. Flower diameter (C1)

In cm; to be measured at anthesis.

107. Floral reward (D1)

The reward might be offered in the flowers or the inflorescences. The reward can be food for the visiting insects, such as pollen (including non-fertile pollen that is produced for reward purposes only), nectar, other sugar-containing liquids, oil (liquid fats), resin (solid or highly viscous secretions), food bodies (solid), or other tissues associated to the flower (e.g. stamens, bracts). Alternatively, food can be destined to larvae when the flowers serve as nurseries, either through dedicated structures (e.g. figs) or not: ovipositing insects of which the larvae feed on the ovules can nevertheless be pollinators. Alternatively, the reward offered by flowers can be shelter (e.g. sleeping places) or heat. Some flowers offer no reward, which is the case for abiotically pollinated flowers, but also for some attractive biotically pollinated, such as in the case of deceptive pollination. The precise mechanism of deceptive pollination can be described as a comment. See also Simpson and Neff (1981).

Perianth characters

207. Symmetry of perianth (D1)

From Schönenberger et al. (2020): “There are many ways in which flowers can be zygomorphic (monosymmetric, with a single plane of bilateral symmetry). Here we record perianth symmetry, regardless of androecium or gynoecium symmetry; thus the character is not applicable when the perianth is absent. We distinguish strict actinomorphy (i.e., polysymmetry, with three or more planes of bilateral symmetry) from spiral actinomorphy. In addition, disymmetry (two orthogonal planes of bilateral symmetry; e.g., *Papaveraceae*) and asymmetry are treated here as separate character states. As for the fusion of the perianth, this character is applied to the perianth as a whole. In case of flowers with two or more perianth whorls, species were scored as actinomorphic if all whorls are actinomorphic and as zygomorphic if one or more whorls are zygomorphic.”

239. Main color of perianth at anthesis (D1)

This is the putative attractive color. If there is more than one color (either within the same organ or among organs), enter as separate character states. The choice is quite reduced compared to all variation in colors that exists, but the goal is to create broad categories for large-scale comparisons. Yellow, green, cream, white, pink, brown, purple, red, grey, blue, orange. If the perianth is absent, this character should remain empty. If the most colourful organs are not the perianth, one should still indicate the colour of the perianth here, and indicate the identity and the colour of the attractive organs in traits 660 “Inflorescence attractive organ” and 661 “Inflorescence attractive colour”. Note that a solitary flower is considered a one-flowered inflorescence.

Male reproductive function

301. Number of fertile stamens (C1)

From Schönenberger et al. (2020): “Number of fertile (functional) stamens in bisexual or male flowers. Staminodes (co-occurring with fertile stamens) are not counted and female flowers are ignored for this character. Stamen number is highly variable within angiosperms and ranges from one (e.g., *Chloranthaceae*, Endress, 1987b) to several thousands (e.g., *Cactaceae*, Barthlott and Hunt, 1993). We record the number of fertile stamens in whorled or spiral flowers as a continuous character (with integer values of 1 and above). In cases of fusion among stamen whorls, the number of stamens may be difficult to determine. In such cases, additional information based on merism, anatomy, development, or comparison with closely related taxa may be taken into account. In synandria of Myristicaceae, for example, the number of fertile stamens can be deduced from the number of thecae present (Sauquet, 2003). In cases where stamen or anther morphology is not fully understood, we recorded the number of fertile stamens only when unequivocally clarified in the literature (e.g., *Malvaceae*, von Balthazar et al., 2004). Equivocal cases were left as missing data.”

Female reproductive function

401. Number of structural carpels (C1)

From Schönenberger et al. (2020): “Number of fertile or sterile carpels in bisexual or female flowers, recorded as a continuous character (with integer values of 1 and above). Contrary to the number of stamens (character 301), the number of co-occurring carpellodes (sterile carpels) is counted here because this number is often more easily obtained from the literature than the actual number of fertile carpels. However, consistently with our treatment of sexual dimorphism, the number of carpellodes in male flowers is ignored for this character. In multicarpellate, unilocular gynoecia with complete carpel fusion up to the stigma (e.g., *Primula*), it may be difficult to assess the number of carpels unequivocally. In such cases, we have scored the number of carpels only if it is well established based on anatomical or developmental investigations. Similarly, in gynoecia where one or more carpels are reduced (e. g., in the pseudomonomerous gynoecia of some *Arecaceae*, Stauffer et al., 2002), the total number of structural carpels was only scored when unequivocally determined in the literature. Contrary to the perianth and the androecium, we do not have a separate character for gynoecium merism here. This is because gynoecia with two or more whorls are rare in angiosperms and gynoecium merism therefore usually equals the number of carpels per flower."

403. Fusion of ovaries (C1)

From Schönenberger et al. (2020): "Degree of ovary fusion expressed as a fraction of the total length of the ovary (from the floral base to the apex of the ovary). Fusion of styles and stigmas is not taken into account here. Not applicable when there is a single carpel."

411. Number of ovules per functional carpel (C1)

From Schönenberger et al. (2020): "Number of ovules per carpel recorded as a continuous character (with integer values of 1 and above). Reduced (sterile) carpels are not taken into account here."

Inflorescences

660. Inflorescence attractive organ (D1)

The attractive function can be fulfilled by the perianth or other organs. These can be bracts, leaves, stamens, or other organs (indicate in a comment), multiple organs (please leave a comment) or none. Note that single flowers should be considered as one-flowered inflorescences.

661. Inflorescence attractive color (D1)

In many cases, the attractive function of the inflorescence is achieved by the perianth; thus, the attractive color is the same as the one scored in "239. Main color of perianth at anthesis", and can be ignored here. In other cases, indicate the color of the bracts, stamens, etc. as scored in "660. Inflorescence attractive organ". Single flowers can be considered as one-flowered inflorescences. If different colours exist, indicate either the dominant colour (if applicable) or each colour separately.

Note: for literature cited in Schönenberger et al. (2020), see that publication for full references.

2 Appendix S2: Supplementary figures

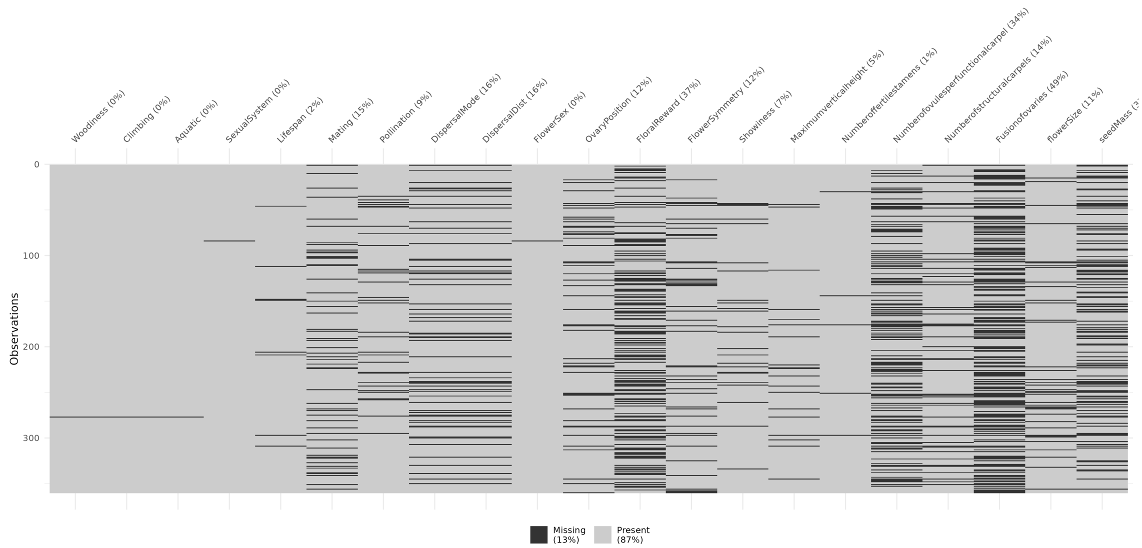

Figure S1: The amount of missing data per trait after the dataset has been cleaned. Traits are in columns and each line is a species in our data set. Black lines in each column indicate where data are missing.

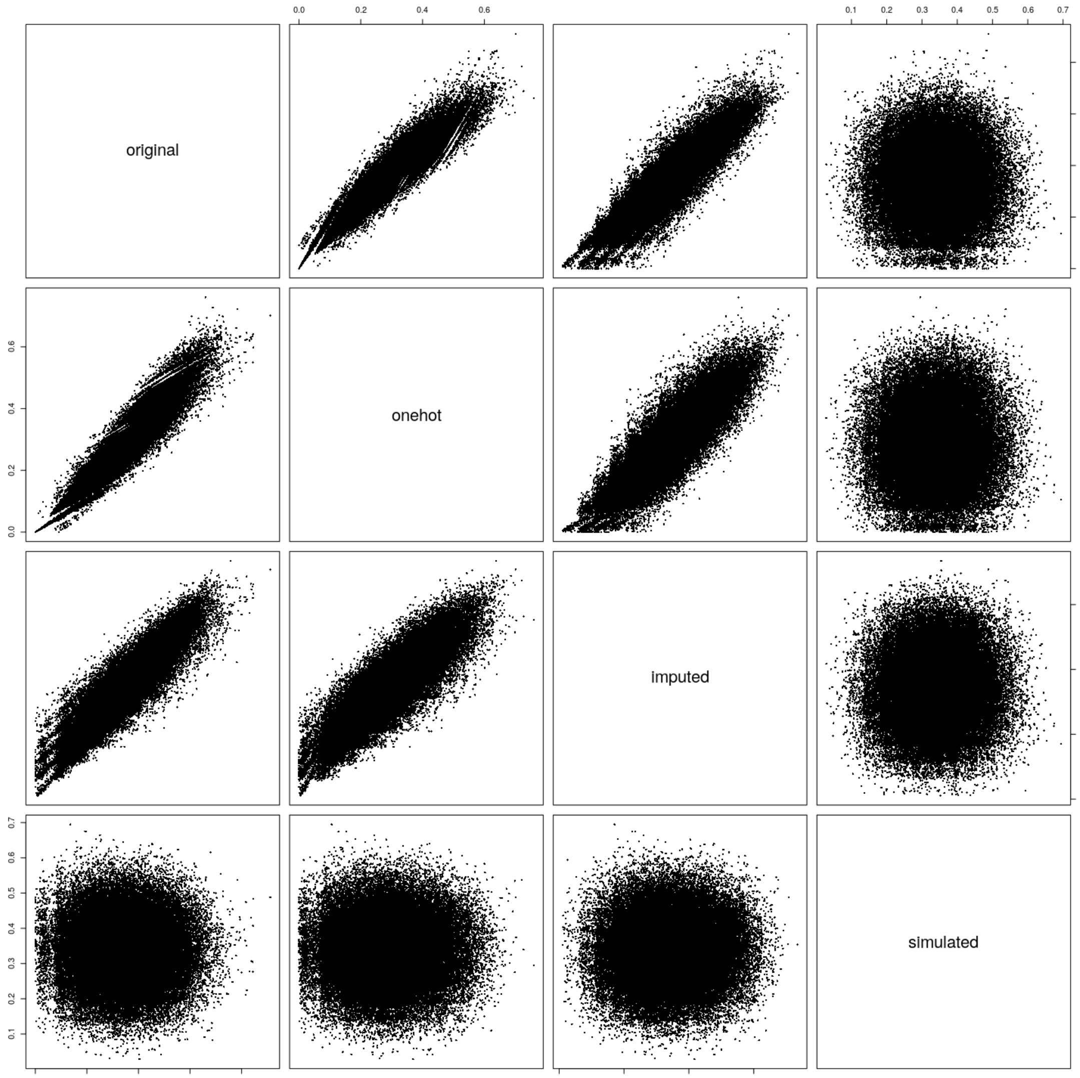

Figure S2: The relationships between different pairwise distance matrices. Along the diagonal are labels indicating each different parwise distance matrix used in this paper, and whether it is on the X or Y axis of each scatterplot. These were made from different data sets, from the top left: the original data set, a one-hot encoded data set, the original data set where missing data were imputed and finally a data set that was simulated using a phylogenetic tree and models of trait evolution (see the methods of the main text for further details).

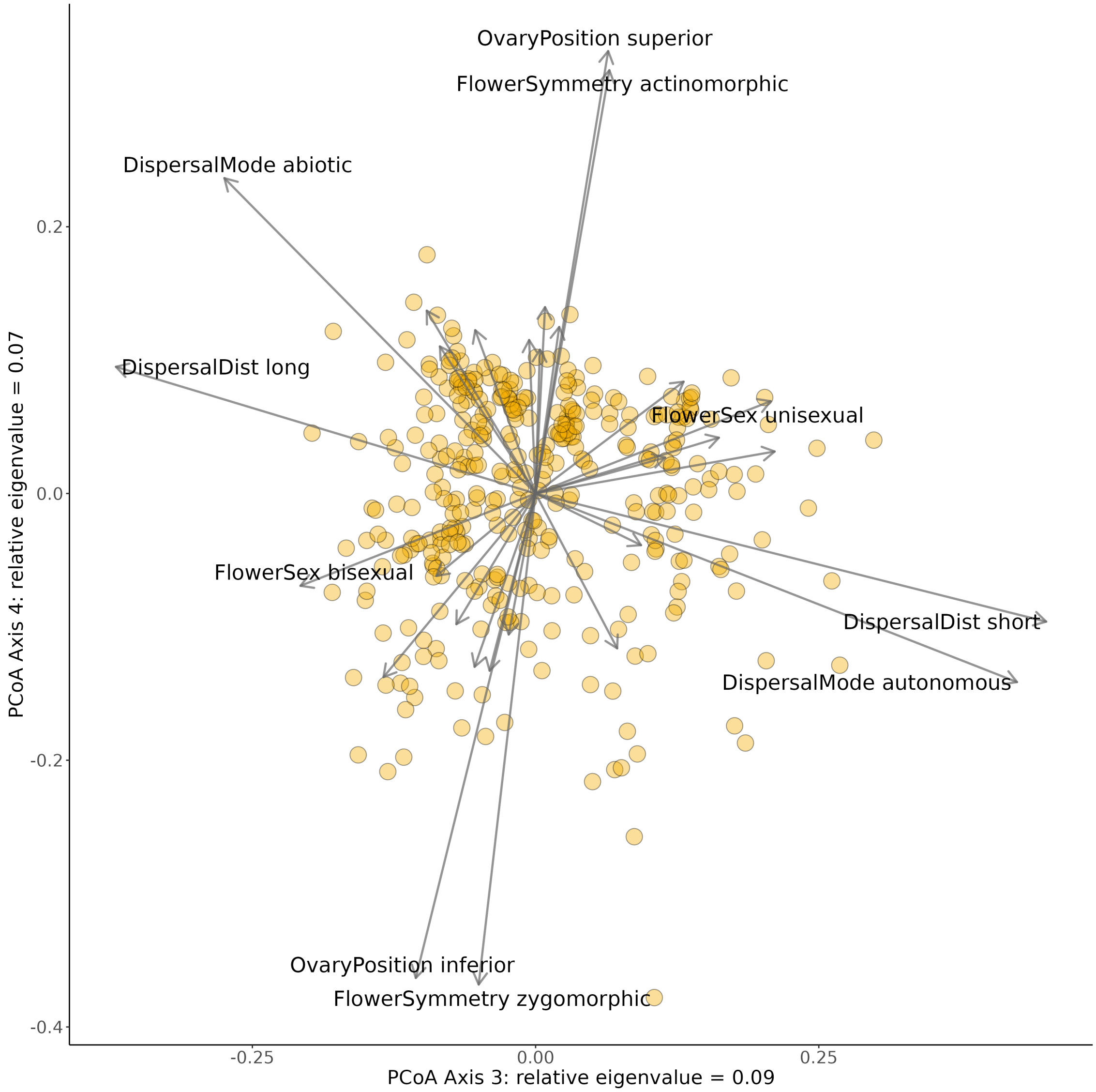

Figure S3: Third and fourth axes of the principal coordinate analysis on our one-hot encoded data set. These axes represent approximately 16% of variation. Using one-hot encoding allowed us to plot arrows indicating how different states and traits affect the trait space, as well as which traits are acting in a similar manner. The length of the arrows indicates the strength of effect of the state and labels were added to arrows representing those states with large effects.

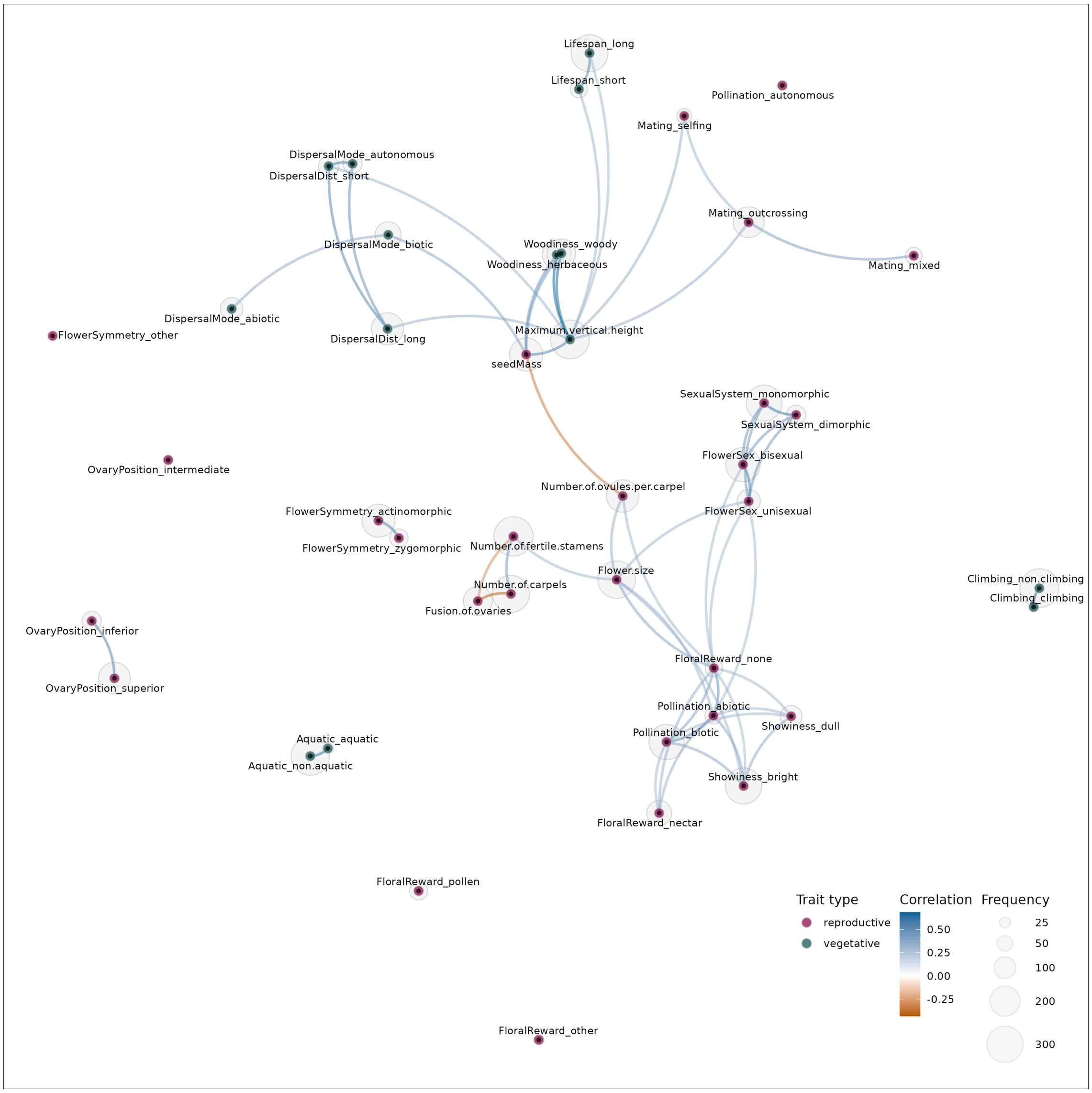

Figure S4: A correlation network of the relationships between the different traits in our one-hot data set. Each point represents the location of a trait on two axes from an Non-metric Multi-dimensional Scaling (NMDS) analysis. The colored circle around the point indicates whether the trait is reproductive or vegetative and the size of the grey circle indicates the number of species with that state in the data set. Correlations between two qualitative variables were calculated using χ^2 tests, between two quantitative variables with Pearson's correlation coefficient and qualitative vs quantitative using analysis of variance (ANOVA). Lines between points indicate a correlation with an absolute value > 0.3, with colours indicating the strength and direction of this correlation. Note that only Pearson's correlation coefficient (between quantitative variables) allows negative values.

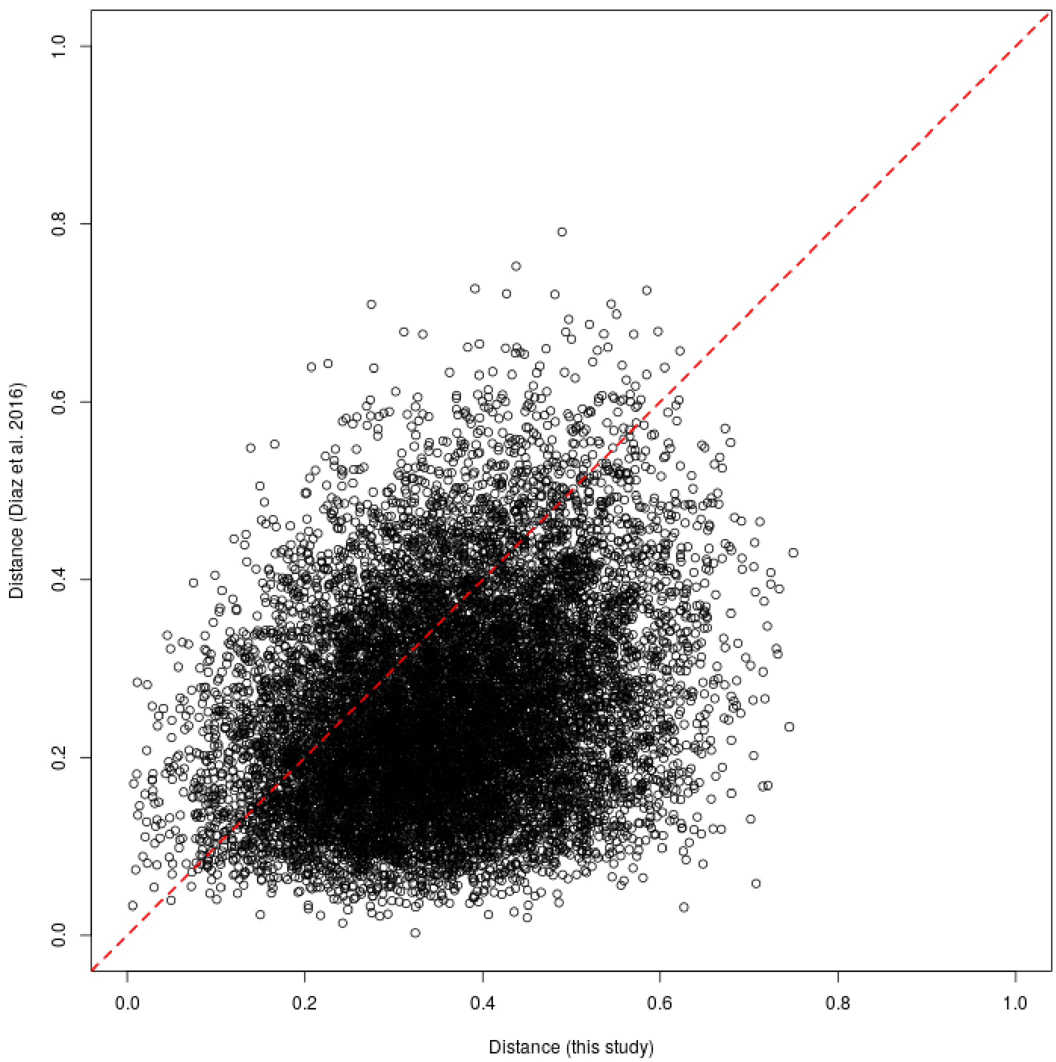

Figure S5: Gower's distances calculated for our data set vs those calculated for a data set derived from Díaz et al., 2016. A total of 159 species were shared across these two data sets and used to make this plot. The dotted red line represents the identity line, which indicates that distances were generally larger in our data set.

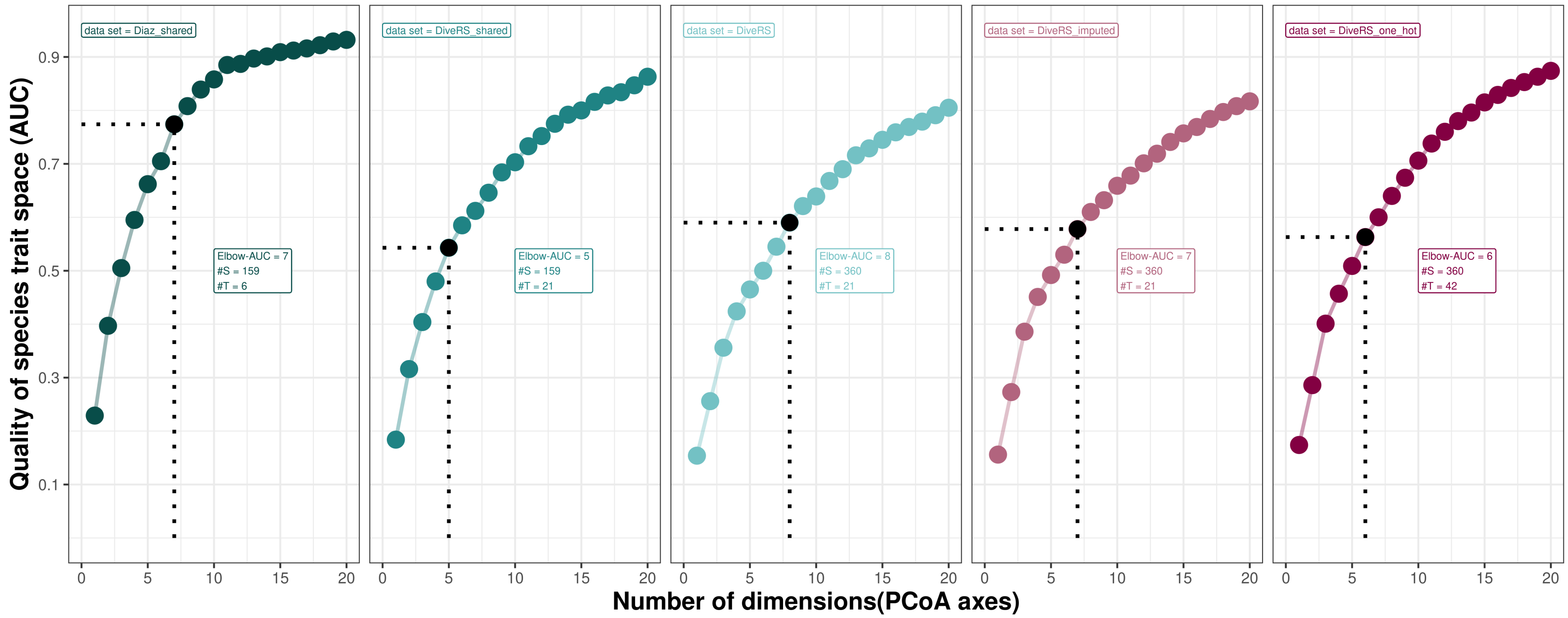

Figure S6: Comparisons of trait space quality (measured using AUC) among data sets, indicated in the top left of each panel: ‘Diaz shared’ contains species present in both Díaz et al., 2022 and our dataset, using the traits of Díaz et al., 2016; ‘DiveRS shared’: same as previous, but using the traits from the current study; ‘DiveRS’: original data set used in the current study; ‘DiveRS imputed’: same as ‘DiveRS’ but missing data is replaced by imputed data; ‘DiveRS one hot’: same as ‘DiveRS’, but using one-hot encoding. Dashed lines and the black point 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.

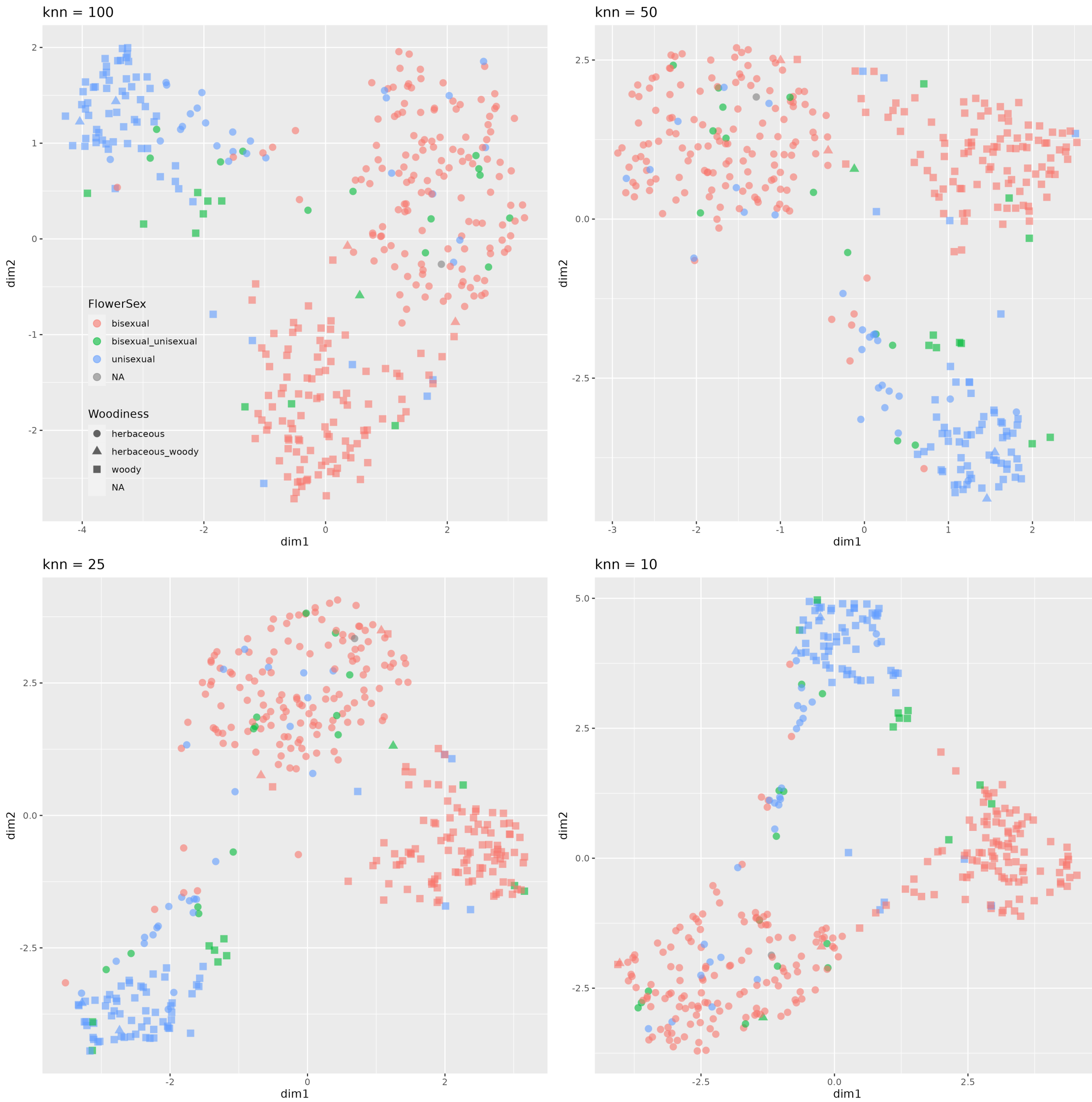

Figure S7: The distribution of species along the first two axes of a UMAP (uniform manifold approximation) analyses. Neighbourhood sizes were varied from 100 to 10, shown in the top left of each plot. Larger values allow for a more global view of the manifold and smaller values mean that local patterns are better preserved. Points are coloured by flower sex and shapes represent the woodiness trait. As neighbourhood size is decreased three major groups more clearly distinguish themselves.

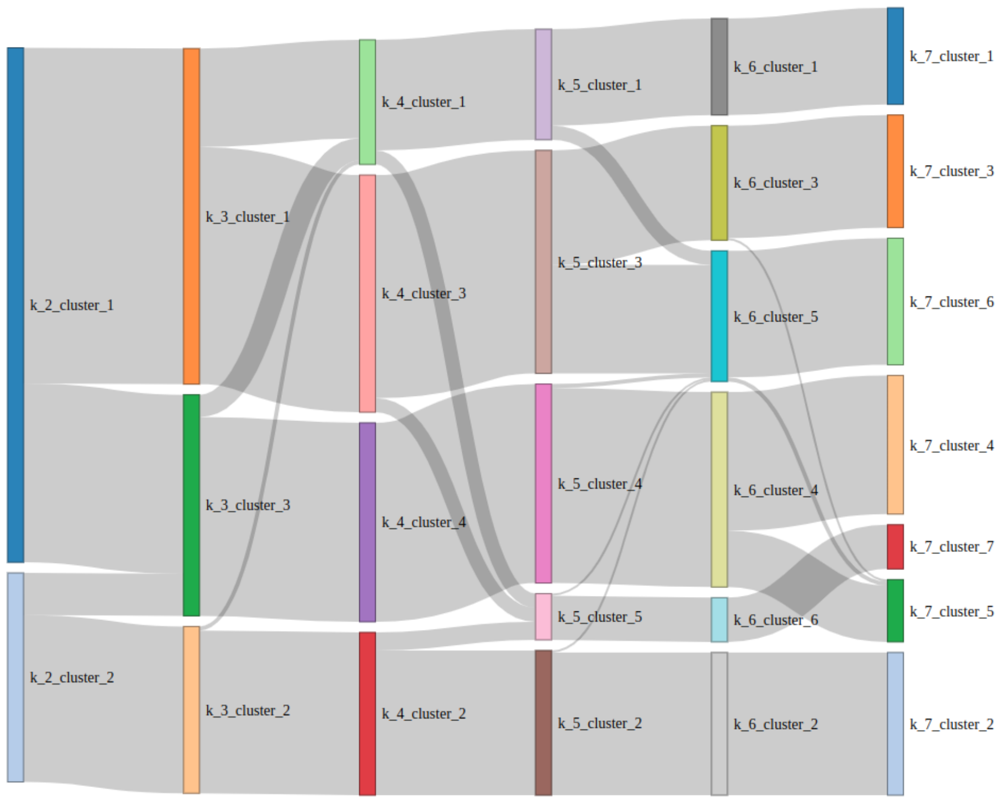

Figure S8: A Sankey plot depicting flows between clusters as values of 'k' (number of clusters) changes when running paritioning around medoids (PAM) clustering. Each set of vertical bars is a cluster (from two to seven clusters) and the grey bands between bars show how species moved between clusters when 'k' is changed. Note that individual species are not represented here, merely the number of species that move from one cluster to another (i.e. one species cannot be tracked through all values of 'k').

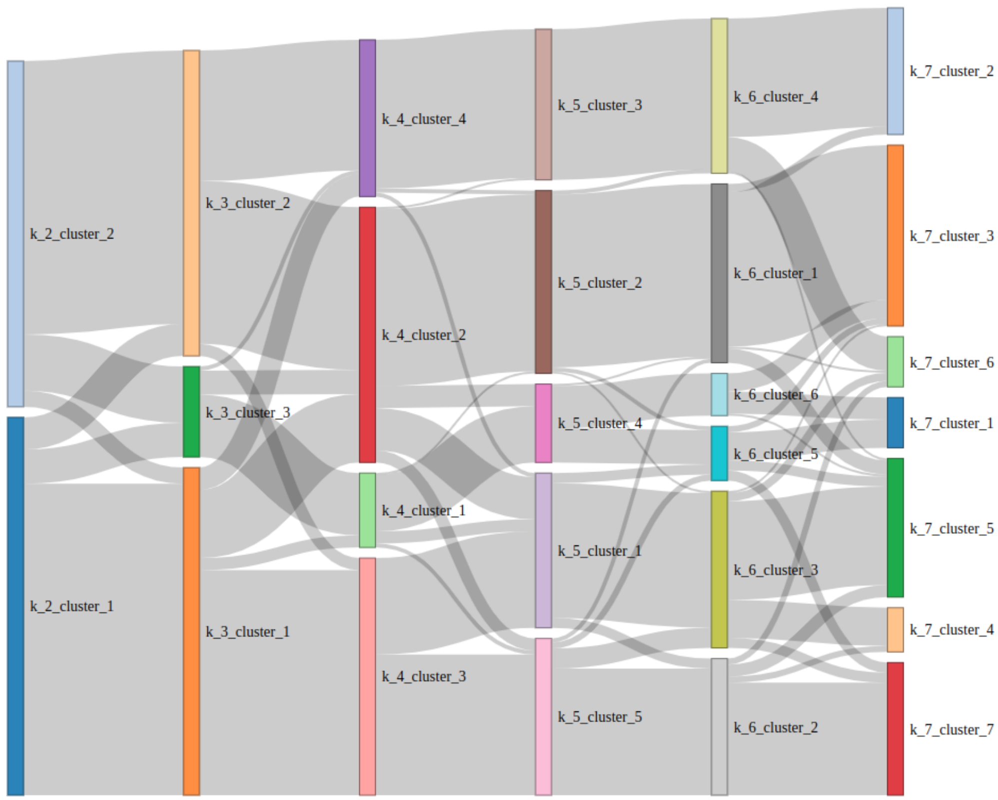

Figure S9: A Sankey plot depicting flows between clusters as values of 'k' (number of clusters) changes when running K-prototypes clustering. For additional information, see Fig. S8.

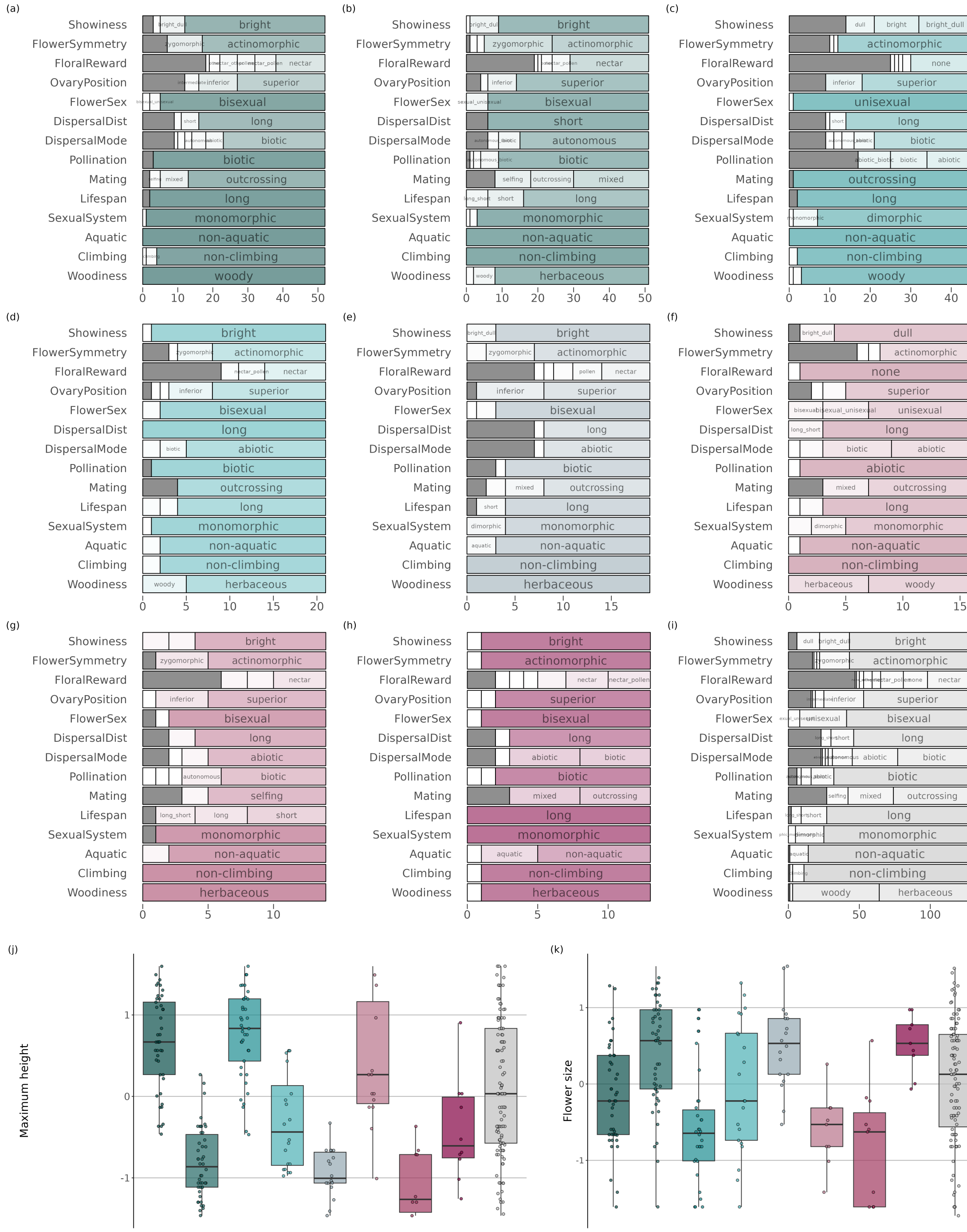

Figure S10: Panels (a-i) depict horizontal stacked barplots of the frequency of states of each qualitative trait in each robust group from K-prototypes clustering analyses. On the X-axis is the number of species in the cluster (and with each state). Dark grey portions of bars indicates missing data. Panels (j) and (k) show the distribution per cluster (after log transformation and scaling) of values for maximum vertical height and flower size respectively.

Figure S11: The location of species on the first two axes of a principal coordinates analysis (PCoA). Points are coloured by cluster membership derived from hierarchical clustering using the ‘wardD2’ method. Ellipses (level = 0.95) represent the area occupied by these clusters. Points are labelled with names of a random subset of species.

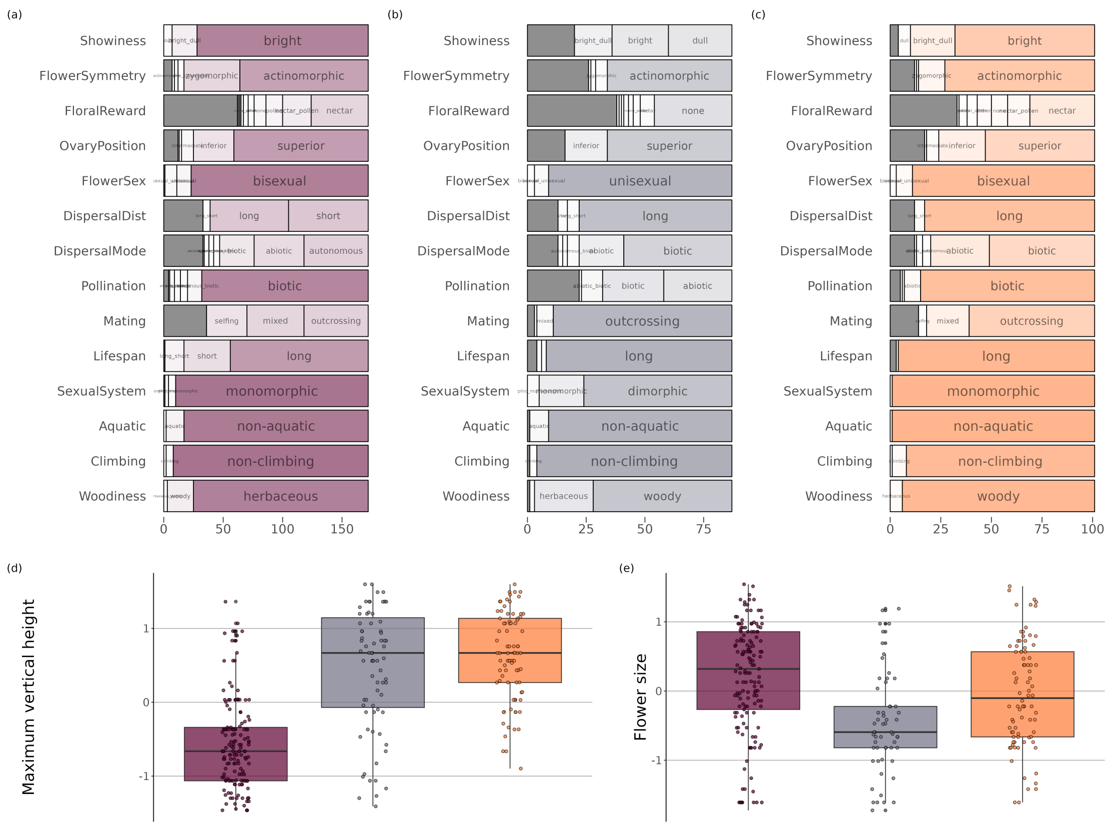

Figure S12: Panels (a-c) show horizontal stacked barplots of the frequency of states of each qualitative trait in each cluster from a hiercharical clustering analysis (k=3, method = 'wardD2'). On the X-axis is the number of species in the cluster (and with each state). Colours correspond to those in Fig. S11 and Fig 1. Dark grey portions of bars indicates missing data. Panels (d) and (e) show the distribution per cluster (after log transformation and scaling) of values for maximum vertical height and flower size respectively.

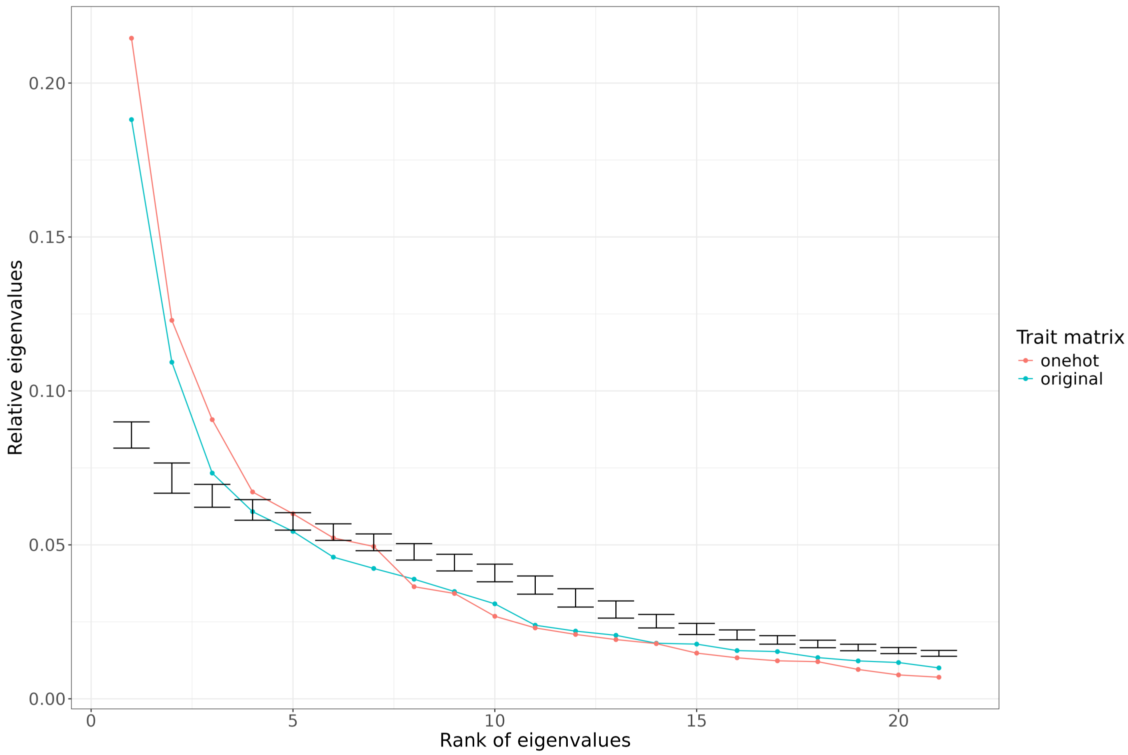

Figure S13: Simulated vs observed eigenvalues from principal cooordinate analyses (PCoA). The black error bars indicate the range of relative eigenvalues for each PCoA axis from 1000 simulated data sets. Blue and red lines show the empirical realtive eigenvalues from our original and one-hot data sets respectively.

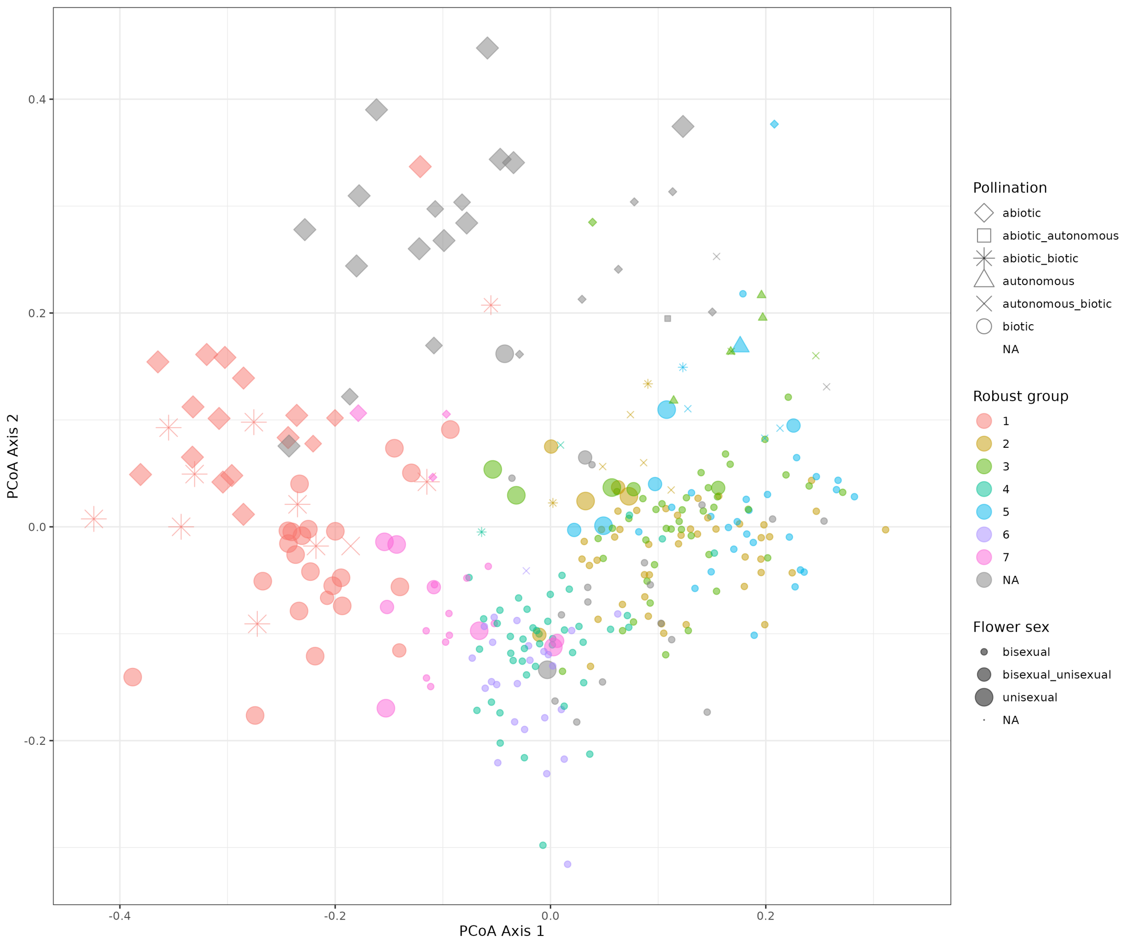

Figure S14: The location of species on the first two axes of a prinipal coordinates analysis (PCoA) on the original data set. Shapes indicate the states for the pollination trait, the flower sex trait is represented by point size and the robust group is represented by point colour.

3 Appendix S3: Supplementary tables

Table S1: Plant traits included in this study. The traits actually used in the analyses are presented, i.e. after simplification for categorical traits and removing traits and species with more than 50% missing data. The original trait is the trait recorded in PROTEUS according to the scoring instructions available in Appendix S1, except for seed mass. For numerical traits, the ranges (minimum - maximum) are given before transformation, with their units if applicable. Note that averages are used per species so count data (numbers) are not necessarily integers.

Plant part	Trait	Type (states or ranges [unit])	Original trait(s)
Entire plant:	Woodiness	categorical (woody, not woody)	Habit
vegetative	Aquatic	categorical (aquatic, not aquatic)	Habit
	Climbing	categorical (climbing, not climbing)	Habit
	Lifespan	categorical (long, short)	Habit, Life history
	Maximum vertical height	numerical (0.01 - 50 m)	Maximum vertical height
Entire plant:	Sexual system	categorical (monomorphic, dimorphic)	Plant sexual system
reproductive	Mating	categorical (selfing, mixed, outcrossing)	Phenotypic mating system, Self-incompatibility system, Outcrossing rate, Plant sexual system[1]
Entire flower	Floral reward	categorical (pollen, nectar, other, none)	Floral reward
	Flower sex	categorical (unisexual, bisexual)	Floral structural sex
	Flower size	numerical (0.08 - 80 cm)	Flower length, Flower diameter[2]
	Flower symmetry	categorical (actinomorphic, zygomorphic, other)	Symmetry of perianth
	Pollination	categorical (biotic, abiotic, autonomous)	Pollination syndrome
	Showiness	categorical (bright, dull)	Main color of perianth at anthesis, Inflorescence attractive organ
	Ovary position	categorical (superior, intermediate, inferior)	Ovary position
Female organs	Fusion of ovaries	numerical (0 - 1)	Fusion of ovaries
	Number of ovules per carpel	numerical (0.5 - 940)	Number of ovules per functional carpel
	Number of structural carpels	numerical (1 - 115)	Number of structural carpels
Male organs	Number of fertile stamens	numerical (0.5 - 425)	Number of fertile stamens
Dispersal	Dispersal distance	categorical (long, short)	Dispersal syndrome
	Dispersal mode	categorical (biotic, abiotic, autonomous)	Dispersal syndrome
	Seed mass	numerical (0.001 - 32835 g)	Seed mass

[1]For the original traits 'Self-incompatibility system' and 'Plant sexual system', only the stricltly outcrossing states (self-incompatibility and dioecy) were used to determine the mating system

[2]If both were available, the largest dimension was used.

Table S2: The number of species per order used in this study.

Order	Frequency	Order	Frequency
Malpighiales	28	Dioscoreales	3
Caryophyllales	25	Pandanales	3
Lamiales	25	Piperales	3
Ericales	23	Proteales	3
Asterales	16	Boraginales	2
Brassicales	15	Cardiopteridales	2
Asparagales	14	Celastrales	2
Poales	13	Dipsacales	2
Malvales	12	Geraniales	2
Alismatales	11	Zygophyllales	2
Ranunculales	10	Acorales	1
Myrtales	9	Amborellales	1
Rosales	9	Aquifoliales	1
Sapindales	9	Arecales	1
Solanales	9	Austrobaileyales	1
Apiales	8	Berberidopsidales	1
Gentianales	8	Bruniales	1
Saxifragales	8	Buxales	1
Fagales	7	Canellales	1
Santalales	7	Ceratophyllales	1
Cucurbitales	6	Chloranthales	1
Laurales	6	Crossosomatales	1
Zingiberales	6	Dilleniales	1
Liliales	5	Huerteales	1
Magnoliales	5	Icacinales	1
Nymphaeales	5	Metteniusales	1
Commelinales	4	Paracryphiales	1
Cornales	4	Petrosaviales	1
Fabales	4	Picramniales	1
Oxalidales	4	Trochodendrales	1
		Vitales	1

Table S3: The correlation of each trait (original data set) with the first four principal coordinate analysis (PCoA) axes. Correlations of quantitative traits are represented by Pearson's coefficient and those of qualitative traits were estimated using analysis of variance (ANOVA).

Trait	Axis.1	Axis.2	Axis.3	Axis.4
Aquatic	0.014	0.145	0.061	0.031
Climbing	0.004	0.022	0.01	0.008
DispersalDist	0.218	0.022	0.635	0.011
DispersalMode	0.264	0.1	0.664	0.047
FloralReward	0.351	0.526	0.123	0.032
FlowerSex	0.546	0.119	0.051	0.083
flowerSize	0.267	-0.406	0.022	-0.11
FlowerSymmetry	0.136	0.016	0.018	0.392
Fusionofovaries	0.02	-0.151	-0.05	-0.101
Lifespan	0.21	0.121	0	0.005
Mating	0.348	0.086	0.042	0.059
Maximumverticalheight	-0.63	-0.367	0.095	0.137
Numberoffertilestamens	-0.062	-0.344	-0.032	0.236
Numberofovulesperfunctionalcarpel	0.381	-0.232	-0.05	0.048
Numberofstructuralcarpels	0.032	-0.209	-0.054	0.138
OvaryPosition	0.01	0.071	0.171	0.286
Pollination	0.263	0.54	0.076	0.049
seedMass	-0.456	-0.282	0.036	0.255
SexualSystem	0.465	0.06	0.079	0.062
Showiness	0.18	0.331	0.046	0.129
Woodiness	0.432	0.261	0.007	0.06

Table S4: Values of a range of functional diversity indices for the three clusters identified using the hierarchical clustering (method = 'wardD2'). 'FDis' = functional dispersion, 'FMPD' = functional mean pairwise distance, 'FNND' = functional nearest-neighbour distance, 'FEve' = functional evenness, 'FRic' = functional richness, 'FDiv' = functional divergence, 'FOri' = functional originality, 'FSpe' = functional specialization. Further information on each of these functional indices can be found in the mFD package (Magneville et al., 2022).

	Species richness	FDis	FMPD	FNND	FEve	FRic	FDiv	FOri	FSpe
Cluster 1	172	0.417	0.424	0.229	0.785	0.284	0.781	0.22	0.401
Cluster 2	87	0.47	0.476	0.288	0.771	0.283	0.742	0.273	0.564
Cluster 3	101	0.304	0.32	0.197	0.782	0.086	0.694	0.183	0.336

Table S5: Phylogenetic signal and associated p-values calculated using δ (Borges et al., 2019) for each one-hot qualitative trait.

	deltas	pvals
Woodiness_herbaceous	7.440	0
Woodiness_woody	6.317	0
Climbing_climbing	1.481	0.52
Climbing_non.climbing	0.956	0.77
Aquatic_aquatic	179.441	0
Aquatic_non.aquatic	9.926	0
SexualSystem_dimorphic	1.779	0
SexualSystem_monomorphic	1.382	0.1
Lifespan_long	1.629	0.44
Lifespan_short	3.810	0
Mating_mixed	0.410	0.81
Mating_outcrossing	0.245	0.87
Mating_selfing	1.976	0
Pollination_abiotic	30.489	0.01
Pollination_autonomous	3.141	0.19
Pollination_biotic	24.268	0
DispersalMode_abiotic	0.347	0.44
DispersalMode_autonomous	0.727	0.7
DispersalMode_biotic	0.020	0.47
DispersalDist_long	0.362	0.81
DispersalDist_short	0.562	0.67
FlowerSex_bisexual	3.013	0
FlowerSex_unisexual	2.773	0
OvaryPosition_inferior	3.196	0
OvaryPosition_intermediate	3.196	0.23
OvaryPosition_superior	0.562	0.65
FloralReward_nectar	4.829	0
FloralReward_none	17.342	0
FloralReward_other	2.516	0.08
FloralReward_pollen	0.356	0.61
FlowerSymmetry_actinomorphic	5.432	0
FlowerSymmetry_other	2.117	0.35
FlowerSymmetry_zygomorphic	7.947	0
Showiness_bright	2.048	0.34
Showiness_dull	2.504	0

Table S6: Two different measures of phylogenetic signal for quantitative traits and their associated p-values.

trait	lambda	K	p_lambda	p_K
Maximum.vertical.height	0.775	0.737	0	0.001
Number.of.fertile.stamens	0.915	0.904	0	0.001
Number.of.ovules.per.carpel	0.599	0.684	0	0.003
Number.of.carpels	0.947	0.937	0	0.001
Fusion.of.ovaries	1.085	1.943	0	0.001
Flower.size	0.744	0.739	0	0.001
seedMass	0.762	0.765	0	0.001

Table S7: Transition rates for each one-hot encoded trait and the size of the tree used to estimate these rates.

states	rates (transitions/lineage/million years)	tree_sizes
Woodiness_herbaceous	0.00415	359
Woodiness_woody	0.00435	359
Climbing_climbing	0.00072	359
Climbing_non.climbing	0.00060	359
Aquatic_aquatic	0.00056	359
Aquatic_non.aquatic	0.00056	359
SexualSystem_dimorphic	0.00359	359
SexualSystem_monomorphic	0.00302	359
Lifespan_long	0.00159	352
Lifespan_short	0.00209	352
Mating_mixed	0.00300	307
Mating_outcrossing	0.00584	307
Mating_selfing	0.00189	307
Pollination_abiotic	0.00181	329
Pollination_autonomous	0.00075	329
Pollination_biotic	0.00138	329
DispersalMode_abiotic	0.00672	302
DispersalMode_autonomous	0.00429	302
DispersalMode_biotic	0.01417	302
DispersalDist_long	0.00390	302
DispersalDist_short	0.00521	302
FlowerSex_bisexual	0.00370	359
FlowerSex_unisexual	0.00515	359
OvaryPosition_inferior	0.00192	315
OvaryPosition_intermediate	0.00082	315
OvaryPosition_superior	0.00250	315
FloralReward_nectar	0.00396	227
FloralReward_none	0.00303	227
FloralReward_other	0.00119	227
FloralReward_pollen	0.00492	227
FlowerSymmetry_actinomorphic	0.00247	315
FlowerSymmetry_other	0.00045	315
FlowerSymmetry_zygomorphic	0.00251	315
Showiness_bright	0.00143	336
Showiness_dull	0.00463	336

References

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

David, P., Pujol, B., Viard, F., Castella, V., & Goudet, J. (2007). Reliable selfing rate estimates from imperfect population genetic data. *Mol. Ecol.*, *16*(12), 2474–2487. https://doi.org/10.1111/j.1365-294X.2007.03330.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

Gao, H., Williamson, S., & Bustamante, C. D. (2007). A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, *176*(3), 1635–1651. https://doi.org/10.1534/genetics.107.072371

Koelling, V. A., Monnahan, P. J., & Kelly, J. K. (2012). A Bayesian method for the joint estimation of outcrossing rate and inbreeding depression. *Heredity (Edinb).*, *109*(6), 393–400. https://doi.org/10.1038/hdy.2012.58

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

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

Perez-Harguindeguy, N., Diaz, S., Garnier, E., Lavorel, S., Poorter, H., et al. (2016). New handbook for standardised measurement of plant functional traits worldwide (vol 61, pg 167, 2013). *Aust. J. Bot.*, *64*(7-8), 715–716. https://doi.org/10.1071/BT12225_CO

Ritland, K. (2002). Extensions of models for the estimation of mating systems using n independent loci. *Heredity (Edinb).*, *88*(4), 221–228. https://doi.org/10.1038/sj.hdy.6800029

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. *Am. J. Bot.*, *107*(10), 1433–1448. https://doi.org/10.1002/ajb2.1538

Simpson, B. B., & Neff, J. L. (1981). Floral rewards - alternatives to pollen and nectar. *Ann. Mo. Bot. Gard.*, *68*(2), 301–322. https://doi.org/10.2307/2398800