Island colonization in flowering plants is determined by the interplay of breeding system, lifespan, floral symmetry, and arrival opportunity

Annie N. Zell[1] (ID) , Charlotte H. Miranda[1] (ID) , Erin L. Grady[1] (ID) , Dena L. Grossenbacher[1,†] (ID) ,
and Boris Igić[2,‡] (ID)

[1]Department of Biology, California Polytechnic State University, San Luis Obispo, CA 93407, U.S.A.

[2]Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, U.S.A.

Authors for correspondence:

[†]*dgrossen@calpoly.edu*

[‡]*boris@uic.edu*

(Last compiled on October 12, 2024.)

Running head: Predictors of island colonization

Keywords: breeding systems, Baker's law, island biogeography, life history traits, floral symmetry.

Summary

- Among flowering plants, self-compatibility, longer lifespan, and generalized pollination syndrome are each thought to increase the lifetime odds of finding a mate, particularly in isolated locales. An accumulated body of evidence supports the role of breeding system in island colonization, but less is known about the impact of other traits and their interactions during establishment.

- We employ a global dataset of 3,222 flowering plant species from 169 families to estimate the effects of traits on the probability of island occurrence. Our analyses additionally account for taxonomic group membership and assess the role of island arrival opportunity.

- Self-compatibility is strongly associated with island colonization. A longer lifespan and generalized pollination syndrome are also associated with increased island colonization, although this is influenced by their interaction with breeding system. The probability of island colonization is highly dependent on taxonomically conserved unmeasured traits and arrival opportunity. As expected, mate limitation appears to increase with dispersal distance, but perhaps at a finer scale than we examine here.

1

- We find that arrival opportunity and breeding system are the primary drivers of island colonization relative to other life history traits we account for here, lending additional support for the positive role of uniparental reproduction in establishment following long-distance dispersal.

Introduction

Long-distance dispersal of organisms to oceanic islands represents a dramatic ongoing natural experiment. Islands may contain differentiated and non-random assemblages of species in comparison to a nearby mainland. Understanding which factors govern the composition of these insular assemblages, why species vary in their colonization ability, and how particular traits or ecological agents of selection impact species distributions are each among the fundamental and long-standing questions in evolution and ecology (Darwin 1859; Wallace 1880; Baker 1955; MacArthur and Wilson 1967). Consequently, numerous studies have sought to identify morphological, physiological, and life history traits that affect the propensity for dispersal and establishment at a variety of spatial and temporal scales (Wallace 1880; Baker 1965; Barrett 2016; Carvajal-Endara *et al.* 2017).

Island colonization can be conceptually divided into three sequential stages: dispersal/arrival, establishment, and subsequent evolution. Following dispersal but preceding establishment and subsequent adaptation, viable and fecund migrants are vulnerable to reproductive failure due to limited mating opportunities. The risk is exaggerated for plants that cannot fertilize themselves and require access to other mates (Baker 1955). It is instructive to consider an extreme example: the fate of a single seed of a sexually reproducing plant species, freshly dispersed to a barren island so distant that it is unlikely to encounter additional conspecific individuals. In addition to the usual vagaries of a new setting—e.g., adequate conditions for germination and growth—the absence of potential mates ensures that only those individuals capable of uniparental reproduction may successfully establish a population. Herbert Baker (1955) observed that any colonization scenario involving "long-distance dispersal" ensures a relative advantage for individuals capable of self-fertilization. Variants of this intuitive observation had been previously recorded by Federico Delpino (Darwin 1876) and many others (e.g., Baker 1948; Dobzhansky 1950), but it was Baker's brief papers (Baker 1955, 1967) that popularized the key ideas and inspired the moniker "Baker's law" (Stebbins 1957). Baker's law is perhaps best interpreted as a probabilistic statement focused on a filtering process which selectively favors propagules capable of uniparental reproduction in the course of colonization (Pannell *et al.* 2015). It is a special case of reduced population growth at low densities ("Allee effect"), specifically due to declines in reproductive fitness. Recent studies suggest that it is likely common in the initial phase of island conquest by plants (Grossenbacher *et al.* 2017; Razanajatovo *et al.* 2019), despite a raft of practical complications in assessing its magnitude (Baker 1965; Carlquist 1974; Pannell 2015). Its effect is likely context-dependent, influenced by

species-specific trait correlations and interactions, and perhaps associated with stochastic error for any given island species assemblage.

The magnitude and variance of breeding system filtering may be affected by any factors that alter the strength of mate limitation, as well as other requirements of successful colonization (Pannell *et al.* 2015). For instance, long lifespans could ease mate limitation by increasing the probability of reproductive overlap among arriving individuals, while generalized pollination syndromes could alleviate it by enhancing pollen movement and deposition by island fauna (Wallace 1878, 1880; Baker 1967). Likewise, the rate of dispersal and the number of simultaneously dispersed individuals within a dispersal unit—usually measured as seeds per fruit and collectively referred to as "propagule pressure"—may introduce an autocorrelation among arriving migrant seed probabilities (Baker 1967; Simberloff 2009). Changes in propagule pressure can alter the population density and consequently influence the relative colonization success of selfing and outcrossing species (Carlquist 1974). The rate of successful dispersal is itself a function of island distance, as well as a variety of abiotic and biotic factors. Baker (1955) conditioned the filtering effect on "long-distance" dispersal, which he saw as a vague proxy for the limited number of migrant arrivals per generation. Few studies have attempted to disentangle the filtering of breeding systems from correlated and confounding variables in the context of Baker's law (Pannell 2015; see Razanajatovo *et al.* 2019). Integrating a range of traits with island biogeography may improve our understanding of the factors that generally influence organismal distributions, and especially those that affect island colonization following long-distance dispersal. Moreover, a clearer focus on the effects of other factors may help discern the extent to which similar dynamics operate in a wide variety of contexts, such as biological invasions, metapopulation dynamics, and natural range expansions (Pannell and Barrett 1998; Cheptou 2012; Pannell *et al.* 2015).

Here, we examine if and how a variety of factors alter the strength of breeding system filtering in the course of island colonization. Specifically, we employ a large and carefully curated dataset of reproductive traits, as well as global geographic occurrence data, to assess whether species with traits that decrease mate limitation are disproportionately represented on islands compared to mainlands. We use generalized linear mixed models to test how specific trait combinations influence the likelihood of island occurrence, while partially controlling for taxonomic relationships. We expect that traits such as longer lifespans and generalized pollination syndromes, which ease mate limitation, may reduce selection pressure on propagules with the potential for uniparental reproduction (self-compatibility, SC) during the initial phase of island establishment. The predictive

value of certain traits for island occurrence may likewise scale with island arrival opportunity, which we infer from the distance of mainland species' ranges to the nearest island. Including these key traits and geographic metrics in our models clarifies the processes that govern island colonization
85 and global patterns of species distributions.

Materials and Methods

Data Collection, Taxonomy, and Classification

We assembled a large database of plant traits (breeding system, lifespan, floral symmetry) and geographic occurrences. It is comprised of previously collected breeding system databases, be-
90 ginning with Raduski *et al.* (2012), which was selectively enlarged by targeting three families in Grossenbacher *et al.* (2017), now merged, and greatly expanded by additional published quantitative breeding system data. For Raduski *et al.* (2012) and Grossenbacher *et al.* (2017), we surveyed the literature (as well as provided some original data) accumulating two types of data: (i) closely examined reports of qualitative estimates of breeding systems, and (ii) quantitative esti-
95 mates with transparent calculation procedures from which relative success of selfed and outcrossed hand-pollinations could be calculated (Index of Self-Incompatibility or ISI). We calculated ISI as the relative fruit set following controlled self- and cross-pollinations:

$$\text{ISI} = 1 - \frac{\text{self-pollination fruit set}}{\text{cross-pollination fruit set}}$$

ISI is an imperfect proxy of SI because many unrelated factors can affect fruit set along with various sources of experimental error and bias, particularly when the reported index instead uses
100 the relative seed set (detailed in Raduski *et al.* 2012). Although it is strongly bimodal, species with consistently intermediate values of ISI do occur (Raduski *et al.* 2012). To facilitate their inclusion in the dataset, we discretized them using a traditionally employed cut-off. SI taxa were defined by having ISI$\geq$0.8, and species with ISI$<$0.8 were considered SC. This conservative cut-off may lump SC species expressing inbreeding depression with SI species. Therefore, as in Grossenbacher *et al.*
105 (2017), we conducted analyses using a stricter threshold (ISI$\geq$0.9), which re-assigns 6% of species in range 0.8$\leq$ISI$<$0.9 to SI, but did not significantly influence any of our main findings (Figure S1).

We also expanded the focus on new reproductive traits, lifespan and floral symmetry, which may influence propensity for island arrival and establishment. We conducted taxonomic resolution to ensure that species-trait observations are not confused by taxonomic changes, For each datum,

5

we recorded the associated taxon name, as provided in the original source, and used Taxonomic Name Resolution Service (Boyle *et al.* 2013, TNRS v4.1) to resolve binomials. We recorded data on lifespan and floral symmetry following searches in a variety of topical references (POWO 2023; eFloras 2023; EOL 2023; FNA 2023; GBIF 2023; SolSource 2023). We filled any remaining missing data using scanned herbarium sheets and botanical literature (Azani *et al.* 2017; Funk *et al.* 2009; Kubitzki 2004, 2007, 2014; Simpson 2019; Stern 2014). Lastly, we obtained species-level data for each trait from the TRY database (Kattge *et al.* 2020), the World Checklist of Vascular Plants (accessed via POWO 2023), and Botanical Information and Ecology Network (BIEN; Maitner *et al.* 2018; `https://bien.nceas.ucsb.edu/bien`), and compared them to our manually collected trait data to flag any discrepancies.

We recorded "lifespan" as an imperfect binary proxy for variable lifetime reproductive bouts. The basic rationale is that many reproductive bouts may ease mate limitation on islands by increasing the probability of arrival and overlap of reproductively compatible individuals over a lifetime. Species were placed in three lifespan categories: annual, biennial or perennial. For all analyses, annual and/or biennial species were lumped to represent shorter lifespan, and we refer to them as "annuals," whereas perennials represent longer, albeit highly variable, lifespan (scored as "perennials").

We collected floral symmetry data as a proxy for generalized and specialized pollination syndromes, because radial flowers are associated with a greater diversity of floral visitors relative to bilateral flowers (Yoder *et al.* 2020). Species with radially symmetric flowers, indicative of a more generalized pollination syndrome, may have an advantage when establishing on islands, where the functional diversity of pollinators is typically lower compared to mainlands (Inoue 1993; Hiraiwa and Ushimaru 2017). Most species floral symmetries were scored as either "bilateral" or "radial." If flowers were borderline bilateral or radially symmetric, we scored symmetry as "slightly bilateral" (e.g., radially symmetric perianth with bilaterally symmetric androecium or gynoecium). For species with tight clusters of small flowers (e.g., composite capitula), we considered the symmetry of the entire head rather than each individual flower.

Geographic distribution

Following the same classification scheme as Grossenbacher *et al.* (2017), we categorized plant geographical regions as either mainlands or islands using TDWG encoding (Brummit 2001, https://www.tdwg.org/standards/109). TDWG divides the global terrestrial surfaces into ge-

ographic units at four nested levels, with unique codes for each region and nested sub-region. Mainlands were categorized at TDWG Level 1 (North America, South America, Europe, Asia, or Australia). Islands were categorized at a finer resolution, generally corresponding to the TDWG Level 4 (see below).

145   As our aim is to assess factors that influence natural patterns of island establishment in the absence of human-mediated dispersal and colonization, non-native occurrence data were removed. We manually recorded native occurrences for each species in our dataset by checking, in order, World Checklist of Vascular Plants (WCVP; Govaerts *et al.* 2021), USDA-GRIN (Wiersema 2023), and GBIF (GBIF 2023). WCVP defines the *native distribution* as the area where focal species

150 has "been present since before the last Ice Age or arrived by natural colonization" and *introduced distribution* as the area "where plants were introduced accidentally or deliberately by humans." If information in these databases did not match, we also checked the primary taxonomic literature to verify the native range of the species. We checked the accuracy of our island and mainland presence data using the native occurrences and native distribution for each species from BIEN.

155   Dispersal rates likely vary with distance between mainland species' ranges and closest islands. Therefore, we calculated a novel proxy for island arrival opportunity based on the distances between each species' native mainland occurrences and the nearest island. For each species, we report the median of these distances to minimize the effects of skewness or outliers in occurrence data. Island-occurring taxa with high species-island distances likely underwent long-distance dis-

160 persal events, while island-occurring taxa with lower species-island distances underwent relatively shorter dispersal events. Species with extensive ranges, especially along continental coastlines, are closer to islands, affording them greater opportunities for island dispersal compared to species with ranges localized to deep interior areas. We calculated the species-island distance metric using `Generate Near Table` (ArcGIS Pro v3.0.0; Environmental Systems Research Institute, Redlands,

165 California, U.S.A.) to obtain the shortest distance between each species' mainland occurrence record and the coastline of any island (Sayre *et al.* 2019). This metric only represents a rough proxy of island arrival opportunity as it has several potential flaws. Given the biased sampling of species occurrence data, the location where a particular species is most commonly documented may not reflect where it is most common. Mainland-only species with low species-island distances may have

170 certain traits or climatic requirements that prevent successful island establishment, despite their close proximity to islands. Despite these caveats, we included this species-island distance metric in a subset of our models to account for the variable dispersal opportunity between species' mainland

ranges and islands, and to evaluate its effect on the strength of island colonization filtering. Specifically, we expect greater effects of mate limitation with increasing distances between species and islands.

The strength of a colonization filter may also depend on the geological history of an island (Grossenbacher *et al.* 2017). For each designated island group, as defined by TDWG Level 4, we recorded whether the group is a part of continental plate, surmising that such islands may have been intermittently connected to the nearby mainland at some point in recent geological history. Conversely, oceanic islands (commonly of volcanic origin) are unlikely to have had recent land bridges connecting them to mainland regions. We expect stronger colonization filtering for oceanic islands given that taxa on these islands evidently colonized through long-distance dispersal. The detailed island classification scheme is described in Grossenbacher *et al.* (2017).

Statistical analyses

To determine the strength of an island colonization filter imposed by mate limitation, we assessed whether SC species are more likely to occur on islands relative to SI species. We used generalized linear mixed models (GLMM), closely following the approach of Grossenbacher *et al.* (2017). All analyses were performed in R (v4.2.1; R Core Team. 2023). GLMMs were implemented in the `lme4` package (v1.1-30; Bates *et al.* 2015), using `glmer` with a logistic link function.

The initial model, which we then expand upon (Table 1), is the same as in Grossenbacher *et al.* (2017). The categorical response variable was island "presence" (species that occur on both mainlands and islands; including those endemic to islands in some analyses) or "absence" (species recorded as occurring exclusively on mainlands). The predictor in the each model was breeding system (SI or SC). We included plant family or genus as a random effect to account for taxonomic non-independence of observations (species). We replicated the same GLMM analyses to evaluate the robustness of our results with the following modifications. To account for the possibility of secondary trait evolution on islands, Model 1b excludes island endemics. To assess the effects of increased taxonomic resolution, Model 1c includes genus nested within family as a random effect. In Model 1d we increased the cutoff for classifying breeding system by changing the ISI threshold from 0.8 to 0.9. To test whether the geological origin influenced our results for island occurring species, the response variable in Model 1e was changed to oceanic island versus continental island.

For Model 2 through Model 8, we added variables to the model that we hypothesized interact with breeding system to predict island presence. Model 2 includes breeding system interacting with

two traits that we used as a proxy to assess the strength of a mate limitation filter on islands:
lifespan and floral symmetry. To account for island arrival opportunity or variable proximities to
islands across each species' range, Model 3 includes the interaction between breeding system and
species-island distance. Models 4-7 build on Model 3, including or excluding trait interactions with
breeding system and species-island distance. Model 8 includes a three-way interaction between
breeding system, lifespan and species-island distance.

To obtain the probability of island presence for breeding system, as well as interacting
categorical traits, we extracted predicted probabilities and their 95% confidence intervals using the
`emmeans` package (v1.8.11) function `emip` (Lenth 2023). We used the Akaike Information Criterion
(AIC) scores to guide model fitting, comparing models with the `aictab` function in the `aicmodavg`
package (v2.3-3).

We used the `MuMln` package (v1.47.1) function `r.squaredGLMM` to calculate the marginal
R^2 to obtain the proportion of variance explained by only the fixed effects, and conditional R^2
to obtain the proportion of variance explained by both fixed and random effects (Johnson 2014;
Grossenbacher *et al.* 2017). Specifically, to compare the relative magnitudes of each fixed effect
as a predictor in our selected model, Model 5, we estimated the proportion of variance explained
by breeding system, lifespan, floral symmetry, and species-island distance (`partR2` function in the
`partR2` package, v0.9.1; Stoffel *et al.* 2021). Following methods outlined in Stoffel *et al.* (2021), we
pooled breeding system interactions with each main effect.

More detailed statistical methods, data-collecting procedures, and methods employed to
filter trait and geographic occurrence data are described in Supporting Information (Methods S1).

Results

Summary statistics

We assembled a database with breeding system, lifespan, floral symmetry for 3222 species, and
species-island distance for 2931 species. Most of our analyses (Models 1c-5) exclude 217 island
endemics and 74 species with missing occurrence data. The remaining 2931 species are from 1019
genera and 162 families. Of this final total, 785 (27%) species are recorded on both islands and
mainlands, while the remaining 2146 (73%) are exclusively recorded on mainlands. For breeding
system, 1569 (54%) are SC and 1362 (46%) are SI (Fig. 1a). For lifespan, 577 (20%) are annual
or biennial species, while the remaining 2354 are perennial (80%; Fig. 1a). For floral symmetry,
715 species (24%) are bilaterally symmetric, and 2216 (76%) are radially symmetric (Fig. 1a).

235 Additional data summaries, including model parameters and outputs, are presented in Supporting Information (Notes S1, Tables S1-S14).

Model selection

Among comparable Models 1b and 2–8 (Table 1), we find that Model 5 (K = 10 parameters, AIC = 3011) has the lowest AIC and carries 48% of the cumulative model weight. The remainder is
240 distributed among two similar models, with Model 6 accounting for 19% and Model 7 for 32% of the model weight. Model 6, which adds an interaction between floral symmetry and species-island distance, has the highest number of parameters (K = 11; AIC = 3013). Model 7, which omits the breeding system and species-island distance interaction, has fewer parameters (K = 9) than models 5 or 6, but yields a slightly higher AIC value (3012). The four remaining comparable models (1b,
245 2, 3, 8) received very low support. Thus, we selected Model 5 as our best model for downstream display and analyses, noting that the alternative models yield qualitatively similar results (Table S12).

Breeding system

Breeding system is consistently associated with island colonization. Among the 785 species recorded
250 on both islands and mainlands, 498 are SC (63%). GLMMs, which use breeding system to predict island presence while accounting for taxonomic relationships, show that SC species have a higher probability of occurring on islands compared to SI species (0.30 vs. 0.18, respectively; Table 1, Model 1b; Fig. 2a). This relationship remains significant ($p < 0.001$) when including or excluding island endemics, increasing the phylogenetic resolution to the genus level, and changing the breeding
255 system threshold to a more conservative ISI$\geq$0.9 value for classification of SI species (Table 1, Models 1a-d). SC species are also more likely to occur on oceanic islands compared to SI species (z-value = -2.98, $p = 0.003$; Table S12, Model 1e), although this relationship is not statistically significant after a conservative p-value adjustment ($p = 0.15$ with Bonferroni correction; Table 1, Model 1e).

Lifespan

260 Among 785 species that occur on both island and mainland, 30 (4%) are SI-annuals, 159 (20%) are SC-annuals, 257 (33%) are SI-perennials, and 339 (43%) are SC-perennials. In our best model, we find a significant breeding system by lifespan interaction in predicting island colonization (Table 1, Model 5). SC-perennials and SC-annuals are more than three times as likely to occur on islands as SI-annuals (0.25, 0.26 vs. 0.08, respectively; Fig. 2b). Pairwise comparisons show that there is no

significant difference in probabilities of island occurrence between SC-annuals and SC-perennials, meaning lifespan does not significantly affect island colonization among SC plants (z-ratio = 0.66, $p > 0.9$; Table S14). In contrast, SI-perennials have a significantly higher probability of island presence than SI-annuals (0.16 vs. 0.08, respectively; z-ratio = -3.32, $p < 0.005$; Table S14), strongly suggesting that the likelihood of self-incompatible plants occurring on islands depends on their lifespan.

Floral symmetry

Of the 785 species, which occur on both island and mainland, 43 (5%) are SI with bilateral floral symmetry, 93 (12%) are SC with bilateral floral symmetry, 244 (31%) are SI with radial floral symmetry and 405 (52%) are SC with radial floral symmetry. We find there is a significant breeding system and floral symmetry interaction in predicting island colonization (z-value = 2.42, $p = 0.01$; Table S12, Model 5), although not significant after a study-wide Bonferroni correction. SC species with radial floral symmetry are more than twice as likely to occur on islands as species with other trait combinations (SC-bilateral, SI-radial, SI-bilateral; Fig. 2c). Furthermore, pairwise comparisons within SC species reveal a significant difference in island occurrence probabilities for species with bilateral and radial floral symmetry (SC-bilateral vs. SC-radial; z-ratio = -5.46, $p < 0.001$; Table S14). It appears that floral symmetry affects island presence for SC species, but far less than it does for SI species.

Species-Island distance

Species-island distance is a significant predictor of presence on islands ($p < 0.001$, Table 1 Models 4-5). This metric, a proxy for island arrival opportunity, is calculated as the median distance between each species' mainland occurrence and the nearest island (Fig. 1b). First, and most intuitively, as species-island distance increases the probability of species presence on islands decreases (Fig. 3a). In other words, species with mainland ranges that are far away from any islands are less likely to occur on islands. Second, in terms of island colonization, the advantage of SC species increases with species-island distance (Fig. 3b). This suggests that reducing island arrival opportunity increases the relative probability of colonization for SC species, relative to SI species. The island arrival opportunity metric, however, does not have a statistically significant interaction with breeding system (z-value = 1.67, $p = 0.09$, Table 1, Model 5). We find that SC species have an island colonization advantage for all values of species-island distances.

Perhaps surprisingly, our analyses reveal a significant interaction between species-island dis-

tance and lifespan (z-value = 4.05, $p < 0.001$, Table 1, Model 5). The change in the relative advantage of perennials over annuals vs. species-island distance is illustrated in Fig. 3c. At closer species-island distances (below 1000 km) annual species with greater access to islands show a relatively higher probability of island presence, compared with perennials. At species-island distances greater than 1000 km, it is instead the perennial species that have a higher probability of island presence (Fig. 3a,c).

Magnitude of model predictors

We assessed the relative magnitudes of model predictors and their interactions by comparing the proportion of variance explained (marginal R^2) between full models (Table 1) and among individual predictors in a single model (Fig. 4). In Model 1a, which includes the full dataset, breeding system as the sole predictor for island presence explains 4% of the variance (Table 1). In Model 1b, excluding island endemics decreases the percentage of variance explained by breeding system to 3%. In Model 2, breeding system interactions with floral symmetry and lifespan explain 5% of the variance. In Model 3, breeding system interacting with species-island distance explains 10% of the variance. In Model 4, breeding system interacting with floral symmetry, lifespan, and species-island distance explains 13% of the variance. Finally, in Model 5, the same fixed effects as Model 4 plus the significant interaction between lifespan and species-island distance explains 15% of the variance. These values did not change in Models 6 & 7. In Model 8, the percentage explained variance by the three way interaction between breeding system, lifespan and species-island distance decreased to 13%.

The proportion of variance explained by both random and fixed effects (conditional R^2; Table 1) depends on the level of nesting of random effects. In Model 1b, family as a random effect plus breeding system as a fixed effect explains 23% of the variance (Table 1). In Model 1c, changing the random effect to genus nested within family nearly doubles the percentage explained variance to 41% (Table 1). We did not complete downstream genus-level analyses, because species-genus-trait combinations at this resolution yield low data occupancy.

Using our best model, Model 5, we used `partR2` to assess the relative strengths of each main effect lumped with its interactions (Stoffel *et al.* 2021). The marginal R^2 calculated for full model comparisons (Table 1) differs slightly from that calculated for within-model comparisons due to differences in package methods (within the range of the 95% confidence intervals). We found that species-island distance explains approximately 8% of the variance, breeding system 5%, lifespan

2%, and floral symmetry 1% (Fig. 4). Plant family, included as a random effect (conditional R^2; Table 1), explains nearly the same proportion of the variance as the trait predictors combined, pointing to a strong potential role for unmeasured factors.

³³⁰ ## Discussion

A pioneering propagule may find itself growing at a location so isolated from conspecific individuals that it is virtually guaranteed to experience mate limitation once it reaches reproductive maturity. For most flowering plants, the extent of mate limitation is dependent on both the rate of arrival of subsequent propagules, as well as successful pollen transfer. Therefore, during the establishment phase of colonization, obligately outcrossing species are expected to be relatively less successful, disfavored by natural selection (Baker 1955). Previous studies have found that the strength of the breeding system filter is modest (Grossenbacher *et al.* 2017; Razanajatovo *et al.* 2019). In other words, a relatively high proportion of species—likely to have been SC at the time of their arrival on distant islands—succeed in insular settings. Baker, Carlquist, and others were aware that the relative advantages of uniparental reproduction are influenced by other traits or factors, both directly and indirectly (Baker 1953; Carlquist 1966; Baker 1967). For instance, shorter dispersal distances or effective dispersal syndromes elevate island arrival opportunity and may subsequently increase the number of available mates. Other traits, such as long lifespan and reliance on generalist pollinators, may also blunt the disadvantages of obligate outcrossing, but they have primarily been subjects of informed speculation (Wallace 1878; Baker 1953, 1967). The approach we take here examines multiple sources of mate limitation, which potentially affect the strength of the breeding system filtering effect predicted by Baker's law.

Our study yields five major observations. First, despite many possible sources of error, as well as conservatively constructed models and treatment of data chronically underestimating its effects, breeding system is a significant predictor of island colonization. Second, longer lifespan is associated with greater probability of island occurrence among SI species, suggesting that longevity may alleviate the effects of mate limitation. Third, species with radially symmetric flowers are associated with greater probability of island occurrence, implicating generalist pollinators in easing mate limitation. Fourth, species-island distance is significantly associated with island colonization probability and strongly interacts with lifespan, but does not detectably interact with breeding system. This suggests that our metric of island arrival opportunity, however flawed, captures an important aspect of propagule pressure. It also indicates that island arrival opportunity and lifes-

13

pan jointly affect the probability of co-existing reproductive adults and the resulting (absence of) mate limitation. Finally, breeding system and other traits explain a fairly small proportion of the total variance in island colonization in our models. Collectively, these traits explain approximately as much as the random effect attributed to taxonomic family membership. Along with these observations, we discuss the many limitations of our data and approaches, as well their considerable potential.

Breeding system is a modest predictor of island colonization

We find that SC species are more likely than SI species to successfully colonize islands (Fig. 2a), a key prediction of Baker's law. Mate limitation exacted on obligate outcrossers during establishment appears to have significant effects on colonization. Grossenbacher *et al.* (2017) previously assessed the probability of island occurrence for 1560 species in three plant families (Asteraceae, Brassicaceae, and Solanaceae) and also found that SC species were significantly more likely to occur on islands than SI species (probabilities 0.46 vs. 0.18, respectively). Similarly, Razanajatovo *et al.* (2019) expanded the ISI dataset to include 1752 species, including 161 families, and found significant advantages for SC and autogamous species. Across these studies, which included a variety of methods for scoring breeding system and island colonization, the common result is a detectable filtering effect, albeit one with a modest magnitude.

Greater lifespan facilitates island colonization

Relatively greater lifespan and woodiness are well-characterized features of insular floras (Carlquist 1974; Lens *et al.* 2013). We reasoned that uniparental reproduction in combination with (reproductive) longevity may affect mate limitation and the probability of establishment on islands. Consequently, we examined the individual and joint effects of coarsely recorded lifespan, as a binary trait. Our results show that lifespan has consistent positive effects on island colonization among SI but not among SC plants (Fig. 2b). That is, the effect of lifespan depends on breeding system in predicting island presence. One possibility is that uniparental reproduction results in seed production sufficient to ensure population viability, regardless of lifespan.

Conversely, the higher probability of island presence for SI-perennials relative to SI-annuals indicates that multiple reproductive bouts associated with longer lifespan may ease mate limitation for species incapable of uniparental reproduction. This may be caused by two overlapping processes. First, longer lifespan may increase the probability of arrival of additional reproductively compatible propagules over an individual lifetime. Although we do not account for it here, clonal reproduction

14

may also engender similar advantages. Second, the repeated opportunity for reproduction is particularly likely to be beneficial if pollinators are scarce (Wallace 1878). This process may contribute to the commonly observed high prevalence of species with convergent insular woodiness, along with increased competition for light, reduced climactic seasonality or decreased herbivory (Böhle *et al.* 1996; Zizka *et al.* 2022). Although our study is focused on traits of colonizing species rather than those secondarily evolved within island endemics, the results broadly support the possibility that mate limitation is alleviated by longer lifespan. Thus, reduced mate limitation may be an underestimated force, partly responsible for generating the pattern of greater prevalence of perennials among island colonizing plants. This relationship may be further clarified with a fine-scale examination of the effects of longevity, perhaps employing proxies such as woodiness or clonality.

Floral symmetry may increase colonization success by reducing reliance on specific guilds

We find that floral symmetry interacts with breeding system to significantly influence island colonization (Fig. 2c). Specifically, SC species with radially symmetric flowers—commonly associated with generalized pollination—are more than twice as likely to be found on islands than SC species with bilaterally symmetric flowers or SI species with both radially and bilaterally symmetric flowers. Although quantifying the effect of biotic interactions on species distributions is difficult (Wisz *et al.* 2013), our results suggest that pollinators may play a role in shaping island floras. SI species apparently experience mate limitation regardless of the composition of insular pollinator communities. Mate limitation of SC species instead may strongly depend on local pollinator communities. For instance, it is possible that some proportion of SC species are autogamous, while the majority remain severely pollen-limited in the absence of pollinator visits (Raduski *et al.* 2012).

Although it has long been thought that islands with few pollinators may filter out species dependent on specialized animal-mediated pollination (Wallace 1878), few studies have addressed this theory, especially in the context of plant breeding systems. Breeding systems are thought to be linked with pollination syndromes (Darwin 1877; Fenster and Marten-Rodriguez 2007; Wolowski *et al.* 2013), partly because increased selfing can ease mate limitation during pollinator transitions (Wessinger and Kelly 2018). For example, Martén-Rodríguez *et al.* (2015) examined visitation rates and autonomous self-fertilization among island and mainland species in Caribbean Gesneriaceae. They found that specialized hummingbird-pollinated species on islands had lower visitation rates relative to hummingbird-pollinated species on mainlands, which may explain shifts to bat- or mixed-

pollination systems on islands. Similar studies that compare island and mainland populations within species could further clarify the interaction of pollination syndrome and breeding system and their role in island colonization.

Shorter species-island distance increases chances of island occurrence

An intuitive extension of Baker's basic argument dictates that there ought to be a detectable relationship between the extent of geographic isolation, mate limitation, and breeding system filtering. In nearly all of his papers, the consequential dispersal scale is given as "long-distance" (Baker 1955; Carlquist 1966). The phrase is almost always given in quotes, giving room to non-literal interpretation, perhaps hinting that distance is merely correlated with dispersal probability. In other words, as propagule dispersal distance increases beyond the ordinary extent of dispersal and gene flow, it is proportionately likely that the dispersed propagules will be reproductively isolated from potential mates. Alternatively, extreme dispersal distance may be relatively unimportant, because reproductive assurance tips the scales even at short distances in favor of pioneering selfers (Lloyd 1980; Pannell and Barrett 1998). If this is the case, we expect a general association between breeding system and tendency to experience bottlenecks in various contexts (succession, invasion, metapopulation dynamics, variable pollinator service).

We devised a coarse metric of island arrival opportunity, "species-island distance," which is calculated as the median distance to nearest island of all recorded species occurrences. We find that species-island distance is a significant predictor of island presence in our generalized linear mixed models (Table 1 Models 4-8). Intuitively, there is a negative relationship between island arrival opportunity, measured by species-island distance, and the probability of species presence on islands (Fig. 2a; cf. MacArthur and Wilson 1967). The advantage of SC species en route to island colonization increases with the species-island distance (Fig. 3b). That is, the relative probability of colonization by SC species, compared with SI species, is negatively correlated with island arrival opportunity. The species-island distance metric, however, shows only a weak or no interaction with breeding system (Table 1 Models 4-6). While island-occurring species show strong breeding system filtering, it seems that incremental changes in arrival opportunity do not differentially affect SC and SI plants. One interpretation is that even local dispersal—perhaps simply exceeding ordinary pollen travel and not necessitating "long-distance"—is sufficient to cause breeding system filtering. Therefore, dispersal across low distances may yield a large decline in colonization success and drive the differences in island colonization patterns.

Our analyses reveal a significant interaction between species-island distance and lifespan (Table 1, Model 6). At closer species-island distances (below 1000 km), annual species with greater access to islands show a relatively higher probability of island presence, compared with perennials. After ca. 1000 km, it is instead the perennial species that have a higher probability of island presence (Fig. 3a,c). Perennials are seemingly more likely to colonize islands than annuals under conditions of limited arrival opportunity. This interaction highlights the possible role of perennial species' ability to reproduce multiple times, critical in scenarios where rare arrival events restrict mate availability.

Our metric is a coarse proxy for arrival opportunity. Seed dispersal by birds, wind, and ocean currents is not equally likely between all equidistant points around the globe (Kling and Ackerly 2021). Therefore, dispersal probability between different pairs of mainlands and islands is perhaps only weakly correlated with physical distance. Given these and other possible limitations, it would be valuable to explore the properties of our species-island distance metric at different scales, particularly while accounting for dispersal traits and physical factors (wind, currents).

Traits and species-island distance explain less variance than taxonomic group membership

Our analyses include a method of "phylogenetic correction," which relies on nesting species-level observations within taxonomic family or genus, helping remove the effects of uneven sampling across taxonomic groups (Grossenbacher *et al.* 2017). The combined fixed and random effects explain approximately twice as much variance as fixed effects (distance, breeding system, floral symmetry and lifespan) alone. Island colonization may therefore be affected by unmeasured or imprecisely measured traits that are phylogenetically conserved. For instance, taxonomic characters that define a particular genus or family could impact dispersal and establishment.

Confounding factors and biases

The association between breeding system and island colonization may be obscured by filtering at other stages or by other causes, unrelated to mate limitation. For example, if SI species disproportionately possess traits that either promote dispersal or enhance colonization success (unrelated to mate limitation), a pattern seemingly opposing the one predicted by Baker could arise. Carlquist (1966) observed that if the arriving SI species harbor greater genetic diversity than SC species, they may display enhanced ability to colonize novel habitats. Clearly, many other unmeasured variables may obscure the effect of Baker's law. Our ability to detect the strength of the effect may be additionally confounded by various sources of error and bias, related to the scoring of traits and

geographic distributions, as well as a variety of non-random sampling artifacts (e.g., ascertainment or acquisition bias).

Contemporary approaches to study of Baker's law employ ever-larger databases of breeding systems. Establishing whether individuals, populations, and species are SC is difficult, even with small sample sizes and geographic areas (Charlesworth 1985; Raduski *et al.* 2012). The limited number of studies and study sites within single species ensure that breeding systems are pragmatically stereotyped across species' ranges. For example, a taxon that occurs on both mainland and islands (island-occurring or island-colonizing species) may have only been studied for breeding system in a mainland population. Nevertheless, breeding systems may vary within and between populations. In virtually all detailed studies, SC individuals are found in otherwise SI species (reviewed in Stone 2002; Igić *et al.* 2008). This pattern is thought to almost exclusively occur due to the high rate of recurrent segregating loss-of-function mutations in SI populations that cause SC (Raduski *et al.* 2012; Goldberg and Igić 2012). Such mutations, often associated with polyploidization, only rarely become fixed and generally fail to yield lineage transitions (Robertson *et al.* 2011).

Phenotypic estimates of breeding systems, derived from limited crossing designs, often conflate a number of unrelated phenomena, including agamospermy, inbreeding depression, and maternal effects (Charlesworth 1985; Raduski *et al.* 2012). Therefore, it is possible that SI species contain some number of SC island populations, whether sourced by filtering segregating variation or secondarily (*de novo*) evolved on islands. These polymorphisms are sometimes papered over by the discretizing breeding system scoring schemes, which practically assign a single binary state to each species. Nevertheless, neither experimental error nor segregating mutants are likely to conservatively bias our results (away from predictions of Baker's law) by artificial omission of filtered occurrences. Practical mistakes in scoring breeding systems would effectively blur the detectable signature of selection, and therefore instead systematically underestimate the magnitude of breeding system filtering with increased mate limitation.

Our approach of contrasting breeding systems in mainland-only species (M) versus those than occur on both mainlands and islands (M+I; island-occurring) is practically necessary to minimize the confounding effects of secondary trait evolution (modification of obligate outcrossing on islands after establishment). This approach may systematically underestimate of the filtering effect of Baker's law. The fact that M+I species retain their conspecific taxonomic status may be attributable to ongoing gene flow. It may also be attributable to recent arrival and represent populations little more than ghosts of dispersal past, simply awaiting their turn on the ratchet of

18

extinction. Another potential source of bias may arise if breeding systems and island distribution are non-randomly studied or reported in the literature, so that the numbers of SI island colonists are under-reported, which could affect the inference of probability of island occurrence. This is unlikely given the relative publication value of reporting SI colonists. The scale of these problems is unclear.

Finally, the prevailing taxonomic practices are broadly reliant on identifiable morphological features fixed between populations. This practice could disproportionately favor the splitting of SC species and lumping of SI species (Goodwillie 1999; Raduski and Igić 2021), yielding an artificially inflated rate of island endemism for SC, compared with SI species. Combined with a possibly excessive taxonomic splitting among island-occurring taxa, this practice may underestimate dispersal-and-establishment events by SC species and falsely weaken the signal of Baker's law. We find only a modest difference between the models that include and exclude island endemics (Model 1a vs. 1b), so the latter of these does not seem to strongly affect our main results. While there are many other possible sources of error and bias, as well as unmeasured variables and interactions, we believe that our analyses are generally conservative, lowering the effects of mate limitation during establishment.

Conclusion

We focused here on the problem of determinants of island colonization through the lens of interaction between mate limitation and breeding system. While confirming the predictive power of breeding systems, our approach also identified lifespan, floral symmetry, and island arrival opportunity as potential interacting factors. These findings reveal promising avenues for the study of plant colonization dynamics, providing novelty in a field dominated by decades-old observations, which often conflate underlying processes. For example, our results hint that a complex interplay among many ecological and evolutionary processes is responsible for shaping plant distributions more generally. Reliance on improved conceptual framing, advanced modeling techniques, taxonomically and geographically expanded datasets, and better proxies for both propagule pressure and mate limitation, will help disentangle the processes that explain differential colonization success. A deeper understanding of these dynamics— including dispersal and post-colonization processes— will provide increasingly unified theories of island biogeography and trait evolution.

Acknowledgments

Author Contributions

ANZ, BI, and DLG designed the research; ANZ, BI, CM, DLG, and EG performed the research; ANZ and DLG conducted data analyses; ANZ, BI, CM, EG, and team of undergraduates collected the data; ANZ, BI, CM, DLG, and EG were involved in data interpretation; ANZ and BI wrote and edited the manuscript. CM, DLG, and EG edited manuscript.

Data availability

All raw trait data files on breeding systems, floral symmetry, and lifespan is available on GitHub: https://github.com/dirtyplants/BSdb.

Competing interests

None declared.

Table 1: The effect of breeding system (BS) on island presence or absence as assessed by generalized linear mixed models. Model 1 considers breeding system as the only predictor, and contains five variants (a-e), which differ in data scoring. Models 2-8 include variations of interactions with breeding system, lifespan (LS), floral symmetry (FS) and species-island distance (Distance). Marginal (Mar.) r^2 values denote the proportion of variance explained by predictors (fixed effects). For each species, plant family was included as a random effect (genus level analysis model 1c). Conditional (Cond.) r^2 values denote the proportion of variance explained by breeding system and the random factor (family membership). *In addition to excluding 217 island endemics, 74 species with missing occurrence data were excluded from Models 1b-5. (CAPTION REFERS TO TABLE ON FOLLOWING MANUSCRIPT PAGE.)

Island presence vs. absence	$N_{spp.}$	N_{Family}*	Cond. r^2	Marg. r^2	Predictors	p-value (Bonferroni-corrected)	AIC[†]
Model 1: BS independently							
(a) Including all species	3222	169	0.26	0.04	Breeding system	≪ **0.001**	(3722)
(b) Excluding island endemics‡	2931	162	0.23	0.03	Breeding system	≪ **0.001**	3175
(c) Genus level random effect (Family(Genus))	2931	1181	0.41	0.03	Breeding system	≪ **0.001**	(3066)
(d) SI threshold ISI ≥ 0.9	2931	162	0.23	0.03	Breeding system	≪ **0.001**	(3179)
(e) Islands only (oceanic vs. continental)	1015	112	0.17	0.01	Breeding system	0.15	(1376)
Model 2: BS-FS and BS-LS interactions	2931	162	0.21	0.05	Breeding system	0.73	3148
					Lifespan	1.00	
					Floral symmetry	1.00	
					Breeding system×Lifespan	0.47	
					Breeding system×Floral symmetry	0.68	
Model 3: BS-Distance interaction	2931	162	0.31	0.10	Breeding system	≪ **0.001**	3060
					Distance	≪ **0.001**	
					Breeding system×Distance	1.00	
Model 4: Add BS-FS and BS-LS interactions	2931	162	0.29	0.13	Breeding system	0.31	3027
					Lifespan	0.78	
					Floral symmetry	1.00	
					Distance	≪ **0.001**	
					Breeding system×Lifespan	0.32	
					Breeding system×Floral symmetry	0.66	
					Breeding system×Distance	1.00	
Model 5: Add LS-Distance interaction	2931	162	0.29	0.15	Breeding system	< **0.05**	**3011**
					Lifespan	< **0.05**	
					Floral symmetry	1.00	
					Distance	≪ **0.001**	
					Breeding system×Lifespan	< **0.05**	
					Breeding system×Floral symmetry	0.77	
					Breeding system×Distance	1.00	
					Lifespan×Distance	**0.005**	
Model 6: Add FS-Distance interaction	2931	162	0.29	0.15	Breeding system	0.05	3013
					Lifespan	< **0.05**	
					Floral symmetry	1.00	
					Distance	≪ **0.001**	
					Breeding system×Lifespan	< **0.05**	
					Breeding system×Floral symmetry	0.71	
					Breeding system×Distance	1.00	
					Lifespan×Distance	0.004	
					Floral symmetry ×Distance	1.00	
Model 7: Remove BS-Distance interaction	2931	162	0.29	0.15	Breeding system	0.10	3012
					Lifespan	0.07	
					Floral symmetry	1.00	
					Distance	≪ **0.001**	
					Breeding system×Lifespan	0.065	
					Breeding system×Floral symmetry	0.56	
					Lifespan×Distance	**0.01**	
Model 8: BS-LS-Distance 3-way interaction	2931	162	031	0.13	Breeding system	≪ **0.001**	3040
					Lifespan	0.003	
					Distance	≪ **0.001**	
					Breeding system× Lifespan	0.003	
					Breeding system× Distance	0.55	
					Lifespan× Distance	0.06	
					Breeding system× Lifespan× Distance	0.94	

*Family as a random factor, except in Model 1c, where genus is nested within family as a random factor.

†Parentheses around AIC scores for models that are not comparable or nested within others.

‡In addition to 217 island endemics, 74 species with missing occurrence data were excluded from Models 1b-5.

Literature Cited

Azani N, **Babineau M**, **Bailey CD**, **Banks H**, **Barbosa AR**, **Pinto RB**, **Boatwright JS**, **Borges LM**, **Brown GK**, **Bruneau A**, *et al.* **2017**. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: The Legume Phylogeny Working Group (LPWG). *Taxon* **66**: 44–77.

Baker H. **1948**. Dimorphism and monomorphism in the Plumbaginaceae: I. A survey of the family. *Annals of Botany* **12**: 207–219.

Baker HG. **1953**. Race formation and reproductive method in flowering plants. In: *Cold Spring Harbor Symposia on Quantitative Biology*. vol. 7, 114–143.

Baker HG. **1955**. Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* **9**: 347–349.

Baker HG. **1965**. Characteristics and modes of origin of weeds. In: Baker HG, Stebbins GL, eds., *The Genetics of Colonizing Species*, New York and London: Academic Press. 147–168.

Baker HG. **1967**. Support for Baker's law–as a rule. *Evolution* **21**: 853–856.

Barrett SC. **2016**. Foundations of invasion genetics: the Baker and Stebbins legacy. In: Barrett C, Collautti R, Dlugosch K, Rieseberg L, eds., *Invasion genetics: the Baker and Stebbins legacy*, Wiley Online Library. 1–18.

Bates D, **Mächler M**, **Bolker B**, **Walker S**. **2015**. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**: 1–48.

Böhle U-R, **Hilger HH**, **Martin WF**. **1996**. Island colonization and evolution of the insular woody habit in *Echium* L.(Boraginaceae). *Proceedings of the National Academy of Sciences* **93**: 11740–11745.

Boyle B, **Hopkins N**, **Lu Z**, **Raygoza Garay JA**, **Mozzherin D**, **Rees T**, **Matasci N**, **Narro ML**, **Piel WH**, **Mckay SJ**, *et al.* **2013**. The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics* **14**: 1–15.

Brummit R. **2001**. World geographic scheme for recording plant distributions, edition 2. international working group on taxonomic databases for plant sciences.

Carlquist S. **1966**. The biota of long-distance dispersal. IV. Genetic systems in the floras of oceanic islands. *Evolution* : 433–455.

Carlquist SJ. **1974**. *Island Biology*. New York: Columbia University Press.

Carvajal-Endara S, **Hendry AP**, **Emery NC**, **Davies TJ**. **2017**. Habitat filtering not dispersal

limitation shapes oceanic island floras: species assembly of the Galápagos archipelago. *Ecology Letters* **20**: 495–504.

Charlesworth D. **1985**. Distribution of dioecy and self-incompatibility in angiosperms. In: Greenwood PJ, Harvey PH, Slatkin M, eds., *Evolution: Essays in Honour of John Maynard Smith*, Cambridge: Cambridge University Press. 237–268.

Cheptou P-O. **2012**. Clarifying Baker's law. *Annals of Botany* **109**: 633–641.

Darwin C. **1859**. *On the Origin of Species by Means of Natural Selection*. London: John Murray. Or the Preservation of Favored Races in the Struggle for Life.

Darwin C. **1876**. *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom*. London: John Murray.

Darwin C. **1877**. *The Various Contrivances by which Orchids are Fertilised by Insects*. London: John Murray.

Dobzhansky T. **1950**. Evolution in the tropics. *American Scientist* **38**: 209–221.

eFloras . **2023**. Missouri Botanical Garden (St. Louis, MO) and Harvard University Herbaria (Cambridge, MA).

EOL . **2023**. Encyclopedia of Life [online].

Fenster CB, **Marten-Rodriguez S**. **2007**. Reproductive assurance and the evolution of pollination specialization. *International Journal of Plant Sciences* **168**: 215–228.

FNA . **2023**. Flora of North America North of Mexico [online].

Funk VA, **Susanna A**, **Steussy TF**, **Robinson HE**. **2009**. Classification of Compositae. *Systematics, Evolution, and Biogeography of Compositae* .

GBIF . **2023**. GBIF: The Global Biodiversity Information Facility.

Goldberg EE, **Igić B**. **2012**. Tempo and mode in plant breeding system evolution. *Evolution* **66**: 3701–3709.

Goodwillie C. **1999**. Multiple origins of self-compatibility in *Linanthus* section *Leptosiphon* (Polemoniaceae): phylogenetic evidence from internal-transcribed-spacer sequence data. *Evolution* **53**: 1387–1395.

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**: 215.

Grossenbacher DL, **Brandvain Y**, **Auld JR**, **Burd M**, **Cheptou P-O**, **Conner JK**,

Grant AG, **Hovick SM**, **Pannell JR**, **Pauw A**, *et al.* **2017**. Self-compatibility is over-represented on islands. *New Phytologist* **215**: 469–478.

Hiraiwa MK, **Ushimaru A**. **2017**. Low functional diversity promotes niche changes in natural island pollinator communities. *Proceedings of the Royal Society B: Biological Sciences* **284**: 20162218.

Igić B, **Lande R**, **Kohn JR**. **2008**. Loss of self-incompatibility and its evolutionary consequences. *International Journal of Plant Sciences* **169**: 93–104.

Inoue K. **1993**. Evolution of mutualism in plant-pollinator interactions on islands. *Journal of Biosciences* **18**: 525–536.

Johnson PC. **2014**. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution* **5**: 944–946.

Kattge J, **Bönisch G**, **Díaz S**, **Lavorel S**, **Prentice IC**, **Leadley P**, **Tautenhahn S**, **Werner GD**, **Aakala T**, **Abedi M**, *et al.* **2020**. TRY plant trait database–enhanced coverage and open access. *Global Change Biology* **26**: 119–188.

Kling MM, **Ackerly DD**. **2021**. Global wind patterns shape genetic differentiation, asymmetric gene flow, and genetic diversity in trees. *Proceedings of the National Academy of Sciences* **118**: e2017317118.

Kubitzki K. **2004**. *Flowering Plants. Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales*, vol. VI. Berlin and Heidelberg: Springer.

Kubitzki K. **2007**. *Flowering Plants. Eudicots: Berberidopsidales, Buxales, Crossosomatales, Fabales p.p., Geraniales, Gunnerales, Myrtales p.p., Proteales, Saxifragales, Vitales, Zygophyllales, Clusiaceae Alliance, Passifloraceae Alliance, Dilleniaceae, Huaceae, Picramniaceae, Sabiaceae*, vol. IX. Berlin and Heidelberg: Springer.

Kubitzki K. **2014**. *Flowering Plants. Eudicots: Malpighiales*, vol. XI. Berlin and Heidelberg: Springer.

Lens F, **Davin N**, **Smets E**, **del Arco M**. **2013**. Insular woodiness on the Canary Islands: a remarkable case of convergent evolution. *International Journal of Plant Sciences* **174**: 992–1013.

Lenth RV. **2023**. *emmeans: Estimated marginal means, a.k.a. least-squares means*. https://CRAN.R-project.org/package=emmeans, R package version 1.8.8. edn.

Lloyd D. **1980**. Demography and evolution in plant populations. In: Solbrig O, ed., *Demographic factors and mating patterns in angiosperms*, Blackwell, Oxford. 209–224.

MacArthur RH, **Wilson EO**. **1967**. *The theory of island biogeography*. Princeton, N.J.: Princeton University Press.

Maitner BS, **Boyle B**, **Casler N**, **Condit R**, **Donoghue J**, **Durán SM**, **Guaderrama D**, **Hinchliff CE**, **Jørgensen PM**, **Kraft NJ**, *et al.* **2018**. The `bien` R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods in Ecology and Evolution* **9**: 373–379.

Martén-Rodríguez S, **Quesada M**, **Castro A-A**, **Lopezaraiza-Mikel M**, **Fenster CB**. **2015**. A comparison of reproductive strategies between island and mainland Caribbean Gesneriaceae. *Journal of Ecology* **103**: 1190–1204.

Pannell JR. **2015**. Evolution of the mating system in colonizing plants. *Molecular Ecology* **24**: 2018–2037.

Pannell JR, **Auld JR**, **Brandvain Y**, **Burd M**, **Busch JW**, **Cheptou P-O**, **Conner JK**, **Goldberg EE**, **Grant A-G**, **Grossenbacher DL**, *et al.* **2015**. The scope of Baker's law. *New Phytologist* **208**: 656–667.

Pannell JR, **Barrett SC**. **1998**. Baker's law revisited: reproductive assurance in a metapopulation. *Evolution* **52**: 657–668.

POWO . **2023**. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew.

R Core Team . **2023**. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raduski AR, **Haney EB**, **Igić B**. **2012**. The expression of self-incompatibility in angiosperms is bimodal. *Evolution* **66**: 1275–1283.

Raduski AR, **Igić B**. **2021**. Biosystematic studies on the status of *Solanum chilense*. *American Journal of Botany* **108**: 520–537.

Razanajatovo M, **van Kleunen M**, **Kreft H**, **Dawson W**, **Essl F**, **Pergl J**, **Pyšek P**, **Winter M**, **Weigelt P**. **2019**. Autofertility and self-compatibility moderately benefit island colonization of plants. *Global Ecology and Biogeography* **28**: 341–352.

Robertson K, **Goldberg EE**, **Igić B**. **2011**. Comparative evidence for the correlated evolution of polyploidy and self-compatibility in Solanaceae. *Evolution* **65**: 139–155.

Sayre R, **Noble S**, **Hamann S**, **Smith R**, **Wright D**, **Breyer S**, **Butler K**, **Van Graafeiland K**, **Frye C**, **Karagulle D**, *et al.* **2019**. A new 30-meter resolution global shoreline vector and associated global islands database for the development of standardized ecological coastal

units. *Journal of Operational Oceanography* **12**: S47–S56.

Simberloff D. **2009**. The role of propagule pressure in biological invasions. *Annual Review of Ecology, Evolution, and Systematics* **40**: 81–102.

Simpson MG. **2019**. *Plant Systematics*. Academic Press, third edn.

SolSource . **2023**. PBI Solanum Project Solanaceae Source: A global taxonomic resource for the nightshade family.

Stebbins GL. **1957**. Self fertilization and population variability in the higher plants. *The American Naturalist* **91**: 337–354.

Stern WL. **2014**. *Anatomy of the Monocotyledons Volume X: Orchidaceae*. Oxford University Press, Oxford.

Stoffel MA, **Nakagawa S**, **Schielzeth H**. **2021**. partr2: partitioning r2 in generalized linear mixed models. *PeerJ* **9**: e11414.

Stone JL. **2002**. Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. *The Quarterly Review of Biology* **77**: 17–30.

Wallace AR. **1878**. *Tropical nature, and other essays*. Macmillan and Company.

Wallace AR. **1880**. *Island Life; or, The Phenomena and Causes of Insular Faunas and Floras*. London: Macmillan.

Wessinger CA, **Kelly JK**. **2018**. Selfing can facilitate transitions between pollination syndromes. *The American Naturalist* **191**: 582–594.

Wiersema JH. **2023**. USDA-ARS Germplasm Resources Information Network (GRIN). http://www.ars-grin.gov/misc/tax/.

Wisz MS, **Pottier J**, **Kissling WD**, **Pellissier L**, **Lenoir J**, **Damgaard CF**, **Dormann CF**, **Forchhammer MC**, **Grytnes J-A**, **Guisan A**, *et al.* **2013**. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* **88**: 15–30.

Wolowski M, **Saad CF**, **Ashman T-L**, **Freitas L**. **2013**. Predominance of self-compatibility in hummingbird-pollinated plants in the Neotropics. *Naturwissenschaften* **100**: 69–79.

Yoder JB, **Gomez G**, **Carlson CJ**. **2020**. Zygomorphic flowers have fewer potential pollinator species. *Biology Letters* **16**: 20200307.

Zizka A, **Onstein RE**, **Rozzi R**, **Weigelt P**, **Kreft H**, **Steinbauer MJ**, **Bruelheide H**, **Lens F**. **2022**. The evolution of insular woodiness. *Proceedings of the National Academy of Sciences* **119**:

e2208629119.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Methods S1. Supporting methods.

Notes S1. Supporting results.

Figure S1. ISI distribution with ISI threshold of 0.8 and 0.9 marked.

Table S1. Sample counts representing each listed taxonomic rank.

Table S2. Breeding system discretization counts by threshold ISI values (2931 species).

Table S3. Island absence (Mainland Only) and presence (Mainland+Island) data and filtering.

Table S4. Island classification (geological origin) for all 1015 island-occurring species in our dataset, including island endemics.

Table S5. Lifespan trait species count.

Table S6. Breeding system by lifespan joint trait count.

Table S7. Floral symmetry trait species count.

Table S8. Breeding System by floral symmetry species count.

Table S9. Breeding system by island presence trait joint coverage data table.

Table S10. Breeding system, lifespan, and island presence trait joint coverage.

Table S11. Breeding system, floral symmetry, and island presence trait joint coverage.

Table S12. Extended version of Table 1 from the main text showing generalized linear mixed models assessing the effect of breeding system (BS) on island presence or absence.

Table S13. Model output with trait interaction probabilities from the generalized linear mixed model (Model 5).

Table S14. Model output parameters: pairwise comparisons for breeding system, lifespan, and floral symmetry traits from the generalized linear mixed model (Model 5).

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Figure 1: Summary of data, showing sample sizes for each trait. **(a)** Number of species in each trait category (self-incompatible (SI) vs. self-compatible (SC), annual vs. perennial, bilateral vs. radial, respectively, as illustrated). **(b)** The frequency distribution of species-island distances for species in our dataset, with an inset illustrating how the key metric, median distance to the nearest island of each species' occurrence, was obtained. Error bars show 95% confidence intervals.
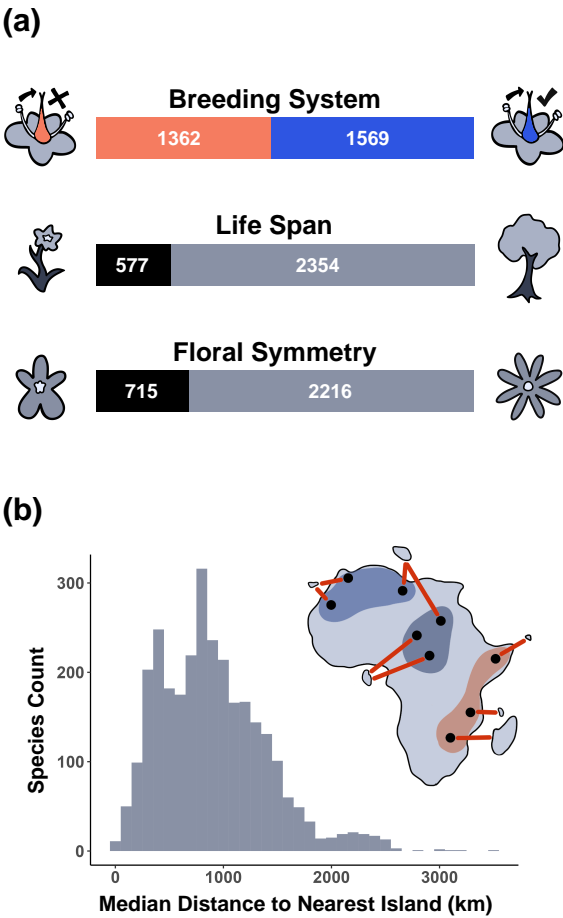
Figure 2: Calculated probabilities of island presence with different predictors (GLMM; Table 1).
(a) Breeding system as the sole predictor in Model 1b. An asterisk marks the significant difference between SI and SC; (z-value = 7.35, $p < 0.001$)). **(b)** Breeding system and lifespan. Trait combinations include: self-incompatible-annual (SI-AN), self-incompatible-perennial (SI-PER), self-compatible-annual (SC-AN) and self-compatible-perennial (SC-PER). The four trait combinations are labeled A-C with an asterisk marking only significant pairwise comparisons (AB*, AC*, AD*, CD, CB*, DB*; Table S14). **(c)** Breeding system and floral symmetry. Trait combinations include self-incompatible-bilateral (SI-BI), self-incompatible-radial (SI-RA), self-compatible-bilateral (SC-BI) and self-compatible-radial (SC-RA). The four trait combinations are labeled A-C with an asterisk marking only significant pairwise comparisons (AB, AC*, AD*, CD*, CB, DB*; Table S14).
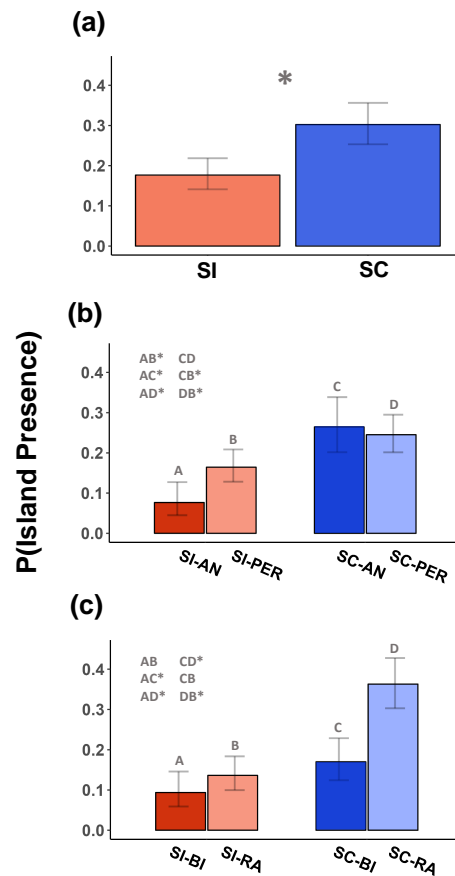
Figure 3: Probability of island presence, predicted by breeding system, lifespan, and species-island distance. The results shown are from Model 5 (GLMM; Table 1). **(a)** The probability of island presence is higher for self-compatible (SC, blue hues) than self-incompatible species (SI, orange hues) while showing the same trend of decrease with dispersal opportunity. The probability of island presence is higher for annuals at closer distances (dark), but decreases for perennials (light) at farther distances. **(b)** The relative island occurrence probability of SC to SI is higher at all distances, and increases with distance. **(c)** The relative island occurrence probability of perennials to annuals is lower at short dispersal distances (high dispersal opportunity) but subsequently increases with species-island distances after circa 1000km. The 95% confidence intervals are estimated from 100 bootstrap replicates.
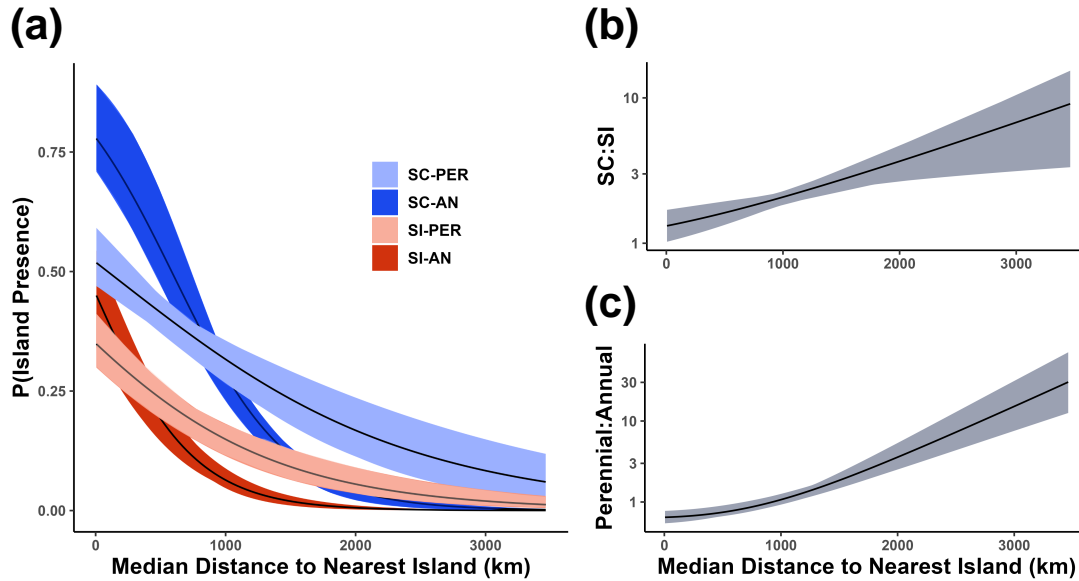
Figure 4: Proportion of variance explained, R^2 (Table 1, Model 5). Main effects on the y-axis are pooled with interactions (Stoffel *et al.* 2021), which is why they add up to more than the full model. Effects are given in decreasing order of means for fixed and random effects (full model, conditional; 23%), main effects alone (full model, marginal; 11%); each main effect, independently (distance 8%, breeding system 5%, floral symmetry 2%, and lifespan 1%). For the full model, conditional R^2 is almost twice as much as the marginal R^2 because a large proportion of variance is explained by random effects (taxonomic nesting).