

Self-compatibility is over-represented on islands

Dena L. Grossenbacher^{1,2,3*}, Yaniv Brandvain^{2*}, Josh R. Auld⁴, Martin Burd⁵, Pierre-Olivier Cheptou⁶, Jeffrey K. Conner⁷, Alannie G. Grant⁸, Stephen M. Hovick⁹, John R. Pannell¹⁰, Anton Pauw¹¹, Theodora Petanidou¹², April M. Randle¹³, Rafael Rubio de Casas¹⁴, Jana Vamosi¹⁵, Alice Winn¹⁶, Boris Igic^{17*}, Jeremiah W. Busch^{3*}, Susan Kalisz^{8*} and Emma E. Goldberg^{18*}

¹Department of Biology, California Polytechnic State University, San Luis Obispo, CA 93407, USA; ²Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN 55108, USA; ³School of Biological Sciences, Washington State University, Pullman, WA 99164, USA; ⁴Department of Biology, West Chester University, West Chester, PA 19383, USA; ⁵School of Biological Sciences, Monash University, Melbourne, VIC 3800, Australia; ⁶CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, CEFE, 34293 Montpellier Cedex 05, France; ⁷Kellogg Biological Station and Department of Plant Biology, Michigan State University, Hickory Corners, MI 49060, USA; ⁸Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA; ⁹Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, OH 43210, USA; ¹⁰Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland; ¹¹Department of Botany and Zoology, Stellenbosch University, Stellenbosch 7602, South Africa; ¹²Laboratory of Biogeography and Ecology, Department of Geography, University of the Aegean, 81100 Mytilene, Lesbos, Greece; ¹³Department of Environmental Science, University of San Francisco, San Francisco, CA 94117, USA; ¹⁴Departamento Ecología, Facultad de Ciencias, Universidad de Granada, Granada, UGR, 18071 Granada, Spain; ¹⁵Department of Biological Sciences, University of Calgary, Calgary, AB, Canada T2N 1N4; ¹⁶Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA; ¹⁷Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA; ¹⁸Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55108, USA

Summary

Author for correspondence:

Dena L. Grossenbacher

Tel: +1 530 204 7824

Email: dgrossen@calpoly.edu

Received: 16 December 2016

Accepted: 20 February 2017

New Phytologist (2017) 215: 469–478

doi: 10.1111/nph.14534

Key words: Baker's law, biogeography, ecological filtering, island, mainland, self-incompatibility.

- Because establishing a new population often depends critically on finding mates, individuals capable of uniparental reproduction may have a colonization advantage. Accordingly, there should be an over-representation of colonizing species in which individuals can reproduce without a mate, particularly in isolated locales such as oceanic islands. Despite the intuitive appeal of this colonization filter hypothesis (known as Baker's law), more than six decades of analyses have yielded mixed findings.
- We assembled a dataset of island and mainland plant breeding systems, focusing on the presence or absence of self-incompatibility. Because this trait enforces outcrossing and is unlikely to re-evolve on short timescales if it is lost, breeding system is especially likely to reflect the colonization filter.
- We found significantly more self-compatible species on islands than mainlands across a sample of > 1500 species from three widely distributed flowering plant families (Asteraceae, Brassicaceae and Solanaceae). Overall, 66% of island species were self-compatible, compared with 41% of mainland species.
- Our results demonstrate that the presence or absence of self-incompatibility has strong explanatory power for plant geographical patterns. Island floras around the world thus reflect the role of a key reproductive trait in filtering potential colonizing species in these three plant families.

Introduction

Colonization, which requires both dispersal and establishment, shapes species' geographical distributions. Factors that determine successful colonization are difficult to observe directly, but they can be investigated by studying the traits of successful colonists. For example, if dispersal is infrequent or involves single or few propagules, the inability to find a mate should act as a filter prohibiting establishment of a new population. Baker (1955) argued that island floras would bear the mark of this filter, possessing an

over-representation of species with traits that liberate individuals from the necessity of finding mates, such as the capacity for self-fertilization. In principle, this filter applies in many geographic contexts, including invasions, natural range expansions, metapopulation dynamics and island colonization (Pannell & Barrett, 1998; Cheptou, 2012; Pannell, 2015; Pannell *et al.*, 2015). The filter also applies to animals, fungi and plants, and to all traits permitting uniparental reproduction, including asexual-ity, clonality and self-fertilization (Pannell *et al.*, 2015).

The intuitive appeal of Baker's logic is compelling. Clearly a lone colonist is doomed if it cannot reproduce. However, a variety of complicating factors make its predictive utility unclear.

*These authors contributed equally to this work.

The capacity for uniparental reproduction is only one of several traits that may influence establishment following long-distance dispersal, and the ability to reproduce uniparentally may not be important if colonists arrive frequently or in groups, such as in multi-seeded fruits or by rafts or animals (Baker, 1967; Nowak *et al.*, 2014; Schlessman *et al.*, 2014). In addition, if species incapable of uniparental reproduction tend to have traits that promote long-distance dispersal, an association may arise between outcrossing and island presence, a pattern opposing the one envisioned by Baker (Cheptou & Massol, 2009; Auld & de Casas, 2012). Even if the ability to self improves the odds of successful establishment, selection to avoid selfing in subsequent generations (e.g. leading to the evolution of dioecy or other outcrossing mechanisms (Carlquist, 1966; Baker & Cox, 1984; Sakai *et al.*, 1995; Webb *et al.*, 1999; Schlessman *et al.*, 2014)) could obscure lasting signatures of colonization filters on islands.

Previous tests of the island colonization filter

Baker's hypothesis was dubbed 'Baker's law' by Stebbins (1957) on the basis of several examples in which self-compatibility is more common in species that are outside the core geographical range of their genus. Since then, a wealth of studies has accumulated on plant reproductive systems and their geographical distributions. However, this work has often led to conflicting results. To illustrate this, we summarize in this section the key attributes of studies that assess reproductive systems in island taxa to test the core predictions of Baker's law (106 papers from 1962 to 2015; see Supporting Information Table S1). These studies examined a wide range of plants (211 families) and some insects (two families). Numerous islands and archipelagos were included, with considerable attention paid to Hawaii, Galápagos, Juan Fernández, Canary, Réunion and New Zealand (Fig. S1). Although Baker's law has been considered in a variety of spatial contexts, its support should be strongest when colonists are rare, as expected for those arriving on distant islands (Pannell *et al.*, 2015).

Traits that reflect a colonization filter The lack of uniparental reproduction was usually defined by two reproductive traits: self-incompatibility (SI) and dioecy (59% and 32% of studies, respectively). SI systems are genetic mechanisms that cause hermaphrodites to reject their own pollen (de Nettancourt, 1977). SI is common and taxonomically widespread, being found in > 100 flowering plant families (Igic *et al.*, 2008). SI has been lost in thousands of lineages, yielding self-compatible (SC) species, but regaining SI is rare (Stebbins, 1957; Igic & Busch, 2013). Thus, a prevalence of SI on islands would represent strong evidence against Baker's law. Of the studies examining the prevalence of SI on islands, 30 concluded support for Baker's law, whereas seven opposed it, 12 reported mixed patterns, and 14 did not report a conclusion. By contrast, although dioecy is relatively rare in plants, it has likely evolved independently thousands of times (Renner, 2014). Consequently, the prevalence of dioecy on islands could be explained by *in situ* evolution from nondioecious colonist species, and it thus constitutes weaker evidence

against Baker's law (Baker, 1967; Sakai *et al.*, 1995; but see Schlessman *et al.*, 2014). Indeed, studies involving dioecy were more equivocal: six concluded support for Baker's law, whereas seven opposed it, 10 reported mixed patterns, and 11 did not report a conclusion.

Reproductive traits on islands Most previous studies have adopted one of two geographical perspectives to evaluate Baker's law. The majority (76 of 106 studies: 72%) only examined reproductive traits on islands and did not include mainland species or populations as a comparison. The preponderance of this study design may derive from the interpretation of Baker's law that any species incapable of uniparental reproduction on an island is unexpected (Baker (1967) himself accepted that there would be exceptions to his 'rule'). This design may also reflect difficulties in identifying relevant mainland species for comparison, especially for older, oceanic islands distant from continents (e.g. Hawaii). Twenty-eight of these studies concluded support for Baker's law, whereas 13 opposed it, 10 reported mixed patterns, and 25 did not report a conclusion. However, it is difficult to interpret such results without knowing which traits are common in related taxa on mainlands. Studies that did explicitly consider species' traits on mainland regions as well as islands (30 of 106 studies: 28%) allow a comparative test of the hypothesis that a capacity for uniparental reproduction is over-represented on islands. Thirteen of these studies concluded support for Baker's law, whereas four opposed it, four reported mixed patterns, and nine did not report a conclusion.

Sampling and the relevant timescales Most previous studies have focused on few species and islands (Fig. S2), which may allow direct experimental assessment of the capacity for uniparental reproduction. Focusing on island and mainland populations within a single species also reduces the time frame over which evolution on islands could erase the signal of colonization filters. Studies of large taxonomic scope give a complementary, broader view of the importance of mate limitation during colonization. We found only five studies that compared island and mainland reproductive traits for > 100 species, with two reporting support for Baker's law (Igic *et al.*, 2008; Patiño *et al.*, 2013) and three finding opposing or mixed patterns (Baker & Cox, 1984; Newstrom & Robertson, 2005; Lord, 2015).

Our approach

In order to support a uniquely broad and powerful test of the importance of mate limitation as a filter for island colonization, we designed a study that combines the strongest elements of previous work. We compiled data on SI and native species occurrence on continental land masses and islands for > 1500 species from three large, globally distributed plant families (Asteraceae, Brassicaceae and Solanaceae). Our data include species on islands in every ocean and on all continents except Antarctica, and they encompass tens of millions of years of evolution during which hundreds of transitions from SI to SC occurred (Igic *et al.*, 2008; Crawford *et al.*, 2009). We used

these data to conduct a taxonomically controlled, worldwide comparison of the frequencies of breeding systems on mainlands and islands. In support of Baker's hypothesis, we find that SC species are significantly over-represented on islands relative to continental mainlands.

Materials and Methods

Breeding system data collection and classification

We conducted a literature survey to obtain the status of plant taxa in the Asteraceae, Brassicaceae and Solanaceae as self-incompatible (SI) or self-compatible (SC). Relevant sources of information included primary literature, books, published floras and personal communications by experts. Given the diverse methodologies employed to evaluate the breeding systems of plants in thousands of studies to date, we established criteria to quantify the confidence of inferred character states such that it was possible to evaluate the sensitivity of our conclusions to data quality.

Index of self-incompatibility Studies comparing the outcome of manual self-pollinations and cross-pollinations were deemed of highest quality because the fruit set (or, less desirably, seed set) in selfed flowers is relativized and thus takes into account other factors that may influence inference of SI (Raduski *et al.*, 2012). These data, which were available in a relatively small number of species ($n=304$), were used to calculate an index of SI (ISI; Lloyd, 1965; Bawa, 1974):

$$\text{ISI} = 1 - \frac{\% \text{ fruit set in self-pollinations}}{\% \text{ fruit set in cross-pollinations}}$$

SI taxa were defined by the widely used criterion of having $\text{ISI} \geq 0.8$, and species with $\text{ISI} < 0.8$ were considered to be SC. Because the distribution of ISI values is strongly bimodal, the cut-off value only slightly affects the proportion of species in each category (Raduski *et al.*, 2012). To ensure that our results were not sensitive to this specific ISI threshold, we also performed analyses using a stricter definition of SI ($\text{ISI} \geq 0.9$). ISI measures provide the best experimental verifications of incompatibility status, and so we used these estimates in all analyses.

Additional self-incompatibility data Restricting our analyses to species with ISI data excludes many species for which breeding system data are available. We therefore compiled an expanded dataset with a lower standard of evidence to make use of other information about this trait.

In some studies, fruit or seed set was measured in manually self-pollinated flowers without a cross-pollination treatment. This provides information on the potential for self-fertilization. In such cases, we considered plants SC if fruit or seed set was within the range of values observed in open-pollinated or cross-pollinated fruits. If selfed flowers produced very little fruit or seed set, the taxon was categorized as SI, although resource limitation or pollen inviability cannot be excluded. In experiments where

flowers were experimentally bagged to exclude pollinators, fruit or seed set indicates SC, but an absence of data is uninformative because either SI or an inability to autonomously self-pollinate is possible.

In other cases, qualitative reports regarding the incompatibility status of a plant taxon have been made without first-hand data. Horticultural databases and papers in which the incompatibility status was not the primary focus, such as monographs and floras, often report the incompatibility status of a plant taxon without accompanying experimental data. Such expert evaluations likely contain some correct information about breeding systems.

In our dataset, we retained only those species that were consistently reported as either SI or SC. Species for which different studies reported conflicting results and those with among-population variation in breeding system were eliminated. Because it is common for occasional SC individuals to be found within SI species, however, we did not exclude species on that basis. The resulting data included SI or SC status for 295 species in the ISI dataset (159 Asteraceae, 37 Brassicaceae and 99 Solanaceae) and 1560 species in the expanded dataset (783 Asteraceae, 304 Brassicaceae and 473 Solanaceae).

Island classification

Testing Baker's law requires classifying geographical regions as either mainlands or islands (Table S2). We used the finest resolution of the Biodiversity Information Standards (TDWG) classification scheme (<http://www.tdwg.org/standards/109>), which subdivides the world into regions, informed by botanical affinities. Mainlands are North America, South America, Europe, Asia, Africa and Australia (no study species occurred in Antarctica). Our island category includes all TDWG regions that are not contiguous with continental mainlands, such that the entire region is surrounded by water. This categorization does exclude some islands that are close to continents when a single TDWG region includes both a portion of a mainland and a nearby island. For example, although Vancouver Island is surrounded by water, it is excluded from our island category because it is in the same region as mainland British Columbia. The islands excluded by this categorization seem least likely to show a signature of the mate-limitation filter if colonization is frequent due to the proximity of the mainland.

The strength of a colonization filter may depend on an island's location and geological history. We recorded the distance of each island from the nearest mainland, expecting weaker propagule pressure and consequently greater mate limitation for more distant islands. We distinguished islands that formed over continental plates and were potentially connected to the mainland at some point in geological history from islands that formed over oceanic plates and have never been connected to any continental mainland. We expected a stronger colonization filter for oceanic islands because all taxa on these islands must have been colonizers at some time. A few island groups (e.g. those in the Caribbean Sea) included continental and oceanic islands and were classified as continental. Data sources for island geology and distance from nearest mainland included the United Nations Environment

Program database (<http://islands.unep.ch>) and primary literature. In a limited number of cases, distance to nearest mainland was quantified using straight line distance in Google Earth (<https://www.google.com/earth>).

Species name resolution: accepted names and synonyms

For both breeding system and geography, data were recorded under the species name used in the relevant publication. These names were then synonymized to reflect current taxonomy. The primary sources of synonymy were Solanaceae Source (<http://solanaceaesource.org>), the Brassicaceae species checklist (Warwick *et al.*, 2006) and the Global Compositae (Asteraceae) Checklist (<http://compositae.landcareresearch.co.nz>). Names not found in those sources were synonymized with GRIN (<http://www.ars-grin.gov>), Tropicos (<http://www.tropicos.org>) or The Plant List (<http://www.theplantlist.org>), in that order of precedence. Name matching was conducted with our own scripts and with the aid of the Taxonomic Name Resolution Service (<http://tnrs.iplantcolaborative.org>).

Native species distributions on mainlands and islands

For each species included in our dataset, we determined its native distribution on mainlands and islands, considering both its accepted name and all of its synonyms (Tables S3, S4). Native mainland occurrences were recorded at the level of continent (TDWG level 1 codes), whereas island occurrences were recorded at the finest resolution possible (generally TDWG level 3 or 4 codes). Sources for occurrence data included four publicly available databases, a manual data search of primary literature, and published island floras. We developed a data processing pipeline (Fig. S3), and later we describe how data from each source were handled and verified.

Euro+Med PlantBase Occurrence records were generously provided by Eckhard von Raab-Straube. Euro+Med area codes were matched to TDWG codes resolved to the finest unit possible for each record. We identified records as native (Euro+Med status: native) or non-native (Euro+Med status: introduced, cultivated). Non-native records were used to filter GBIF and island flora records (see later) but dropped in all downstream analyses.

United States Department of Agriculture – Agricultural Research Service, Germplasm Resources Information Network (GRIN) Occurrence records were generously provided by John H. Wiersma. Locality descriptions were matched to TDWG codes resolved to the finest unit possible for each record. We identified records as native (GRIN status: native) or non-native (GRIN status: naturalized, adventive, cultivated). Records with status 'other' or 'uncertain' were dropped. Again, non-native records were used to filter GBIF and island flora records (see later) but dropped in all downstream analyses.

Global Biodiversity Information Facility (GBIF) Occurrence records were downloaded at the family level. TDWG level 4

area codes were identified for each record based on latitude and longitude. To avoid including garden records, we eliminated occurrences containing the following words in the location description: garden, garten, jardin, greenhouse and glasshouse. To avoid including occurrences with imprecise locality information, we eliminated records with whole numbers for latitude and longitude. GBIF records generally lack information on native or non-native status. We therefore determined status ourselves in two ways. First, for records occurring in a TDWG area identified in Euro+Med and USDA-ARS GRIN earlier, we used status from those two databases. For remaining records occurring in TDWG areas unique to GBIF for a given species, we used primary literature and other online data sources to determine whether a record was native, non-native or ambiguous. Mainland codes were resolved to the level of continent, whereas island codes were resolved to the finest unit possible. Only native records were retained.

Global Composite Checklist (GCC) Occurrence records with TDWG codes for Asteraceae were generously provided by Christina Flann. We retained all putatively native records (GCC status: indigenous), and as with GBIF records earlier, we further filtered out non-native records using Euro+Med and USDA-ARS GRIN. For remaining records occurring in TDWG areas unique to GCC for a given species, we used primary literature and other online data sources to determine whether a record was native, non-native or ambiguous. Only native records were retained.

Manual data search In order to ensure that the online databases described earlier were adequately identifying native island occurrences, we additionally manually searched the primary literature and other online databases for a subset of species with breeding system information ($n = 843$), and identified native island occurrences and extracted TDWG codes resolved to the finest unit possible for each species.

Island floras In order to further ensure that we adequately documented native island occurrences, we also gathered data by hand from 26 published island floras. Floras were selected based on detailed descriptions provided in Frodin (2001) and literature searches. We focused on floras described as 'complete', printed in English, most recently published for a given island or island group, and accessible in libraries at our home institutions or online. For each flora, we recorded TDWG area codes for species with breeding system information and their synonyms. When status was indicated in the flora, we recorded whether the island occurrence was a native or non-native. For records where status was not indicated but that occurred in a TDWG area identified in Euro+Med and USDA-ARS GRIN, we used status from those two databases. For remaining records occurring in TDWG areas unique to island floras, we used primary literature and other online data sources to determine whether a record was native, non-native or ambiguous. Only native records were retained.

Statistical analyses

In order to determine whether the proportion of SC species is enriched on islands relative to mainlands, we performed an analysis (model 1) asking whether geography (mainland, island) was predicted by incompatibility status (SI, SC) using a generalized linear mixed model (GLMM; 'GLMER' function in the LME4 R package; Bates *et al.*, 2015) with a logistic link function. The 'mainland' category includes species that occupy only mainlands (never occur on an island). The 'island' category contains species with at least one native island occurrence, including species that occur on both mainlands and islands (395 species) and species that are endemic to islands (127 species). To take into account phylogenetic structure, we included family as a random factor and genus within family as a nested random factor.

In order to ensure that our results were robust to the standards used to determine incompatibility status, we replicated the GLMM analysis restricting the dataset to species with highest quality breeding system data (studies reporting ISI values) using two separate cutoffs for determining SI status ($ISI \geq 0.8$ and $ISI \geq 0.9$, model 2 and model 3, respectively). To determine whether results depended on island type (oceanic, continental, near, far), we fitted four separate models comparing the proportion of SC species on mainlands to that of each island type (models 4–7). To determine whether occurrence on different types of islands was predicted by incompatibility status (SI, SC), we used this same basic model with island type (continental vs oceanic; near vs far) as the response variable in two separate models (models 8–9). 'Continental' species were those restricted to continental islands, whereas 'oceanic' species were those with at least one native oceanic island occurrence. 'Near' species were those restricted to islands < 100 km from a mainland, whereas 'far' species were those with at least one native occurrence on an island > 100 km from mainland.

In order to partially rule out alternative explanations for a disproportionate amount of SC on islands, we replicated model 1 while excluding island endemics in two ways. First, excluding all island endemics regardless of their inferred incompatibility status (model 10) removes the species most likely to have originated on islands, rather than colonized them. Second, excluding only SC island endemics (model 11) removes the species most likely to have undergone post-colonization changes in breeding system.

For models 1–11, we calculated marginal R^2 (proportion of variance explained by the fixed factors alone) and conditional R^2 (proportion of variance explained by both the fixed and random factors) values following Johnson (2014), and we determined significance using likelihood ratio tests with single term deletions with the 'mixed' function in the R package AFEX (Singmann, *et al.*, 2016).

In order to explore the individual effects in each family, we replicated the 11 GLMM models for each family separately, with genus included as a random factor. We took into account multiple test comparisons ($n=44$) by applying the Benjamini & Hochberg correction using the 'p.adjust' function in core STATS package in R.

Finally, because the number of species was highly variable among genera, we also performed a sign test with each genus as a single datum. For each genus that contained at least one mainland and one island occurrence ($n=77$ genera), we calculated the proportion of SC species on islands and mainlands (using the model 1 definitions of breeding system and geographic categories). A paired Wilcoxon signed-rank test assessed whether the proportion of SC species on islands was significantly different from that on mainlands using the 'wilcox.test' function in core stats package in R.

Results

Overall, a significantly greater proportion of island species was SC (66%), compared with mainland species (41%) (Fig. 1). A generalized linear mixed model in which incompatibility status predicts island occurrence, with genus nested within family as a random factor, predicted that the probability of an SC species occurring on an island (0.46) was more than twice that of an SI species (0.18) (Tables 1, 2, model 1). The elevated occurrence of SC species on islands remained significant when we limited our dataset to species whose SI or SC status was determined by empirical studies involving self- and cross-pollinations (Tables 1, 2, model 2; Fig. S4).

Self-compatible taxa were predicted by our model to be significantly more likely than SI taxa to be on islands, regardless of whether islands were oceanic or continental, or whether islands were located near or far from mainlands (Tables 1, 2, models 4–7). There was no difference in proportions of SC taxa on islands near vs far from the mainland (Tables 1, 2, model 9; Fig. S5), but the frequency of SC was 20% greater on oceanic than on continental islands (Tables 1, 2, model 8; Fig. S6). Because the ancestors of all plants on oceanic islands must have been colonists at one point in their history, the elevated frequency of SC on oceanic compared with continental islands suggests more intense trait filtering during or after colonization.

Our result that SC species are significantly more common on islands than on mainlands could be due to processes other than trait filtering during colonization. One alternative explanation is that speciation rates are influenced by a combination of incompatibility status and geography. For example, SI species could have a lower speciation rate on islands than mainlands, or taxonomists could tend to over-describe swarms of selfing species on islands. The species most likely to have originated on islands, rather than having colonized them, are island endemics. Excluding all island endemics (8% of species) from our data thus partially controls for *in situ* speciation, and we find that it does not change our conclusion that SC species are over-represented on islands (Tables 1, 2, model 10; Fig. S7). Another explanation for enriched SC on islands is the repeated loss of SI after colonization. The species most likely to have lost SI after colonization are SC island endemics: SI island species were almost surely ancestrally SI, and widespread SC species, under this alternative explanation, must have independently lost SI on mainlands as well as islands. Excluding all SC island endemics (6% of species) thus

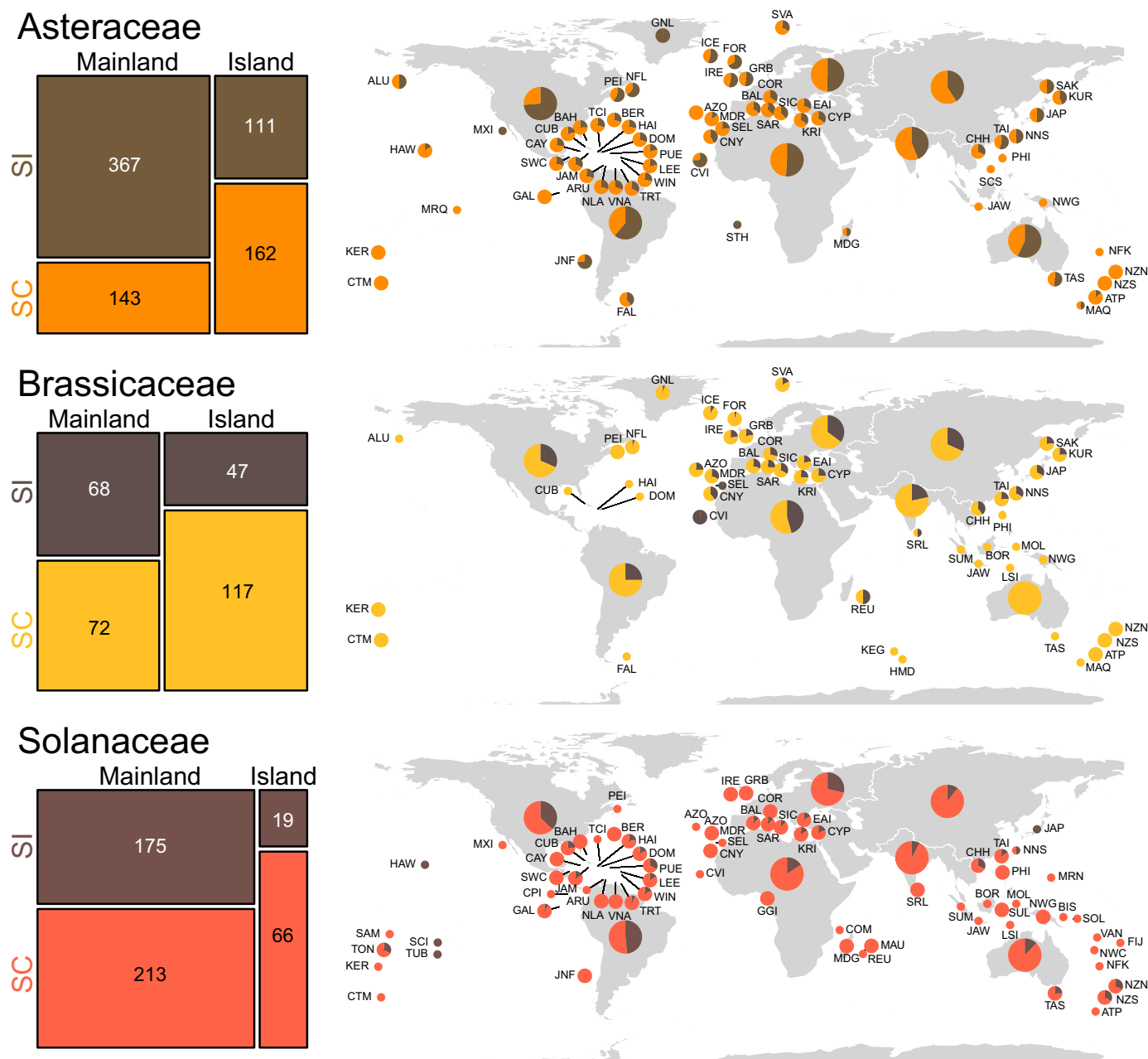


Fig. 1 The distribution of breeding systems among species in three plant families. The 'mainland' category includes species that only occupy mainlands (never occur on an island). The 'island' category includes species that occur on both mainlands and islands (395 species), and species that are endemic to islands (127 species). Numbers indicate the total species count within each category. Adjacent maps indicate the spatial extent of the data and the proportion of self-compatible (SC) and self-incompatible (SI) species on mainlands and islands (large pie charts, mainlands; medium and small pie charts, islands). Medium pie charts indicate islands with ≥ 3 study species; small pie charts indicate islands with < 3 study species. Islands are labeled with Biodiversity Information Standards (TDWG) codes.

largely controls for the possibility that colonists were SI but subsequently evolved SC on islands, and we find that even with this conservative test, SC species are over-represented on islands (Tables 1, 2, model 11; Fig. S8).

Although the earlier analyses included taxonomy as a random factor, we further took into account the phylogenetic nonindependence of closely related species by conducting analyses at a coarser taxonomic scale. A sign test with each genus as a single datum again found a significantly greater representation of SC

species on islands than on mainlands (Wilcoxon signed-rank test, $P < 0.001$; Fig. 2). Complete species-level phylogenies for these families are not available, so a more powerful analysis that thoroughly controls for phylogenetic nonindependence is not currently possible. Both incompatibility status and island status evolve sufficiently quickly, however (as shown by their widespread heterogeneity within genera), that these analyses provide robust evidence for the repeated and consistent association of SC and island presence.

Table 1 Results of 11 separate generalized linear models with a logistic link function and a binomial error distribution, assessing the effect of species' breeding system (SI, self-incompatible; SC, self-compatible) on geography

Model no. and binary response (data subset if any)	No. of spp.	Marg. R^2	Cond. R^2	Prob. on island			<i>P</i>
				SC	SI	LR	
1. Mainland, island	1560	0.060	0.588	0.46	0.18	67.48	<0.0001
2. Mainland, island (only SI breeding system: threshold ≥ 0.8)	297	0.029	0.434	0.39	0.21	4.74	0.03
3. Mainland, island (only SI breeding system: threshold ≥ 0.9)	285	0.051	0.445	0.40	0.18	8.62	0.003
4. Mainland, continental island	1465	0.069	0.213	0.38	0.17	69.9*	<0.0001
5. Mainland, oceanic island	1536	0.105	0.147	0.27	0.09	81.92*	<0.0001
6. Mainland, near island	1440	0.076	0.473	0.27	0.09	55.26	<0.0001
7. Mainland, far island	1556	0.053	0.596	0.39	0.15	55.76	<0.0001
8. Continental island as 'mainland', oceanic island	522	0.011	0.380	0.58	0.46	3.74	0.05
9. Near island as 'mainland', far island	522	0.033	0.072	1.00	1.00	0.01	0.93
10. Mainland, island (all island endemics removed)	1433	0.067	0.202	0.39	0.18	66.50	<0.0001
11. Mainland, island (SC island endemics removed)	1471	0.040	0.172	0.39	0.22	41.99	<0.0001

Family was included as a random factor, as was genus nested within family. In some models (*) genus was dropped as a random factor to achieve model convergence. Bold *P*-values are significant ($P < 0.05$) after Benjamini & Hochberg correction for 44 multiple test comparisons. 'Marg. R^2 ' values are the proportion of variance explained by breeding system (fixed factor). 'Cond. R^2 ' values are the variance explained by breeding system and the random factors of family and genus nested within family for those comparisons where genus was retained. 'Prob. on island' is the predicted probability of an SC or SI species occurring on an island (or on an oceanic or far island, for models 8–9, respectively).

Table 2 For each family (Asteraceae, Brassicaceae and Solanaceae), results of 11 separate generalized linear models with a logistic link function, assessing the effect of species' breeding system (SI, self-incompatible; SC, self-compatible) on geography

Model no. and binary response (data subset if any)	Asteraceae			Brassicaceae			Solanaceae		
	<i>N</i>	LR	<i>P</i>	<i>N</i>	LR	<i>P</i>	<i>N</i>	LR	<i>P</i>
1. Mainland, island	783	29.79	<0.0001	304	12.13	0.0005	473	21.13	<0.0001
2. Mainland, island (only SI breeding system: threshold ≥ 0.8)	159	0.37	0.54	37	7.67	0.006	101	2.88	0.09
3. Mainland, island (only SI breeding system: threshold ≥ 0.9)	151	0.84	0.36	35	5.41	0.02	99	9.60	0.002
4. Mainland, continental island	709	26.50	<0.0001	294	12.19	0.0005	462	16.06	<0.0001
5. Mainland, oceanic island	773	13.91	0.0002	294	5.37	0.02	469	20.98	<0.0001
6. Mainland, near island	698	20.78	<0.0001	287	9.68	0.002	455	24.93	<0.0001
7. Mainland, far island	783	18.80	<0.0001	300	14.58	0.0001	473	17.40	<0.0001
8. Continental island, oceanic island	273	5.31*	0.0212	164	0.53	0.47	85	4.11	0.04
9. Near island, far island	273	2.37	0.12	164	3.10	0.08	85	0.56	0.46
10. Mainland, island (all island endemics removed)	696	19.28	<0.0001	283	12.03	0.0005	454	21.12	<0.0001
11. Mainland, island (SC island endemics removed)	720	13.43	0.0002	291	8.85	0.003	460	14.21	0.0002

Genus was included as a random factor, except where indicated (*) in order to achieve model convergence. Significance of fixed factors was assessed by likelihood ratio tests (LR) using single term deletions. Bold *P*-values are those that were significant ($P < 0.05$) after Benjamini & Hochberg correction for 44 multiple test comparisons.

Discussion

Our study demonstrates that species present on islands are significantly less likely to be self-incompatible (SI) than species present on mainlands. This finding strongly supports Baker's (1955) hypothesis that colonization imposes a filter, in which species incapable of self-fertilization are less likely to be successful island colonists. Although a broad comparative study cannot identify processes directly, our supplemental analyses indicate that the alternative explanations of island speciation or post-colonization loss of SI do not account for the over-representation of self-compatible (SC) in island species.

Beyond documenting a pattern consistent with Baker's law, our analysis also quantifies the power of a species' self-(in)compatibility status to predict island colonization. We found that

the probability of an SI species occurring on an island is less than half that of an SC species. Thus, the power of incompatibility status for predicting island presence or absence is equal to or greater than that of other widely studied traits, such as dispersal characteristics of passerine birds (Lees & Gilroy, 2014) and flowering plants (Vazačová & Münzbergová, 2014; Vargas *et al.*, 2015). Dispersal characteristics may yield a weaker signature of colonization because of their evolutionary lability combined with strong post-arrival selection to reduce dispersal.

Although we found that SC species are over-represented on islands, we also found that islands do harbor many SI species (177 in the present study), and that a few genera have more SI than SC island species (Fig. 2). Traits other than the capacity for self-fertilization could account for the ability of these taxa to pass through Baker's filter. Self-incompatible species that have multi-

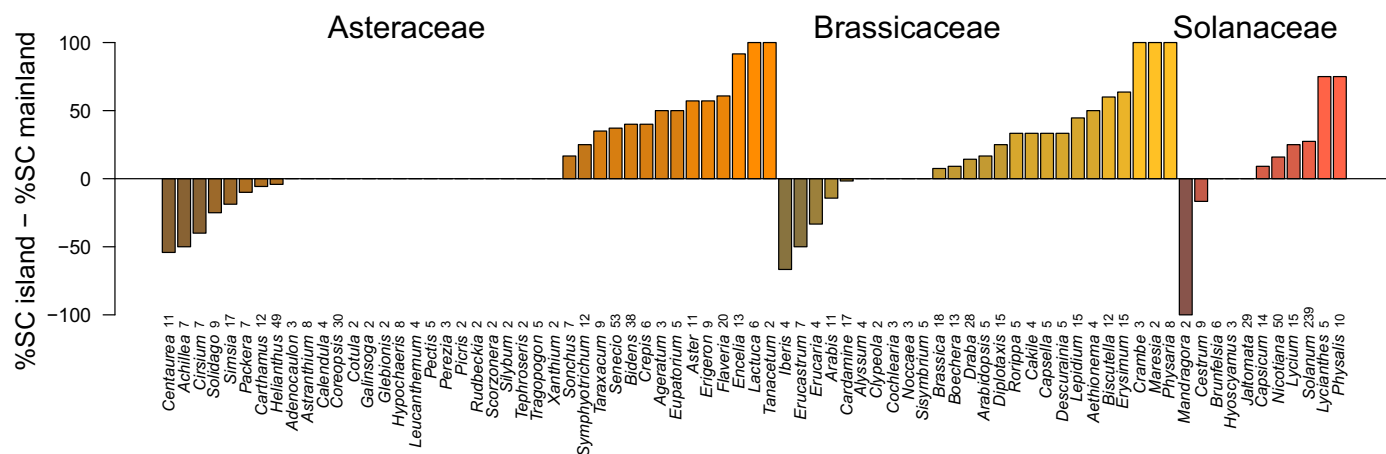


Fig. 2 For genera with at least one mainland and one island occurrence, the percentage of species that are self-compatible (SC) is greater on islands than on mainlands (Wilcoxon signed-rank test, $P < 0.001$). Numbers indicate the total number of species per genus included in the dataset.

seeded fruits, high rates of dispersal, or asexual reproduction may overcome mate limitation during colonization. For example, in the genus *Solanum*, SI is positively correlated with clonality (Vallejo-Marín & O'Brien, 2007), which could facilitate greater establishment on islands relative to nonclonal SC species. If the SI species on islands overcome mate limitation through other means, support for a colonization filter could be even stronger in analyses including additional traits that influence colonization success.

We designated each species in our analyses as either SI or SC even though polymorphism in SI status is not uncommon. Experimental error, recurrent mutation and variation among populations (Raduski *et al.*, 2012) could all underlie reports of such polymorphism. Neither experimental error nor the occurrence of occasional mutants is likely to bias our results because there is no reason to expect either to be geographically structured. Because the colonization advantage attributed to SC is not expected to depend on taxonomic rank, it is reasonable to expect that population-level differences would be in the same direction as the species-level differences we report. Of the few careful comparisons of SI in conspecific island–mainland population pairs, four of six revealed greater SC in island populations (Spears, 1987; Suehs *et al.*, 2003; Schueller, 2004; Aguiar *et al.*, 2012; Bramow *et al.*, 2013).

Islands have long served as iconic natural laboratories for illuminating the ecological and evolutionary processes shaping biological communities (Darwin, 1859; Wallace, 1902; MacArthur & Wilson, 1967; Mahler *et al.*, 2013). We find that one key reproductive trait, the capacity for self-fertilization, predicts species' distribution patterns in three plant families. These findings imply that SC also may be important in other contexts involving colonization, including metapopulation dynamics (Husband & Barrett, 1998; Pannell, 2006), range expansions (Bierzychudek, 1985; Pujol *et al.*, 2009) and invasions (Husband & Barrett, 1998; Pannell, 2006; Barrett *et al.*, 2008; van Kleunen *et al.*, 2008; Burns *et al.*, 2011). In support of this broader view, a recent global study of non-native plants found compelling evidence that SC species are more often naturalized than SI taxa

(Razanajatovo *et al.*, 2016). Available evidence thus supports SC as an important factor facilitating the successful colonization of new environments. Extending this trait-based perspective in the future will help determine whether other traits conferring improved establishment ability – such as clonality, perenniality or many-seeded fruits – also significantly influence species distributions.

Acknowledgements

This paper was produced by a working group titled 'Linking self-fertilization, dispersal, and distribution traits of species: is Baker's law an exception to the rule?' and funded by NSF (EF-0905606). This group convened at the National Evolutionary Synthesis Center (NESCent) and was led by S.K., A.M.R. and P-O.C. D.L.G. was supported by NSF (DEB-1119000). We thank Christina Flann, Eckhard von Raab-Straube and John H. Wiersema for providing geographical data.

Author contributions

D.L.G. and Y.B. led the collection and curation of occurrence records; J.W.B. and B.I. led the collection and curation of self-incompatibility records; D.L.G., Y.B., B.I., J.W.B., R.R.d.C. and E.E.G. developed and performed statistical analyses; D.L.G., Y.B., B.I., J.W.B., S.K. and E.E.G. crafted early drafts and finalized the manuscript; J.R.A., M.B., J.K.C., A.G.G., S.M.H., J.R.P., A.P., T.P., J.V., A.W., A.M.R., P-O.C. and S.K. contributed to gathering data and editing the final.

References

- Aguiar JM, Pansarin LM, Ackerman JD, Pansarin ER. 2012. Biotic versus abiotic pollination in *Oeceoclades maculata* (Lindl.) Lindl. (Orchidaceae). *Plant Species Biology* 27: 86–95.
- Auld JR, de Casas RR. 2012. The correlated evolution of dispersal and mating-system traits. *Evolutionary Biology* 40: 185–193.
- Baker HG. 1955. Self compatibility and establishment after 'long distance' dispersal. *Evolution* 9: 347–349.

- Baker HG. 1967. Support for Baker's law – as a rule. *Evolution* 21: 853–856.
- Baker HG, Cox PA. 1984. Further thoughts on dioecism and islands. *Annals of the Missouri Botanical Garden* 71: 244–253.
- Barrett SCH, Colautti RI, Eckert CG. 2008. Plant reproductive systems and evolution during biological invasion. *Molecular Ecology* 17: 373–383.
- Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48.
- Bawa KS. 1974. Breeding systems of tree species of a lowland tropical community. *Evolution* 28: 85–92.
- Bierzychudek P. 1985. Patterns in plant parthenogenesis. *Experientia* 41: 1255–1264.
- Bramow C, Hartvig I, Larsen SB, Philipp M. 2013. How a heterostylous plant species responds to life on remote islands: a comparative study of the morphology and reproductive biology of *Waltheria ovata* on the coasts of Ecuador and the Galapagos Islands. *Evolutionary Ecology* 27: 83–100.
- Burns JH, Ashman T-L, Steets JA, Harmon-Threatt A, Knight TM. 2011. A phylogenetically controlled analysis of the roles of reproductive traits in plant invasions. *Oecologia* 166: 1009–1017.
- Carlquist S. 1966. The biota of long-distance dispersal. IV. Genetic systems in the floras of oceanic islands. *Evolution* 20: 433–455.
- Cheptou P-O. 2012. Clarifying Baker's law. *Annals of Botany* 109: 633–641.
- Cheptou P, Massol F. 2009. Pollination fluctuations drive evolutionary syndromes linking dispersal and mating system. *American Naturalist* 174: 46–55.
- Crawford DJ, Lowrey TK, Anderson GJ, Bernardello G, Santos-Guerra A, Stuessy TF. 2009. Genetic diversity in Asteraceae endemic to oceanic islands: Baker's law and polyploidy. In: Funk VA, Susanna A, Stuessy TF, Bayer RJ, eds. *Systematics, evolution, and biogeography of Compositae*. Vienna, Austria: International Association for Plant Taxonomy, 139–151.
- Darwin C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London, UK: John Murray.
- Frodin DG. 2001. *Guide to standard floras of the world: an annotated, geographically arranged systematic bibliography of the principal floras, enumerations, checklists and chorological atlases of different areas*. Cambridge, UK & New York, NY, USA: Cambridge University Press.
- Husband BC, Barrett SCH. 1998. Spatial and temporal variation in population size of *Eichhornia paniculata* in ephemeral habitats: implications for metapopulation dynamics. *Journal of Ecology* 86: 1021–1031.
- Igic B, Busch JW. 2013. Is self-fertilization an evolutionary dead end? *New Phytologist* 198: 386–397.
- Igic B, Lande R, Kohn JR. 2008. Loss of self-incompatibility and its evolutionary consequences. *International Journal of Plant Sciences* 169: 93–104.
- Johnson PCD. 2014. Extension of Nakagawa & Schielzeth's R²GLMM to random slopes models. *Methods in Ecology and Evolution* 5: 944–946.
- van Kleunen M, Manning JC, Pasqualetto V, Johnson SD. 2008. Phylogenetically independent associations between autonomous self-fertilization and plant invasiveness. *American Naturalist* 171: 195–201.
- Lees AC, Gilroy JJ. 2014. Vagrancy fails to predict colonization of oceanic islands. *Global Ecology and Biogeography* 23: 405–413.
- Lloyd DG. 1965. Evolution of self-compatibility and racial differentiation in *Leavenworthia* (Cruciferae). *Contributions from the Gray Herbarium of Harvard University* 195: 3–134.
- Lord JM. 2015. Patterns in floral traits and plant breeding systems on Southern Ocean islands. *AoB Plants* 7: lv095.
- MacArthur R, Wilson EO. 1967. *The theory of island biogeography*. Princeton, NJ, USA: Princeton University Press.
- Mahler DL, Ingram T, Revell LJ, Losos JB. 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science* 341: 292–295.
- de Nettancourt D. 1977. The genetic basis of self-incompatibility. Monographs on theoretical and applied genetics. In: *Incompatibility in angiosperms*. Berlin: Springer, 28–57.
- Newstrom L, Robertson A. 2005. Progress in understanding pollination systems in New Zealand. *New Zealand Journal of Botany* 43: 1–59.
- Nowak MD, Haller BC, Yoder AD. 2014. The founding of Mauritian endemic coffee trees by a synchronous long-distance dispersal event. *Journal of Evolutionary Biology* 27: 1229–1239.
- Pannell JR. 2006. Effects of colonization and metapopulation dynamics on the evolution of plant sexual systems. In: Harder LD, Barrett S, eds. *Ecology and evolution of flowers*. New York, NY, USA: Oxford University Press, 223–238.
- 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 SCH. 1998. Baker's law revisited: reproductive assurance in a metapopulation. *Evolution* 52: 657–668.
- Patiño J, Bisang I, Hedenäs L, Dirkse G, Bjarnason ÁH, Ah-Peng C, Vanderpoorten A. 2013. Baker's law and the island syndromes in bryophytes. *Journal of Ecology* 101: 1245–1255.
- Pujol B, Zhou S-R, Vilas JS, Pannell JR. 2009. Reduced inbreeding depression after species range expansion. *Proceedings of the National Academy of Sciences, USA* 106: 15 379–15 383.
- Raduski AR, Haney EB, Igić B. 2012. The expression of self-incompatibility in angiosperms is bimodal. *Evolution* 66: 1275–1283.
- Razanajatovo M, Maurel N, Dawson W, Essl F, Kreft H, Pergl J, Pyšek P, Weigelt P, Winter M, van Kleunen M. 2016. Plants capable of selfing are more likely to become naturalized. *Nature Communications* 7: 13 313.
- Renner SS. 2014. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany* 101: 1588–1596.
- Sakai AK, Wagner WL, Ferguson DM, Herbst DR. 1995. Origins of dioecy in the Hawaiian flora. *Ecology* 76: 2517–2529.
- Schlessman MA, Vary LB, Munzinger J, Lowry PP. 2014. Incidence, correlates, and origins of dioecy in the island flora of New Caledonia. *International Journal of Plant Sciences* 175: 271–286.
- Schueler SK. 2004. Self-pollination in island and mainland populations of the introduced hummingbird-pollinated plant, *Nicotiana glauca* (Solanaceae). *American Journal of Botany* 91: 672–681.
- Singmann H, Bolker B, Westfall J, Aust F. 2016. *afex: analysis of factorial experiments*. R package version 0.16-1. [WWW document] URL <https://CRAN.R-project.org/package=afex> [accessed 1 November 2017].
- Spears EE. 1987. Island and mainland pollination ecology of *Centrosema virginianum* and *Opuntia stricta*. *Journal of Ecology* 75: 351–362.
- Stebbins GL. 1957. Self fertilization and population variability in the higher plants. *American Naturalist* 91: 337–354.
- Suehs CM, Médail F, Affre L. 2003. Invasion by South African *Carpobrotus* (Aizoaceae) taxa in the Mediterranean Basin: the effects of islands on plant reproductive systems. In: Child LE, Brock JH, Brundu G, Prach K, Pysebreve K, Wade PM, Williamson M, eds. *Plant invasions: ecological threats and management solutions*. Leiden, the Netherlands: Backhuys Publishers, 247–263.
- Vallejo-Marín M, O'Brien HE. 2007. Correlated evolution of self-incompatibility and clonal reproduction in *Solanum* (Solanaceae). *New Phytologist* 173: 415–421.
- Vargas P, Arjona Y, Nogales M, Heleno RH. 2015. Long-distance dispersal to oceanic islands: success of plants with multiple diaspore specializations. *AoB Plants* 7: plv073.
- Vazačová K, Münzbergová Z. 2014. The importance of species traits for species distribution on oceanic islands. *PLoS One* 9: e101046.
- Wallace AR. 1902. *Island life, or, the phenomena and causes of insular faunas and floras: including a revision and attempted solution of the problem of geological climates*. London, UK: Macmillan and Co.
- Warwick SI, Francis A, Al-Shehbaz IA. 2006. Brassicaceae: species checklist and database on CD-ROM. *Plant Systematics and Evolution* 259: 249–258.
- Webb CJ, Lloyd DG, Delph LF. 1999. Gender dimorphism in indigenous New Zealand seed plants. *New Zealand Journal of Botany* 37: 119–130.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Islands around the world where Baker's law has been empirically tested.

Fig. S2 Number of previously published studies by total number of species or islands examined.

Fig. S3 Schematic of geographical data pipeline with data sources for each study species and its synonyms.

Fig. S4 The distribution of breeding systems on islands vs mainlands after restricting the dataset to just those species with highest quality breeding system data.

Fig. S5 The distribution of breeding systems on continental vs oceanic islands.

Fig. S6 The distribution of breeding systems on islands near vs far from mainlands.

Fig. S7 The distribution of breeding systems on islands vs mainlands after excluding all species endemic to islands.

Fig. S8 The distribution of breeding systems on islands vs mainlands after excluding just SC species endemic to islands.

Table S1 Information extracted from publications using a Web of Science search to find all papers citing Baker (1955) and including the word 'island*' in the title or abstract

Table S2 Data table containing island information

Table S3 Data table containing species' level breeding system and geographic information

Table S4 BibTeX file of all data sources for gathering breeding system and geographic data

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.



About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <26 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**