bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Running title: Range size is predicted by mating system

Geographic range size is predicted by plant mating system

Dena Grossenbacher[1], Ryan Briscoe Runquist[1], Emma E. Goldberg[2], and Yaniv Brandvain[1]

[1]Department of Plant Biology, University of Minnesota, St. Paul, MN, 55108, USA; [2]Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN, 55108, USA

Author emails: dgrossen@umn.edu, rbriscoe@umn.edu, eeg@umn.edu, ybrandva@umn.edu

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Species' ranges vary enormously, and even closest relatives may differ in range size by several orders of magnitude. With data from hundreds of species spanning 20 genera and generic sections, we show that plant species that autonomously reproduce via self-pollination consistently have larger geographic ranges than their close relatives that generally require two parents for reproduction. Further analyses strongly implicate autonomous fertilization in causing this relationship, as it is not driven by traits such as polyploidy or annual life history whose evolution is sometimes correlated with the transition to autonomous self-fertilization. Furthermore, we find that selfers occur at higher maximum latitudes and that disparity in range size between selfers and outcrossers increases with time since their separation. Together, these results show that autonomous reproduction — a critical biological trait that eliminates mate limitation and thus potentially increases the probability of establishment — increases range size.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The enormous variation in species' geographic distributions has puzzled ecologists and evolutionary biologists for generations. Why does one species range across an entire continent while its close relative has a very narrow geographic distribution (Darwin 1859, p6; Brown et al. 1996; Gaston 2003, ch3)? There are a host of potential ecological and evolutionary explanations, including species' geographic location, age, niche breadth, environmental tolerance, competitive ability and key life history traits such as body size, dispersal ability, and mating system (reviewed by Brown et al. 1996; Gaston 1996; Gaston 2003). Most of these potential explanations have been the subject of intensive research programs across a wide range of taxa, but none are universally supported. Here we show that plant species possessing one key trait — the ability to autonomously reproduce via self-pollination — consistently have larger geographic ranges than their close relatives that generally require two parents for reproduction.

To understand the potential impacts of autonomous reproduction, or any trait, on species' range size, consider the mechanistic processes that underlie a species' range. We envision a species' range as a set of occupied points on a grid. Over time, a local population (a point) might go extinct because individuals cannot tolerate the local environment, they are outcompeted by another species, or because of a stochastic demographic event (Sexton et al. 2009). As some local populations disappear, new populations can emerge by range expansion — the process by which species spread to new geographic locations. To achieve this, individuals must first disperse to new sites. Upon arrival, individuals must tolerate any novel biotic and abiotic environmental conditions they encounter (Sexton et al. 2009) and, for most sexual species, individuals

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

must find a suitable mate in order to reproduce and establish (Baker 1955; Stebbins 1957; Pannell & Barrett 1998). Together, the action of local extinction, recolonization and range expansion generates a species' range.

Given equivalent amounts of time for range expansion, directional differences in geographic range size beyond those expected by stochasticity are likely due to traits that influence one or more of these key factors: dispersal, environmental tolerance, and mate availability. For example, increased body size in birds and mammals is thought to increase species' environmental tolerance by allowing individuals to maintain homeostasis over a wider range of environmental conditions, and it may increase dispersal ability and home-range size (Gaston 2003, p106-113). Indeed, many studies have found that body size correlates positively with species' range size, although it generally explains only a small fraction of the overall variation in range size (Gaston 2003; Agosta 2013). Similarly, self-fertilization and other forms of autonomous reproduction may affect the basic biological factors that shape range size. Although the role of autonomous reproduction in dispersal is controversial (Cheptou & Massol 2009; Hargreaves & Eckert 2014), its role in mate availability is particularly clear, and it may influence environmental tolerance.

There are two primary and opposing hypotheses regarding the effect of selfing, like other forms of autonomous reproduction, on species' geographic distributions. On the one hand, the reproductive assurance provided by selfing is predicted to increase successful colonization and establishment (Baker 1955). This is because selfers, unlike outcrossing species, are not mate-limited at low population densities and may successfully reproduce when even a single propagule lands in a new habitat (Baker

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1955; Stebbins 1957; Pannell & Barrett 1998). Under this scenario, selfers will rapidly expand their ranges and will occupy marginal habitats that support only small populations or have unpredictable pollinators. They may also face reduced risk of extinction upon colonizing new locations as the capability of autonomous reproduction can shield plants from Allee effects (reviewed in Goodwillie et al. 2005), and because the history of selfing has purged rare deleterious recessive alleles that would otherwise be exposed in small, isolated populations (Pujol et al. 2009; see Haag & Ebert 2004 for an exploration of this hypothesis in asexual populations). The logical extension of this argument is that selfers will have larger geographic ranges than outcrossers, given similar amounts of time for range expansion (Henslow 1879, p391; Lowry & Lester 2006; Randle et al. 2009).

On the other hand, the limited genetic variation within selfing species may constrain their ability to adapt to many habitats and therefore result in more limited ranges relative to their outcrossing relatives. For instance, new populations established by selfers may have reduced genetic diversity compared to populations established by outcrossers (Hamrick & Godt 1996; Crawford et al. 2008). This lack of genetic diversity may limit the niche breadth of selfing populations and prevent them from colonizing and adapting to novel environments (Stebbins 1957; Hamrick & Godt 1996; Crawford and Whitney 2010; Sheth & Angert 2014). In addition to limited genetic diversity, species that reproduce autonomously have trouble shedding deleterious mutations (Lande & Schemske 1985), and therefore competitively superior outcrossing relatives may further diminish the realized niche of selfing species. Ultimately, the evolutionary genetic consequences of self-fertilization may constrain the geographic ranges of selfers

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

relative to outcrossers (Lowry & Lester 2006; Randle et al. 2009).

Here, we test these alternative hypotheses by asking whether, across pairs of sister species of flowering plants from around the world, selfing or outcrossing plants have larger ranges. Data from 194 sister species across 20 clades consistently show that selfing species have larger ranges than their outcrossing relatives. Further analyses strongly implicate autonomous fertilization in causing this strong relationship, as it is not driven by traits such as polyploidy or annual life history whose evolution is correlated with the transition to autonomous self-fertilization. To better understand this pattern, we consider two 'follow-up' questions. We ask whether species' latitudinal distributions are affected by mating system, finding that selfers occur at higher maximum latitudes than outcrossers. Finally, with divergence time estimates from time-calibrated phylogenies, we show that the disparity in range size between selfers and outcrossers increases with the time since their separation. Together, our results show that autonomous reproduction — a critical biological trait that influences the probability of establishment — has a major influence on range size.

Materials and methods

We identified genera or generic sections that met the following criteria: 1) a published, species-level phylogeny that contained at least one predominantly selfing (or functionally selfing, e.g., asexual) species and one predominantly outcrossing species, and 2) DNA sequence data for at least 50% of the species within the clade available on GenBank (http://www.ncbi.nlm.nih.gov/genbank/). This resulted in 21 clades whose combined native distributions spanned every continent except Antarctica (see Figure S1

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in Supporting Information). On average, clades contained 35 ±7 (±1SE) extant species, 80 ±4.6 percent of which were included in our phylogenies.

For the analyses described below, all data and R scripts will be made available from the Dryad Digital Repository.

Estimating phylogenies

We generated time-calibrated phylogenies for all 21 genera or generic sections using publicly available sequence data. We reconstructed phylogenies for two reasons: 1) most previously published phylogenies consisted of only a single topology or consensus tree, making it difficult to incorporate uncertainty into our analysis, and 2) most trees were not time calibrated. Prior to estimating the phylogenies, for each clade separately, we downloaded nrITS sequences for species within the clade from GenBank and aligned them using the MUSCLE package in R, version 3.8.31-4 (Edgar 2004). We simultaneously estimated the phylogenetic relationships and the absolute divergence times among species in a Bayesian framework in BEAST version 1.6.2 (Drummond et al. 2012). To estimate absolute divergence times, we used the mean and range substitution rate for herbaceous and woody plants at the nuclear ribosomal internal transcribed spacer locus (nrITS) (Kay et al. 2006), because fossils are not known for any of the clades in the analysis. The substitution rate was set to a normally distributed prior for herbaceous lineages with mean of 4.13×10^{-9} subs/site/yr and standard deviation of 1.81×10^{-9}, and for woody lineages with mean of 2.15×10^{-9} subs/site/yr and standard deviation of 1.85×10^{-9}.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To accommodate heterogeneity in the molecular evolutionary rate among branches, we used an uncorrelated log-normal (UCLN) relaxed clock model. The prior model on branch lengths was set to a Yule process of speciation. The prior model on substitutions and the number of MCMC generations varied by clade (see Table S1). Posterior samples of parameter values were summarized and assessed for convergence and mixing using Tracer v. 1.5 (Rambout et al. 2014). Of the 21 clades identified above, all met convergence except for *Leavenworthia*, which was excluded from subsequent analyses. For the remaining 20 clades, the minimum estimated sum of squares (ESS) for the posterior was >1100, and the minimum ESS across all other parameters was >600 (Table S1).

For all ensuing analyses, we identified sister species in a subset of 9000 trees from the posterior distribution for each clade. For each sister species pair, we recorded the average divergence time and the posterior probability of that pair as the proportion of trees that contained that pair, providing a measure of phylogenetic uncertainty. Since our phylogenies sampled, on average, only 80% of extant taxa, these sister pairs may not represent “true” extant sisters, but they are recently diverged groups representing independent evolutionary replicates.

Estimating mating system, ploidy, and life history

We collated 54 studies that described the mating systems of species from the 20 genera or generic sections identified above. Most published studies classified species as predominantly selfing, mixed mating, or predominantly outcrossing. An exception to this were species in *Oenothera* sect. *oenothera,* which were classified as either

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

functionally asexual, due to a permanent translocation whereby plants self-fertilize but do not undergo segregation and recombination, or sexual (Johnson et al. 2009). Sexual *Oenothera* sect. *oenothera* species are partially or wholly self-incompatible, and are assumed to be outcrossing relative to the asexual species. Different traits are more reliable indicators of mating system in different taxa, and so the methods for mating system classification sometimes varied among clades, but they were generally consistent within clades (described in Table S2). To extend our data set, we occasionally classified taxa that were missing from the primary studies using the same traits and metrics as those used for other species within that clade (see Table S2). Only sister pairs with one selfing and one outcrossing species were included in the ensuing analyses, hereafter termed "selfing-outcrossing sister pairs".

While our focus is on mating system, correlated traits such as polyploidy (Stebbins 1950; Barringer 2007, Robertson et al. 2011) and perennial or annual life history (Barrett et al. 1996) may coevolve with mating system. To ensure that these traits did not drive a relationship between mating system and range size, we gathered published information on ploidy and life history when possible. Species were classified as diploid, polyploid, or mixed when both diploid and polyploid individuals were known. Species' life histories were classified as annual, perennial, or mixed when both annual and perennial individuals were known.

Estimating geographic range size and latitudinal distributions

We downloaded all known species occurrence records for the clades from the Global Biodiversity Information Facility (http://www.gbif.org) and filtered for quality by excluding

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

records with coordinate accuracy <100 km, coordinates failing to match the locality description, and taxonomic misidentifications (verified by the authors and taxonomic specialists of each clade). We checked species' epithets against the most recently published taxonomies and corrected synonyms and spelling errors. We included only coordinates from the native range of species. Coordinates outside the native species range were identified using published monographs and online databases that report native and invaded ranges (e.g., GRIN database, http://www.ars-grin.gov/).

We used the filtered occurrence data to estimate species' range size using a grid approach. In this approach, we divided the world into a series of rectangular cells by grid lines that follow degree longitude and latitude using the "raster" R package version 2.3-0 (Hijmans et al. 2011). We calculated range size as the summed area of occupied grid cells for a given species. In order to assess whether the ensuing analyses were sensitive to the spatial scale at which species' ranges are estimated, range size was calculated across a range of cell sizes, 0.05, 0.1, 0.5 and 1 decimal degrees, representing grid cells of approximately 25, 100, 2500, and 10000 km^2 respectively.

In addition to range size, we quantified three components of species' latitudinal distributions from the filtered occurrence data: absolute minimum latitude, absolute midpoint latitude (midpoint between minimum and maximum latitude), and absolute maximum latitude. None of the species in our dataset have distributions spanning the equator, so minimum latitude was always greater than zero.

Analyses

To determine whether selfers have larger or smaller ranges than their outcrossing sister

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

species, we performed linear mixed effects models using the *lme4* R package (Bates et al. 2012). We performed all analyses 4 times using the 4 different range estimates (described above) to ensure that our results were robust to the spatial scale at which range size was determined. We also ran all analyses including only selfing-outcrossing sister pairs that did not differ in ploidy (N=127) or annual/perennial life history (N=112) to ensure that our results were not being driven by these potential correlated traits. For all analyses, we used natural log-transformed range size as the dependent variable to improve normality of residuals and homogeneity of variance. We treated mating system as a fixed effect and included random intercepts for sister pairs and clades, as well as by-clade random slopes. To incorporate phylogenetic uncertainty into our model, we included a weighting factor for each sister pair that was equal to their posterior probability (the proportion of phylogenetic trees that contained a given sister pair). 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) were calculated according to Johnson (2014). Significance was determined using likelihood ratio tests with single term deletions implemented in the R package, *afex* (Singmann 2014).

To determine whether selfers occupy different latitudes than their outcrossing sister species, we used the same basic model described above with absolute minimum, midpoint, and maximum latitude treated as response variables in 3 separate models.

To determine whether range size or latitude were affected by divergence time and whether this varied by mating system, we added divergence time and its interaction with mating system as fixed effects to the above model, as well as by-clade random

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

slopes for divergence time. To meet model assumptions, divergence time was natural log-transformed prior to analysis.

Results

The phylogenetic analysis identified 194 sister species across 20 clades that differed in mating system. Within clades, the number of selfer-outcrosser sister pairs ranged from 1-68, and their posterior probabilities ranged from <0.01 – 1.0.

Selfing-outcrossing sister species displayed tremendous variation in range size, from having nearly equivalent range areas to differing by more than 3 orders of magnitude. Mating system shifts explained a significant proportion of this variation (16-21%) with selfers having, on average, ~1.5-2 times larger ranges than their outcrossing sister species (Table 1; Fig 1). This effect was significant across all spatial scales, and it was robust to excluding sister pairs that differed in ploidy and annual/perennial life history ($P \leq 0.01$ in all cases; see Table S3).

Mating system also explained the latitudinal distributions of sister species. Selfers had higher maximum latitudes relative to their outcrossing sister species by ~1 decimal degree on average (Table 1; Fig 2A). In contrast, midpoint and minimum latitudes did not vary by mating system (Table 1; Fig. 2B,C). When we excluded sister pairs that differed in ploidy or annual/perennial life history, the direction and magnitude of the effect was similar but it was no longer significant for maximum latitude ($P > 0.22$, see Table S2). This loss of significance likely reflects a simple loss of power, but it could also be attributable to polyploidy and annual/perennial life history directly influencing species' latitudinal distributions.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

With increasing time since divergence, selfing species tended to increase their ranges more than their outcrossing sister species (Table 2; Fig 3). In contrast, the latitudinal distributions of species were not significantly affected by divergence time, nor by the interaction between divergence time and mating system (Table 2).

Discussion

Our analysis across 20 clades reveals that selfers have geographic ranges on average twice the size of their outcrossing sister species. This difference increases with divergence time between selfing-outcrossing pairs, and it may be partially attributable to the colonization of high latitude regions by selfers. Together, our results provide strong support for the idea that the increased colonization ability in selfers associated with the escape from mate limitation (Baker 1955; Stebbins 1957; Pannell & Barrett 1998) allows selfers to expand and extend their geographic ranges relative to outcrossers (Randle et al. 2009). It appears that this effect overwhelms potentially opposing forces, such as genetic constraints, that may limit the geographic ranges of selfers relative to outcrossers. We note that although this correlative study cannot fully exclude alternative interpretations of this pattern, our results are not driven by two plausibly explanatory traits, annual life history and polyploidy, which are often correlated with self-fertilization. Another potentially correlated trait is dispersal ability. Although we do not assess traits related to dispersal, theory suggests predominant selfers may have narrower dispersal kernels, going against the results found here (Cheptou & Massol 2009; Hargreaves & Eckert 2014).

Henslow (1879) was the first to observe a relationship between mating system

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and geographic range size when he noted that most widespread members of the British flora tended to be selfers. However, it was more than 100 years until this pattern was systematically tested in three North American plant genera, with mixed results. In *Collinsia*, Randle et al. (2009) found that automatically selfing species had larger ranges than their outcrossing sister species. The same trend was found in *Oenothera* sect. *Oenothera*, but it was not statistically significant and relied on very different comparative methods (Johnson et al. 2009). In *Clarkia*, the trend was variable depending on how the data were subsetted, possibly due to poor phylogenetic resolution or to confounding correlated traits such as ploidy (Lowery & Lester 2006). All three genera were included in the present study and supported the overall pattern of greater range size for selfing than outcrossing sister species (Fig 1).

Our results have implications regarding the niche breadth of selfing species. In contrast to the idea that reduced genetic variation in selfers will prevent them from establishing and adapting in new environments (Stebbins 1957; Hamrick & Godt 1996; Crawford & Whitney 2010), we found that selfers had larger geographic ranges, achieved in part by occupation of higher latitudes. Latitude is one of the most extreme environmental gradients on the planet, with high latitudes experiencing high seasonality and low biotic diversity relative to low latitudes (reviewed in Mittelbach et al. 2007). The ability to persist across a range of latitudes suggests a greater realized species-wide niche breadth for selfers, which may be partially attributable to ploidy and/or life history. It may or may not, however, reflect an underlying greater environmental tolerance of selfers versus outcrossers. A recent study of thermal tolerance in monkeyflowers included two selfing-outcrossing sister species pairs and found mixed results (Sheth &

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Angert 2014). In one pair, the selfer had greater within-population genetic variation and increased thermal tolerance, while in the other pair the selfer had less genetic variation and reduced thermal tolerance. In both these cases however, only a single population was assessed so it is unknown whether these selfing species as a whole (across all populations) can tolerate more environmental conditions than closely related outcrossing species. Further exploration of the breadth of environments occupied by selfers and outcrossers may uncover axes along which selfers have expanded their realized environmental niches, as well as provide hypotheses concerning their environmental tolerances.

Like self-fertilization, other forms of uniparental reproduction, such as asexuality, may allow plants to evade Allee effects during range expansion and achieve larger ranges. Indeed, the ranges of asexual plants are sometimes larger than the ranges of their sexual relatives (Bierzychudek 1985), and asexual species sometimes occur at extreme latitudes (Bell 1982; Bierzychudek 1985). Numerous hypotheses have been put forth to explain the correlation between asexuality and latitude. For example, asexuals may have ‘generalist genotypes’ and broader environmental tolerance (Lynch 1984), or fluctuating biotic interactions with pathogens and competitors in tropical regions maintains sex in low but not high latitudes (Bell 1982; Glesner & Tilman 1978). Fully excluding these alternative explanations requires experimental work as well as additional correlational studies. For example, evidence of large ranges in pseudogamous apomicts (plants that, for their own asexual propagation, require pollen which is often not their own (Ho□randl 2010)) would argue against our hypothesis that reproductive assurance acts to increase range size.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The age-range relationship, speciation and extinction

The observation that range size increases with time since most recent divergence has been made in several previous studies, although this effect varies widely across taxonomic groups (reviewed in Pigot et al. 2012). Here, we found a similar effect, but only for selfing species (Fig. 3). Although it is tempting to interpret this as evidence of strong directional range size evolution for selfers relative to outcrossers, we caution that the geography of speciation and filtering effects of extinction could also contribute to this pattern. For instance, species that inherit large ranges across speciation events are free to shrink and expand their range considerably; in contrast, species that inherit a small range during speciation will go extinct if their range shrinks considerably. This bias will cause an apparent increase in range size with age, at least for relatively young ages like we consider, simply because species with decreased ranges were lost to extinction. Thus the geography of speciation and the pace of range size evolution can introduce artificial trends in range size evolution (Pigot et al., 2012). This process could be relevant for speciation events involving selfing - outcrossing pairs (e.g., if selfers commonly arise in small populations via budding speciation); however, these ideas have not been explicitly modeled in this context, and therefore precise predictions and directions of biases are unclear.

Our finding that selfers have larger ranges than outcrossers, and that ranges of selfing species increase with age, seems at odds with the long-held idea that selfers are evolutionary dead ends and face high extinction rates relative to outcrossers (reviewed in Igic & Busch 2013). One possible resolution is that selfers do in fact go extinct more frequently, but the ones we observe are those that happen to attain large range size (as

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in the example above). Another resolution is that extinction in selfing plants occurs by rapid extirpations across the entire range, rather than a gradual elimination of populations until the range dwindles and disappears. The rapid extinction of a selfing species across a large range could be due to rapid fluctuation of the geographic distribution of selfers, or to stronger autocorrelation of extinction risk across the range. For example, if selfers have less genetic variation (Stebbins 1957; Hamrick & Godt 1996; Crawford & Whitney 2010; Sheth & Angert 2014) and steadily accumulate deleterious mutations (Lande & Schemske 1985), they may be vulnerable to sudden environmental changes (e.g., novel parasites or climate change) leading to rapid, range-wide extirpation and extinction. A final potential resolution could be the combination of the phylogenetic scale over which selfing species go extinct (e.g., selfers may give rise to other selfers prior to going extinct) and ascertainment bias on the scale of our study. There are many species, both selfing and outcrossing, whose sister species is not of the opposite mating system and which are consequently not included in our dataset.

Other contributors to geographic range size

The search for reasons underlying the massive variation in species' geographic range sizes has a long history that reveals few, if any, universal explanations (Brown et al. 1996; Gaston 1996; Gaston 2003). The results here suggest that mating system is a powerful predictor of range size, explaining up to 20% of the variation in species' geographic ranges. To put this in perspective, body size is one of the best-studied correlates of range size and only explains about 6% of the overall variation in the range size of birds and mammals (average across studies; Agosta et al. 2013), which is typical

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of other predictors of range size (Brown et al. 1996). This highlights that the majority of variation in range size is likely attributed to other factors. We note that the proportion of variation in range size that we explain is likely an overestimate: we chose clades in which mating system was sufficiently variable so as to potentially explain a reasonable portion of the variation. It is not clear how this ascertainment scheme compares to previous investigations of a trait's influence on range size.

In addition to mating system, recent shared ancestry may also play a role in shaping species' geographic ranges (Jablonski 1987; Bohning-Gaese et al. 2006; Martin & Husband 2009). For instance, species may be likely to arise in the same region of the world as their close relatives, which in turn may influence range size, e.g., Rapoport's Rule (Stevens 1989). Additionally, life history traits that are shared due to recent common ancestry (e.g., dispersal characters, size, and traits related to environmental tolerance) may influence species' ranges (reviewed in Brown et al. 1996; Gaston 1996; Gaston 2003). Indeed, in our study, recent shared ancestry at the level of genus and sister pair accounted for a large part of the overall variation in range size. Together, shared ancestry and mating system explained 90-95% of the variation in species' geographic ranges in the present study (see Table 1).

Conclusion

Our observation that selfers have larger ranges than outcrossers is consistent with the idea that mate availability at the colonization stage is likely a major limiting factor on species' range size. This implies that the simple ability to find a mate and establish in a novel habitat may have as great an influence on the species' range as environmental

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tolerance or interspecific competition. Furthermore, this suggests that any trait that increases the odds of finding a mate during colonization, selfing or otherwise (e.g., sperm storage in females), may result in increased geographic range size. Unfortunately, despite the critical role of mate limitation in slowing range expansions (Shaw & Kokko *in press*), studies assessing the impact of mate availability on range size in animals or other taxa are lacking, as much of the last century of research has instead focused on traits influencing dispersal and environmental tolerance (reviewed in Brown et al. 1996; Gaston 1996; Gaston 2003). We suggest that in animals and other taxa, focusing on traits that encourage mate-finding during colonization may be central to understanding the puzzle of geographic range size.

Acknowledgements

We thank Boris Igic, April Randle, Sue Kalisz, and the Baker's Law working group for stimulating discussions. We thank Anne Worley, Barbara Neuffer, Andress Franzke, Jeremiah Busch, and Justen Whittall for expert advice on mating systems and phylogeny construction.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Results of 7 separate linear mixed models analyzing the effect of mating system on species' range size (estimated at 4 spatial scales) and latitudinal distributions.

Response	LR	*P*	Marginal R^2	Conditional R^2	Predicted value	
					Selfer	Outcrosser
Natural log range size						
~25 km^2 grid cell	12.18	<0.001	0.208	0.950	1979	940
~100 km^2 grid cell	11.85	<0.001	0.199	0.947	6494	3224
~2500 km^2 grid cell	10.90	0.001	0.182	0.936	71,232	40,921
~10,000 km^2 grid cell	11.15	<0.001	0.159	0.930	174,312	108,597
Absolute latitude						
minimum	0.20	0.65	0.002	0.976	31.95	32.31
midpoint	1.61	0.20	0.007	0.990	37.49	36.75
maximum	3.86	0.05	0.024	0.989	43.32	41.56

Significance of fixed effects was assessed by likelihood ratio tests (LR) using single term deletions. Marginal R^2 values are the proportion of variance explained by mating system (fixed effect). Conditional R^2 values are the variance explained by mating system and the random effects of clade and sister pair. Predicted values for range size are back-transformed.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Results of 4 separate linear mixed models analyzing the effect of divergence time, mating system, and their interaction on species' range size and latitude.

Response	Fixed effects					
	Divergence time		Mating system		Time*Mating System	
	LR	*P*	LR	*P*	LR	*P*
Natural log range size	3.86	0.05	4.67	0.03	54.37	0.001
Minimum latitude	2.44	0.12	0.53	0.47	2.71	0.10
Midpoint latitude	0.00	0.99	0.00	0.99	2.02	0.16
Maximum latitude	0.00	0.99	3.07	0.08	2.90	0.09

Significance of fixed effects was assessed by likelihood ratio tests (LR) using single term deletions. Range size was estimated in ~100 km^2 grid cells; results are consistent across all spatial scales (not presented).

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Agosta, S.J., Bernardo, J., Ceballos, G. & Steele, M. A. (2013). A macrophysiological analysis of energetic constraints on geographic range size in mammals. *PLoS One*, 8, e72731.

2. Alarcón, M.L., Roquet, C. & Aldasoro, J.J. (2011). Evolution of pollen/ovule ratios and breeding system in Erodium (Geraniaceae). *Syst. Bot.*, 36, 661–676.

3. Baker, H.G. (1955). Self compatibility and establishment after long distance dispersal. *Evolution*, 9, 347–349.

4. Barrett, S.C.H., Harder, L.D. & Worley, A.C. (1996). The comparative biology of pollination and mating in flowering plants. *Philos. Trans. Biol. Sci.*, 1271–1280.

5. Barringer, B. (2007). Polyploidy and self-fertilization in flowering plants. *Am. J. Bot.*, 94, 1527–1533.

6. Bates, Douglas; Maechler, Martin; Bolker, B. & Walker, S. (2014). {lme4}: Linear mixed-effects models using Eigen and S4. *R Packag. version 1.1-7*.

7. Bell, G. (1982). *The masterpiece of nature: the evolution and genetics of sexuality*. CUP Archive.

8. Bierzychudek, P. (1985). Patterns in plant parthenogenesis. *Experientia*, 41, 1255–1264.

9. Blackburn, T.M. & Gaston, K.J. (1996). Spatial patterns in the species richness of birds in the New World. *Ecography*, 19, 369–376.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10. Böhning-Gaese, K., Caprano, T., van Ewijk, K. & Veith, M. (2006). Range size: disentangling current traits and phylogenetic and biogeographic factors. *Am. Nat.*, 167, 555–567.

11. Brown, J.H., Stevens, G.C. & Kaufman, D.M. (1996). The geographic range: size, shape, boundaries, and internal structure. *Annu. Rev. Ecol. Syst.*, 27, 597–623.

12. Cheptou, P. & Massol, F. (2009). Pollination fluctuations drive evolutionary syndromes linking dispersal and mating system. *Am. Nat.*, 174, 46–55.

13. Crawford, D.J., Archibald, J.K., Stoermer, D., Mort, M.E., Kelly, J.K. & Santos-Guerra, A. (2008). A test of Baker's law: breeding systems and the radiation of Tolpis (Asteraceae) in the Canary Islands. *Int. J. Plant Sci.*, 169, 782–791.

14. Crawford, K.M. & Whitney, K.D. (2010). Population genetic diversity influences colonization success. *Mol. Ecol.*, 19, 1253–1263.

15. Darwin, C. (1859). On the origin of species by means of natural selection. *Murray, London*.

16. Drummond, A.J., Suchard, M. a, Xie, D. & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.*, 29, 1969–73.

17. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32, 1792–1797.

18. Gaston, K. (1996). Species-range-size distributions: patterns, mechanisms and implications. *Trends Ecol. Evol.*, 11, 197—201.

19. Gaston, K.J. (2003). *The structure and dynamics of geographic ranges*. Oxford University Press.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. Glesener, R.R. & Tilman, D. (1978). Sexuality and the components of environmental uncertainty: clues from geographic parthenogenesis in terrestrial animals. *Am. Nat.*, 659–673.

21. Goodwillie, C., Kalisz, S. & Eckert, C.G. (2005). The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.*, 36, 47–79.

22. Haag, C.R. & Ebert, D. (2004). A new hypothesis to explain geographic parthenogenesis. *Ann. Zool. Fennici*., 41, 539–544.

23. Hamrick, J.L. & Godt, M.J.W. (1996). Effects of life history traits on genetic diversity in plant species. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.*, 351, 1291–1298.

24. Hargreaves, A.L. & Eckert, C.G. (2014). Evolution of dispersal and mating systems along geographic gradients: implications for shifting ranges. *Funct. Ecol.*, 28, 5–21.

25. Henslow, G. (1879). On the self□fertilization of plants. *Trans. Linn. Soc. London*, 1, 317–398.

26. Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2011). dismo: Species distribution modeling. R package version 0.7–23.

27. Hörandl, E. (2010). The evolution of self-fertility in apomictic plants. *Sex. Plant Reprod.*, 23, 73–86.

28. Igic, B. & Busch, J.W. (2013). Is self□fertilization an evolutionary dead end? *New Phytol.*, 198, 386–397.

29. Jablonski, D. (1987). Heritability at the species level: Analysis of geographic ranges of cretaceous mollusks. *Science*, 238, 360–363.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

30. Johnson, M.T.J., Smith, S.D. & Rausher, M.D. (2009). Effects of plant sex on range distributions and allocation to reproduction. *New Phytol.*, 186, 769–79.

31. Johnson, P.C.D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods Ecol. Evol.*, 5, 944–946.

32. Kay, K.M., Whittall, J.B. & Hodges, S.A. (2006). A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evol. Biol.*, 6, 36.

33. Lande, R. & Schemske, D.W. (1985). The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution*, 39, 24–40.

34. Lowry, E. & Lester, S.E. (2006). The biogeography of plant reproduction: potential determinants of species' range sizes. *J. Biogeogr.*, 33, 1975–1982.

35. Lynch, M. (1984). Destabilizing hybridization, general-purpose genotypes and geographic parthenogenesis. *Q. Rev. Biol.*, 257–290.

36. Martin, S.L. & Husband, B.C. (2009). Influence of phylogeny and ploidy on species ranges of North American angiosperms. *J. Ecol.*, 97, 913–922.

37. Mittelbach, G.G., Schemske, D.W., Cornell, H. V, Allen, A.P., Brown, J.M., Bush, M.B., *et al.* (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.*, 10, 315–331.

38. Pannell, J.R. & Barrett, S.C.H. (1998). Baker's law revisited: reproductive assurance in a metapopulation. *Evolution*, 657–668.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39. Pigot, A.L., Owens, I.P.F. & Orme, C.D.L. (2012). Speciation and extinction drive the appearance of directional range size evolution in phylogenies and the fossil record. *PLoS Biol.*, 10, e1001260.

40. Pujol, B., Zhou, S.-R., Vilas, J.S. & Pannell, J.R. (2009). Reduced inbreeding depression after species range expansion. *Proc. Natl. Acad. Sci.*, 106, 15379–15383.

41. Rambaut, A., Suchard, M., Xie, D. & Drummond, A. (2014). Tracer v1.6. *Available from http//beast.bio.ed.ac.uk/Tracer*.

42. Randle, A., Slyder, J. & Kalisz, S. (2009). Can differences in autonomous selfing ability explain differences in range size among sister taxa pairs of Collinsia (Plantaginaceae)? An extension of Baker's. *New Phytol.*, 618–629.

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

44. Sexton, J.P., McIntyre, P.J., Angert, A.L. & Rice, K.J. (2009). Evolution and ecology of species range limits. *Annu. Rev. Ecol. Evol. Syst.*, 40, 415–436.

45. Shaw, A.K. & Kokko, H. 201X. Dispersal evolution in the presence of Allee effects can speed up or slow down invasions. *Am. Nat.*

46. Sheth, S.N. & Angert, A.L. (2014). The evolution of environmental tolerance and range size: a comparison of geographically restricted and widespread Mimulus. *Evolution*, 2917–2931.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47. Singmann, H. (2014). afex: Analysis of factorial experiments. R package version 0.9-109.

48. Stebbins, G.L. (1950). *Variation and evolution in plants*. Columbia University Press, New York, London.

49. Stebbins, G.L. (1957). Self fertilization and population variability in the higher plants. *Am. Nat.*, 337–354.

50. Stevens, G.C. (1989). The latitudinal gradient in geographical range: how so many species coexist in the tropics. *Am. Nat.*, 240–256.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1.

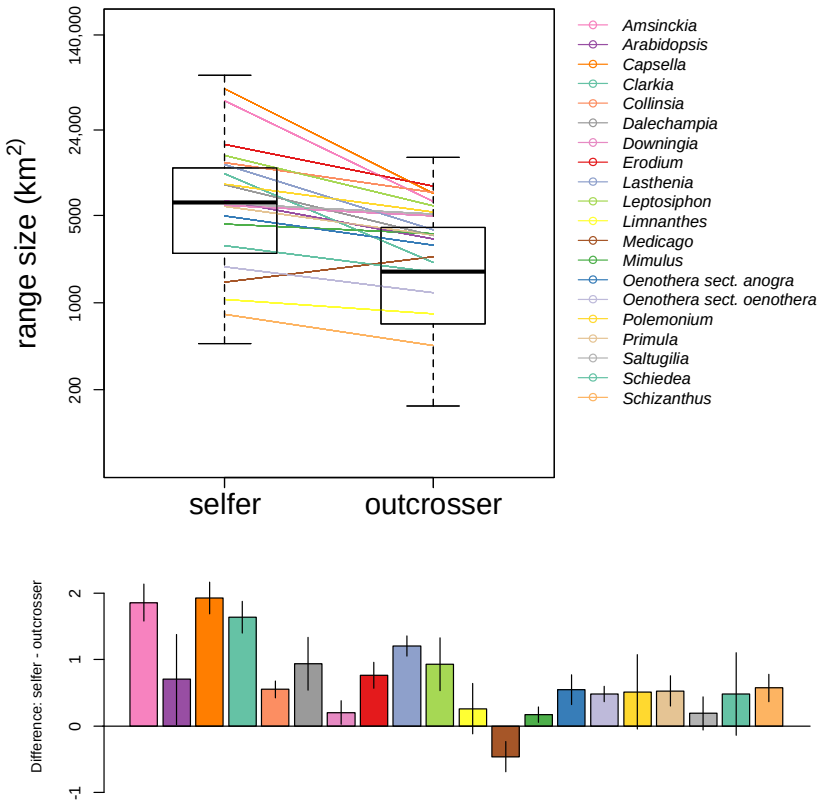

Figure 1. Top panel: Box plots of predicted range size of selfing and outcrossing sister species at the ~100 km^2 grid cell spatial scale. Colored line segments indicate predicted slopes for the 20 clades. Vertical axis is natural logarithmic scale (back-transformed km^2). Bottom panel: The predicted average difference in range size (ln(selfer)– ln(outcrosser)) for each of 20 clades, with vertical lines representing standard errors. See Table 1 for statistical results.

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2.

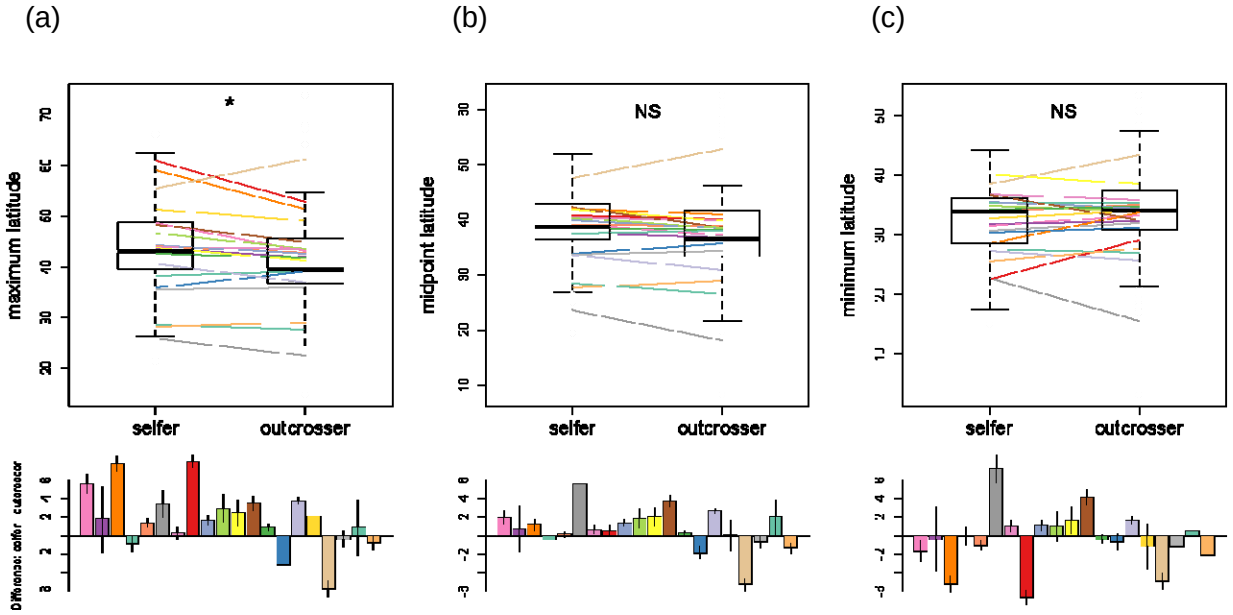

Figure 2. Box plots of predicted absolute latitudinal distributions of selfing and outcrossing sister species, (a) maximum latitude, (b) midpoint latitude, (c) minimum latitude. Colored line segments indicate predicted slopes for the 20 clades (see Fig. 1 legend). Bottom bar charts represent the predicted average difference in latitude (selfer - outcrosser) for each of 20 clades, with vertical lines representing standard errors. * P <0.05 (see Table 1 for full statistical results)

bioRxiv preprint doi: https://doi.org/10.1101/013417; this version posted January 5, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3.

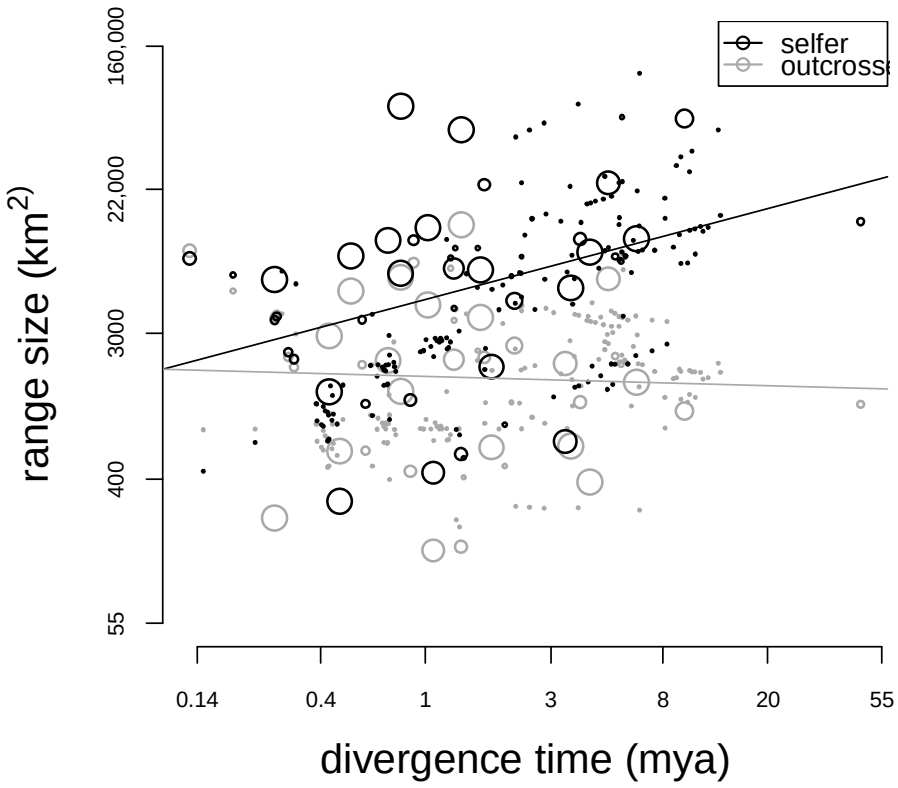

Figure 3. Range size by divergence time for selfing and outcrossing sister species. The size of open circles represents the confidence that species pairs are sisters. The line segments represent the linear regression results for selfers and outcrossers (black and gray lines respectively). Range size and divergence time axes are natural logarithmic scale (back-transformed). See Table 2 for statistical results.