THE PHYLOGENETIC DISTRIBUTION AND FREQUENCY OF SELF-INCOMPATIBILITY IN FABACEAE

Lucy E Delaney[1,*] and Boris Igić*

*Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60607, USA

Editor: Hervé Sauquet

Premise of research. Surprisingly little is known regarding the phylogenetic distribution and frequency of self-incompatibility (SI) in some of the largest angiosperm families, including the legumes (Fabaceae). The ecological and economic significance of the family has motivated an enormous but scattered output of literature concerning its breeding systems and reproductive biology. Since the last synthesis of this literature, we have gained a clearer understanding of the various mechanisms responsible for SI, but data on their number and phylogenetic distribution remain limited.

Methodology. We compiled species-level information on legume breeding systems, pollination syndromes, longevity, stature, stem composition, geographic location, and ploidy. The resulting database contains entries for 1167 unique species across 184 genera, accounting for nearly 5% of the family. Using these data, we assess the phylogenetic distribution and frequency of SI in Fabaceae and evaluate the potential correlation between SI and three other traits: ploidy, stem composition, and longevity.

Pivotal results. We estimate that approximately one-fifth of legume species have SI, with the remainder being self-compatible. Appropriate controlled pollinations are available for only a small portion of species in our database, and genetic studies are limited to fewer than 0.5% of the family. Moreover, our survey reveals no convincing evidence that a single SI mechanism operates in Fabaceae. Despite uneven coverage and a sample of ca. 5% of Fabaceae, the results of our phylogenetic analyses strongly suggest that longevity is statistically significantly correlated with breeding systems, while ploidy is not.

Conclusions. Our collection reveals a great variety of SI expressions in the family. Compelling evidence indicates that one of the best-studied SI systems—RNase-based SI—is homologous across core eudicots and ancestral to the group. Previous studies surmise that this system may operate in Fabaceae, but this remains an open question. Whatever its genetic causes among legumes, the reduced seed and fruit set after selfing appears to evolve dynamically and exhibit uncommon variation. We highlight the need for empirical work evaluating SI and outline areas where future efforts may provide disproportionate rewards.

Keywords: self-incompatibility, breeding systems, legumes, life history traits, polyploidy.

Online enhancements: supplemental PDF and data.

Introduction

Unlike most sexually reproducing groups, the great majority of flowering plants are hermaphrodites. Although they harbor potential for self-fertilization, flowering plants also often maintain mechanisms that reduce selfing or prevent it entirely (Barrett 1998). Such mechanisms are part of the enormous variation observed in the anatomical and physiological features, collectively known as plant breeding systems, that govern patterns of reproduction in angiosperms (Neal and Anderson 2005). While some features reduce selfing by morphological or temporal means (e.g., herkogamy or dichogamy), in the present article we focus on self-incompatibility (SI)—a general term for a suite of postmating, prezygotic, physiological mechanisms that allow a plant to recognize and reject its own pollen. SI is one of the most common and effective means by which higher plants prevent self-fertilization, but unlike features such as herkogamy or dioecy, SI is largely an invisible trait; there is rarely any obvious characteristic morphology associated with it. While all SI systems involve the coordinated action of male and female components facilitating the recognition and rejection of self-pollen, several different mechanisms

[1] Author for correspondence; email: ldelan5@uic.edu.

Manuscript received June 2021; revised manuscript received September 2021; electronically published November 12, 2021.

 © 2021 The University of Chicago. All rights reserved. Published by The University of Chicago Press. https://doi.org/10.1086/717329

with distinct genetic and molecular underpinnings are known to exist (Silva and Goring 2001).

Current estimates indicate that a variety of SI mechanisms evolved at least 35 times, and this is likely a great underestimate (Igić et al. 2008). The underlying molecular machinery generally functions to inhibit germination of self-pollen grains on the stigma by preventing elongation of the self-pollen tube within the style, often controlled by a single multiallelic locus called the S-locus (de Nettancourt 2001). Gametophytic SI (GSI) is the most common type of SI and involves the rejection of self-pollen as determined by the pollen haplotype. One particular flavor of GSI (RNase-based SI [RSI]) is found in several distantly related core eudicot families—including Cactaceae, Plantaginaceae, Rosaceae, Rubiaceae, Rutaceae, and Solanaceae—with astonishing genetic and molecular similarities, evidence that this system may be homologous and therefore ancestral to core eudicots (Igić and Kohn 2001; Steinbachs and Holsinger 2002; Ramanauskas and Igić 2017, 2021; Liang et al. 2020).

The evolution of SI is thought to be associated with the increased evolutionary success of angiosperms (Zavada and Taylor 1986; Franklin-Tong and Franklin 2003; Goldberg et al. 2010). Traits that influence mating patterns, including genetic relatedness among mated pairs, affect the amount and the distribution of genetic variation in space and over time (Barrett 2013). Such traits have broad consequences for the evolutionary pathways available to organisms. Thus, the presence or absence of SI has two important effects. First, strict outcrossing achieved by SI may commonly increase net diversification rates, the balance of speciation and extinction rates (Goldberg et al. 2010). Second, patterns of genetic variation determine the raw material available for the action of natural selection, so SI also ought to influence the distribution of all other traits.

Indeed, SI has long been observed to causally or indirectly correlate with other traits, including ploidy (Charlesworth 1985), stem composition (Arroyo 1981; Barrett 1998), stature (Scofield and Schultz 2006), and longevity (Hamrick and Godt 1996), among others. Associations between breeding system and traits related to life span, like stem composition and longevity, are thought to arise partly as a result of the increased levels of genetic load present in long-lived plant species (Scofield and Schultz 2006). It is generally observed that woody and perennial species are disproportionately likely to express some form of SI (Barrett 1998; Friedman 2020). Annual species, many of which are weedy, instead predominantly self-fertilize, partly because of load dynamics (Lande and Schemske 1985) but perhaps also to provide reproductive assurance to such species during the relatively brief periods when mating is possible (Hamrick and Godt 1996). Similarly, there appears to be a strong association between increased selfing and polyploidy (Grant 1956; Barringer 2007). This association may likewise provide reproductive assurance for recent polyploids that may be reproductively isolated from their diploid progenitors, in addition to reducing the deleterious effects of high genetic load by expanding the number of gene copies available to an individual organism (reviewed in Ramsey and Schemske 1998). Polyploidy can causally disrupt the molecular mechanism responsible for RSI (Stout and Chandler 1942; Stone 2002), and evidence of correlated evolution of self-compatibility (SC) and polyploidy has been observed across groups with RSI (Miller and Venable 2000; Robertson et al. 2011). Comparative analyses of this kind are valuable tools that provide insight into the forces shaping trait distribution among angiosperms, but such analyses are not possible without good data.

There are enormous gaps in breeding system data in many families that are especially relevant in determining the phylogenetic distribution and frequency of SI across angiosperms. Notably, the molecular mechanism(s) operating in Fabaceae and their phylogenetic distribution within the family remain unknown, despite many years of active literature (e.g., Sims et al. 1999; Casey et al. 2010; Aguiar et al. 2015). Fabaceae contains nearly 7% of all flowering plant species, is abundantly distributed everywhere on earth, except in the extreme polar regions, and exhibits a practically unparalleled diversity of traits (Yahara et al. 2013). The economic and agricultural importance of this group has led to a considerable body of literature regarding its reproductive biology. Arroyo (1981) collated and explored the distribution of breeding systems in Fabaceae with an impressive collection of data containing 364 species, highlighting a general trend of a high frequency of SI in woody tropical groups and a low frequency in herbaceous temperate groups. Neither the molecular causes nor confirmation of the distribution of RSI across core eudicots would be uncovered for nearly a decade following her review (McClure et al. 1989), so her paper could not assess information regarding the mechanism responsible for the action of SI in Fabaceae. While characterization of SI systems is still fraught with challenges (e.g., Allen and Hiscock 2008), recent work takes advantage of the knowledge regarding the distribution of SI and candidate-based sequencing experiments, focusing on the metonymous pistil-expressed genes (T2/S-type RNases) in species that possess this system (Ramanauskas and Igić 2017, 2021; Liang et al. 2020). A nominal sliver of Fabaceae species (six species, five of six with SC, of ca. 19,500) have previously been unsuccessfully screened for candidate S-RNase genes (Sims et al. 1999; Aguiar et al. 2015). The limited available data on the distribution and phenomenon of SI in the family represent a barrier to a broader understanding of the evolutionary causes and consequences of breeding system variation, obscure the mechanisms that may operate, and limit efficient sampling for efforts to uncover the genetic basis of SI in the family.

Here, we aim to examine the distribution and frequency of SI in Fabaceae by recording the currently known literature. We evaluate potentially correlated evolution between breeding system and three traits: ploidy, longevity, and stem composition. Combining the empirical data gathered from pollination experiments, genetic studies, and molecular and physiological work on SI, we assess the overall weight of evidence regarding the claim that RSI may operate in Fabaceae. Finally, we provide a set of recommendations for efficiently uncovering the molecular genetic basis of SI in this and other families.

Material and Methods

Study System

Fabaceae is the third-largest family of flowering plants behind Asteraceae and Orchidaceae and is perhaps the most economically important behind Poaceae (Azani et al. 2017). Its species are distributed worldwide, and many play an indispensable ecological role as nitrogen fixers. Legumes from each major lineage appear in the fossil record during the middle to late Eocene, suggesting that its six major clades are approximately 50–55 Myr old (Lavin et al. 2005).

As currently described, the family contains six subfamilies (Azani et al. 2017). Papilionoideae, the largest and youngest subfamily, contains roughly 14,000 recognized species, or approximately 72% of the family. This subfamily is characterized by the highly specialized papilionoid flower and an increasing trend toward a herbaceous growth form, and it is also the most widespread of the family. Most agriculturally important Fabaceae species belong to this group (e.g., common beans, soy, lentils, peanuts). Caesalpinioideae, which includes Mimosoideae, is a lineage of mostly tropical trees that contains approximately 4400 species, or 22% of the family. The remaining 6% of species belong to the four smallest and oldest subfamilies: Duparquetioideae, Cercidoideae, Detarioideae, and Dialioideae.

Breeding System Data

Using the ISI Web of Knowledge Science Citation Index and Google Scholar, we collected published reports with dozens of search terms related to plant mating and breeding systems within Fabaceae. We also included data from conservation reports, local flora and field guides, books, conference presentations, theses, monographs, and citations recovered within each source. We examined more than 1000 published works. From 366, we assembled a primary database that contained breeding system and associated trait data.

Breeding system data were collected and scored with an accompanying quality score (QS) indicator (Grossenbacher et al. 2017). Instances where the relative success of selfed and outcrossed hand pollinations were reported or those that directly provided a measure along with transparent calculation procedures received a score of 1 (QS = 1; highest quality). These high-quality data were used to calculate an index of SI (ISI). Following Lloyd (1968), we define ISI as

$$\text{ISI} = 1 - \frac{\text{relative selfed success}}{\text{relative outcrossed success}},$$

where relative pollination success is scored as the ratio of fruits (or seeds) set to flowers pollinated. The subtraction from unity differs from Lloyd's formulation, which is an index of SC, and it has the intuitive effect of having the higher ISI denote greater success following cross-fertilization (Raduski et al. 2012). When possible, calculation of ISI with fruit set, compared with seed set, is superior because reduced seed set can occur for reasons other than the prezygotic action of SI, especially inbreeding depression and resource limitation. Therefore, we used pollination experiments that report relative fruit set ($n = 223$). When these data were not available, we resorted to calculating ISI with seed set ($n = 13$). If only manual self-pollinations were performed (without cross-pollinations) and they successfully yielded fruit, these data were assigned a QS of 1 (QS = 1). If hand pollinations were performed but fruit set or seed set was not reported within the study, we assigned QS = 2, and those that simply stated the breeding system with no pollination experiments reported were given the minimum, QS = 3. When reports stated the breeding system with an associated citation that did not address the breeding system of that species, these data were not scored (i.e., "NA"). Our search included papers published through October 2018.

It is possible for ISI to take on negative values in the rare instances ($n = 24$) when the relative success of self-pollinations is higher than that of cross-pollinations. Within any single population, there is no clear biological basis for consistently increased fruit set following self-pollination versus cross-pollination (Lloyd and Schoen 1992; Raduski et al. 2012). Instead, negative ISI values likely result from a variety of sources of error, including finite sample sizes, incompatibility of outcross pollen, variances in pollen loads and other aspects of manual pollination, and mistakes (Raduski et al. 2012). Consequently, we set all negative values of ISI to 0, which helps facilitate downstream analyses. For reports of a variety of summary statistics, we employed an arbitrary approximation: species with ISI > 0.8 were classified as having SI (Bawa 1974).

Chromosome Counts and Ploidy Data

When reported alongside breeding system data, information regarding the chromosome number and ploidy of investigated species was scored from the original sources. When published reports did not indicate the chromosome numbers or ploidy of the plants tested, such data were instead taken from the Chromosome Counts Database (Rice et al. 2015) semiautomatically using the CCDBcurator R package (Rivero et al. 2019). For an initial prune, these records were sorted against a list of all recognized Fabaceae genera from the Plant List (TPL; https://www.theplantlist.org/), yielding a total of 4926 entries. These entries were then formally synonymized using the procedure described below for a total of 4687 entries. When chromosome numbers differed for a single species, the value with the highest number of published reports was chosen (mode, $n = 966$). When there was an equal number of reports indicating differing chromosome numbers, these data were not used ($n = 193$). We constrained chromosome numbers, removing entries with extremely low haploid (<4) values, because the original sources sometimes contain errors because of multivalence or notes of irregular chromosome counts. We observed 17 species (0.3% of the total data set) with haploid chromosome numbers greater than 56. We cross-referenced these entries with published data to ensure that these reports were accurate. Our final data set of chromosome counts included 4202 unique entries.

Stebbins's fraction. Most taxa have chromosome number information but lack ploidy determinations. We inferred species' ploidy designations relative to their generic base with an estimate referred to as "Stebbins's fraction" (Stebbins 1938; Román-Palacios et al. 2019). Sporophytic (2n) counts that are greater than or equal to 3.5 times the lowest haploid (n) base count of the genus are categorized as polyploid (Wood et al. 2009). In addition, we removed genera with data for fewer than three species, as the generic base may not be reliably established. While the use of this method to estimate increases in ploidy level may fail to detect older duplication events, we are primarily concerned with uncovering a pattern of association between recent chromosome doubling and loss of SC, indicative of the action of RSI.

Other Trait Data

We collected information on other traits if other traits of interest were also concurrently reported in the study with breeding system data for each species. Stem composition, longevity, and stature were almost always included within the original published reports, but in the rare instances when they were not, this information was collected from another published report when available.

We also collected data on pollen-to-ovule ratios, outcrossing rates, pollination syndromes (including tripping mechanisms, floral morphology that prevents self-fertilization without a pollinator), geographic locations, elevations, and pollen limitations (calculated from observations of natural and manual pollination success). These data are not evaluated in the present article but are available in the associated database.

Phylogenetic Relationships

We extracted phylogenetic relationships among Fabaceae from a larger data set used to create the broadly inclusive seed plant phylogeny (Smith and Brown 2018). Briefly, a sparse multilocus alignment matrix of species across Fabales was constructed using 4548 taxa from the National Center for Biotechnology Information (retrieved from http://www-personal.umich.edu/~eebsmith/big_seed_plant_datasets/trees/ on March 2, 2018). Tree inference involved constraint (monophyly) of recognized families and a model of sequence evolution partitioned by gene region. We used a list of all recognized Fabaceae genera from TPL to prune nonlegume taxa from the tree. (Taxonomic designations for all species were standardized using the synonym-resolving procedure described below.)

For rate estimates and visualization, we relied on two resulting phylogenetic trees. The first is a species-level tree for trait analysis containing only those species for which our database includes breeding system information and relevant trait data—ploidy, longevity, and stem composition. The second also used the same Fabales phylogeny to construct a genus-level Fabaceae tree by randomly selecting and retaining one tip from each genus. We imposed a semiparametric penalized likelihood molecular clock (Sanderson 2002) as implemented with the chronopl function (R package APE; Paradis et al. 2004).

Trait Correlation Analyses

We employed single-variable phylogenetic logistic regression models (Ives and Garland 2010) to test correlations between breeding system (SI or SC) and three binary traits: ploidy (diploid or polyploid), stem composition (woody or herbaceous), and longevity (annual or perennial). Although the association between annual plants and SC has long been recognized, there is no clear evidence indicating the order of transitions (see Friedman 2020 for a review). Long-lived species (woody or herbaceous perennials) may be associated with higher inbreeding depression, resulting in selection against selfing, while annual life cycles present risks without assured reproductive success (Hamrick and Godt 1996). On the basis of such observations, Barrett et al. (1996) argue that changes in ecological conditions drive transitions in life history traits, which then have important consequences for breeding system evolution (e.g., the repeated evolution of autogamy within annual lineages of *Gilia* species). Thus, stem composition and longevity were used as explanatory variables, along with ploidy, which commonly disrupts GfSI in diploid taxa (Miller and Venable 2000; Zenil-Ferguson et al. 2019). Because there are no woody annual species, we could not test interactions between these terms. It is unlikely that longevity and stem composition are independent, and we consequently performed separate analyses with these traits jointly coded as three states: herbaceous annuals, herbaceous perennials, and woody perennials.

We implemented all phylogenetic regression fitting and tests using the function phyloglm in the R package phylolm (Ho et al. 2016). We examined different combinations of these traits to test for significant interactions but focused only on traits that were significant in single-trait models: stem composition and longevity. We also used the jointly coded stem composition and longevity to assess whether there may be interactions between these traits and ploidy. We assessed the relative performance of each model using the Akaike information criterion (AIC) values from the phyloglm results. We display the results of the best-performing model using the mosaic function from the R package vcdExtra, which also displays Pearson residuals (Friendly 2021).

Taxonomic Name Resolution

We resolved taxonomic names for all species binomials in breeding system ($n = 1330$), chromosome count ($n = 4957$), and phylogenetic ($n = 4368$) data sets using the Taxonomic Name Resolution Service version 5.0 (Boyle et al. 2021) with three sources selected: the USDA (National Plant Data Team 2020), TPL (https://www.theplantlist.org/), and Tropicos (http://www.tropicos.org). Our search settings allowed partial matches and a match accuracy that returned all names found within an edit distance of four (four missed letters, four transposed letters). For all data sets, species designations classified as "accepted" were kept. Species designations classified as "synonym," "illegitimate," "invalid," "no opinion," "misapplied," or "rejected name" were replaced with "accepted" names when available or were removed from the data sets. We also removed species no longer recognized as belonging in the family. (See the supplemental PDF, available online, for additional information.)

All analyses were conducted in R (R Core Team 2020). All data (including the full breeding system database and associated reference list) are available in the supplemental data, available online, and supplemental PDF.

Results

The final assembled data set contains data from 366 reports published from 1886 through 2018, including books, theses, monographs, conservation reports, flora and field guides, and journal papers. There are 1411 entries for 1167 unique currently described species placed in 184 genera. Our data set contains 285 entries of data with a QS of 1, 104 with a QS of 2, and 770 with a QS of 3. Of these, 1050 characterize unique currently described species as having SI or SC.

Breeding System in Fabaceae

We find that 20.6% of tested Fabaceae species have SI and that the highest proportion of species with SI appear in Cercidoideae and Detarioideae (table 1). However, only a nominal portion of these subfamilies have been sampled (less than 1%); Caesalpinioideae exhibits the highest proportion of species with SI with the highest coverage. The family-wide SI percentage is less than the 27% with SI that Arroyo (1981) originally reported with 364 species. We found the breeding system status for 6.0% of Papilionoideae, 3.5% of Caesalpinioideae (including 3.6% of Mimosoideae), 3.5% of Dialioideae, and less than 1% of Cercidoideae and Detarioideae, and we found no

Table 1

Summary of Self-Incompatibility (SI) Frequency and Breeding System Coverage by Subfamily

Subfamily	Total species	Species with data	SI (%)	Coverage (%)
Duparquetioideae	1	0	...	...
Cercidoideae	~335	1	100.0	.3
Detarioideae	~760	3	66.7	.4
Dialioideae	~85	3	33.3	3.5
Caesalpinioideae:	~4400	204	52.0	3.5
Mimosoideae	*~3000*	*110*	*62.7*	*3.7*
Papilionoideae	~14,000	839	12.8	6.0
Total	~19,500	1050	20.6	5.4

Note. Columns list the total number of unique species, the number of species for which data on breeding systems are available, the percentage of species with SI of those tested for breeding system in each subfamily, and coverage as the percentage of the subfamily with the tested breeding system. For comparison with previous studies, the caesalpinioid data originating from the previously recognized mimiosoid subfamily are shown in italics. Approximate total numbers for each subfamily are as given by the Legume Phylogeny Working Group (Azani et al. 2017).

data for Duparquetioideae, for a total family-wide coverage of 5.4%. We were unable to find any reported data for 410 currently recognized genera across the family (see the supplemental PDF for a full list). The vast majority of genera with reported breeding systems have such data for less than one-fifth of each genus (fig. 1). Of the 163 genera in our data set that are also found on the phylogenetic tree, 125 entirely have either SI or SC (fig. A1; figs. A1, A2 available online).

ISI and quality of breeding system reports. We found experimental crossing data sufficient to estimate ISI for 236 species (fig. 2). ISI fruit set data were preferred ($n = 223$), but we used ISI calculated from seed set where fruit yield was not reported ($n = 13$). These data span four subfamilies (Detarioideae, Dialioideae, Caesalpinioideae, and Papilionoideae) and 84 genera. We also added species for which only manual self-pollinations were performed but full fruit set was observed, assigning an ISI value of 0 ($n = 66$ for Papilionoideae and $n = 6$ for Caesalpinioideae). Thus, our total ISI data set contains 308 entries from 95 genera (12.6% of 756 recognized genera). As observed across angiosperms (Raduski et al. 2012), the frequency appears strongly bimodal, with 79.9% of species possessing values of ISI ≤ 0.2 or ISI ≥ 0.8. For species where ISI calculation was possible, more than one-third (37.0%) are classified as having SI on the basis of a cutoff value of 0.8.

We find a significantly higher proportion of species with SI in our ISI data (QS 1) than across our entire data set (37.0% vs. 20.6%; two-sample test for equality of proportions: $\chi^2 = 33.2$, $P < 0.01$). We also observe a higher proportion of species with SI for Papilionoideae in our QS 1 data than across the data set (24.8% vs. 12.8%; two-sample test for equality of proportions: $\chi^2 = 12.2$, $P < 0.01$). This proportional difference in SI data is not significant for Caesalpinioideae (52.0% vs. 54.3%; two-sample test for equality of proportions: $\chi^2 = 0.1$, $P = 0.76$; there is a single QS 1 record from Detarioideae [SC] and Dialioideae [SI]). This likely reflects a bias caused by the fact that SC is easier to ascertain. It is easier to casually establish self-fertilization, especially in a glasshouse environment, because it requires no specific intervention. On the other hand, it may also often be scored in error, especially when unreliably reported on account of copious fruit set in the field. We ignored such reports, as well as those that claimed that entire genera had SI/SC without any accompanying evidence. We cannot be sure, however, that secondary reports did so.

The surveyed literature contains a greater proportion of qualitative SC accounts than qualitative SI accounts (80.0% vs. 58.0%; two-sample test for equality of proportions: $\chi^2 = 49.8$, $P < 0.01$). Of the 900 SC reports, 723 were QS 2 or QS 3 data. Of the 238 SI reports, 138 were QS 2 or QS 3 data. The data also contain a higher proportion of qualitative accounts from Papilionoideae, which has a far lower frequency of species with SI generally (84.6% qualitative from Papilionoideae vs. 39.8% from other subfamilies; two-sample test for equality of proportions: $\chi^2 = 181.2$, $P < 0.01$).

Genetics of pollen rejection and site of pollen tube inhibition. We evaluated the evidence consistent with a variety of plausible sterility and SI mechanisms across the family. We found nine studies that investigated the patterns of genetic inheritance of SI in Fabaceae species (table A1; tables A1, A2 available online). The results of these studies are sometimes vague, uncertain, or difficult to interpret, and they may conflate phenomena that are distinct from SI (e.g., inbreeding depression). Only three genera of Papilionoideae have been examined, and we found no work establishing the genetics of the response for species from other subfamilies.

We also recorded outcomes of studies that tracked the site of pollen tube arrest following self-pollination. Published reports on pollen tube arrest were available for 38 species spanning 27 genera of 2 subfamilies: 16 from Papilionoideae, 10 from Caesalpinioideae (5 from Mimosoideae), and 1 from Detarioideae. Fabaceae species show an apparently high variation in the location of pollen tube arrest following self-pollination, which is reported at the stigma, within the style, or at the ovary (fig. 3). Half of the studied species are reported to express some kind of "late-acting" phenomenon, meaning that the pollen tube arrests at the base of the style, at or near the ovary (late-acting SI [LSI]), or that the cross fails after fertilization (postzygotic failure, distinct from SI). The remainder arrest earlier in the pistil, between the stigmatic surface and the bottom of the style (table A2).

Trait Associations

We evaluated trait correlations between breeding system, ploidy, longevity, and stem composition using phylogenetically corrected logistic regression (phyloglm function in R). After the Fabaceae species-level tree was pruned to retain only taxa for which ploidy, longevity, and stature were also known, our final tree contained 185 species, representing species from Papilionoideae ($n = 132$), Caesalpinioideae ($n = 52$), and Cercidoideae ($n = 1$). This tree and the companion data set were used for all subsequent analyses of trait associations. We report AIC values, significant terms, coefficients of independent variables, and phylogenetic correlation parameters (α) for each model (table 2). Our best-performing models indicate that longevity is significantly correlated with breeding system.

Stem composition and longevity. Pairwise phylogenetically corrected logistic regression analyses using the phyloglm function found evidence of correlations between breeding system

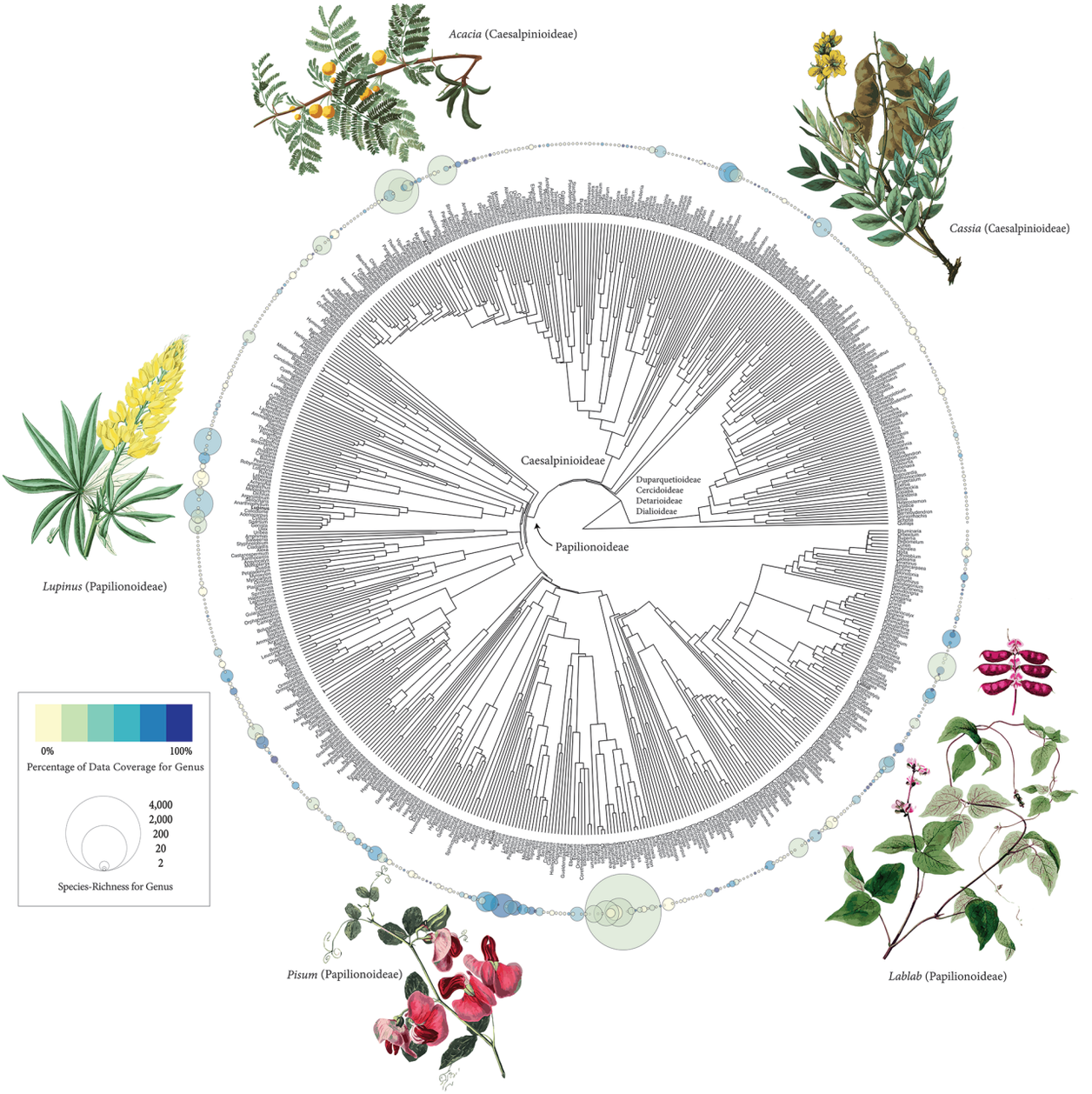

Fig. 1 Genus-level legume phylogeny displaying breeding system data coverage. Each circle is scaled to indicate the relative size of genera. Nodes of Caesalpinioideae and Papilionoideae are labeled, and illustrations indicate representative morphology of selected genera. The color gradient represents 0%–100% breeding system data coverage within that genus. For approximately even coverage, genera were manually binned into each of six percentage-of-coverage categories: 0 ($n = 412$), >0 and <1.7 ($n = 31$), ≥1.7 and <4.0 ($n = 33$), ≥4.0 and <9.0 ($n = 36$), ≥9.0 and <20.52 ($n = 27$), and ≥20.52 ($n = 36$). The largest category included 412 genera with 0% coverage, indicating that breeding system studies are wanting. (The phylogenetic tree is taken from Smith and Brown [2018]; see the text for details.)

and both stem composition ($z = -1.2$, $P < 0.01$) and longevity ($z = -3.3$, $P < 0.05$), but stem composition evaluated independently was one of the lowest-performing models examined. Moreover, when stem composition and longevity are evaluated together in a single phyloglm model (StemComp + Longevity), only longevity is significantly correlated with breeding system ($z = -3.2$; longevity: $P < 0.05$; stem composition: $P = 0.40$). As these terms likely interact, we jointly coded them into a single term, StemCompLong, which was significantly correlated with breeding system in all models evaluated ($P < 0.01$). The best-performing model includes both stem composition and longevity as separate terms as well as ploidy, with longevity being the only significant term ($z = 4.1$, $P < 0.01$). We observed far fewer woody perennials with SC and more perennials with SI,

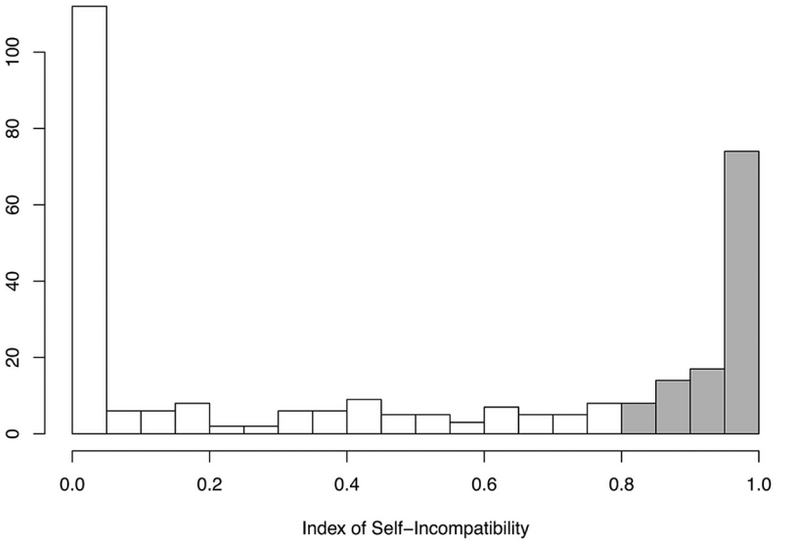

Fig. 2 Distribution of index of self-incompatibility (ISI) values in Fabaceae from the collected literature (308 species). ISI measures the relative selfed and outcrossed success of fruit set per flower after manual pollination. ISI = 0 denotes equal success following self- and cross-fertilization, and ISI = 1 denotes failed fruit set after self-fertilization and success after cross-fertilization. A critical ISI value of 0.80 is often used to classify species as self-compatible (lower) or self-incompatible (higher).

regardless of stem composition, than expected if these traits were independent (fig. 4).

Ploidy. Our final chromosome counts database contains 4144 observations across four subfamilies (Cercidoideae, Detarioideae, Caesalpinioideae, Papilionoideae) and 188 genera, with 3294 diploid species and 850 polyploid. The lowest 2n polyploid value is 16, from the genus *Michrocharis*, where 2n = 8 is the generic base. The highest 2n diploid value is 56, from the genus *Leucaena*, where 2n = 112 is the highest observed chromosome count in the genus. Of these 4144 observations, 626 overlap with our breeding system database. The highest number of observations comes from Papilionoideae ($n = 531$, ca. 3.8% of the subfamily), followed by Caesalpinioideae ($n = 94$, ca. 2.1% of the subfamily) and a single observation from Cercidoideae. Of the 541 diploid species, 442 are classified as having SC and 99 as having SI. Of the 85 polyploid species, 66 are classified as having SC and 19 as having SI.

In the data set we used for phylogenetically corrected logistic regression analysis, we retained 185 species, 33 of which are polyploid. Of the 33 polyploid species, 12 have SI: 1 from Caesalpinioideae and 11 from Papilionoideae. If we change the threshold for categorizing polyploids such that sporophytic counts (2n) that are greater than or equal to four times the lowest haploid (n) count of the genus are coded as polyploid, we retain seven SI polyploids: one from Caesalpinioideae and six from Papilionoideae. We find no evidence of significant correlation between ploidy and breeding system in any models evaluated, even with the greater threshold for assigning polyploids, but we do find evidence of a possible interaction between our jointly coded StemCompLong and ploidy ($z = 2.6$, $P < 0.05$; table 2).

Discussion

Establishing the pattern of genetic control of SI and gathering functional genetic evidence uncovering the molecular actors is time-consuming and expensive. This partly explains the absence of a clear explanation for the identity of genes governing SI as well as its distribution across legumes. Unlike in Brassicaceae or Solanaceae, for example, where the genetic basis of SI is widely shared by species within each family, it seems that the expression of various phenomena related to SI in Fabaceae is both varied and variable. Despite some lack of clarity, our results confirm many of those recovered by Arroyo (1981), and the greater availability of data, including phylogenetic relationships, allows several insights. SI is broadly distributed in Fabaceae, and its variability as a phenomenon, both within and between genera, suggests that no single SI "system"—as it is presently defined and understood—consistently operates or explains this variability. Furthermore, we find no compelling evidence, to date, that the most widespread RNase-based mechanism (RSI) is responsible for causing SI in this family. Such absence of evidence should clearly be interpreted with caution. We report 366 papers with breeding system information, an apparently large number that belies a surprising lack of data on SI frequency, phenomena, and underlying genetics across this very large family. Below, we discuss the principal findings and some shortcomings of our approaches and outline areas where future work may be effectively focused.

Features of SI in Fabaceae

Despite long-standing interest in the genetic basis, phylogenetic distribution, and frequency of SI across Fabaceae, the body of literature has not been systematically collected and summarized in decades. We modeled our search on the previous data set compiled by Arroyo (1981), which contained information on the breeding systems and their correlates for 364 species. With an additional 686 species, we find similar patterns of SI occurrence across Fabaceae: a higher frequency of SI across the primarily woody lineages (Duparquetioideae, Cercidoideae, Detarioideae,

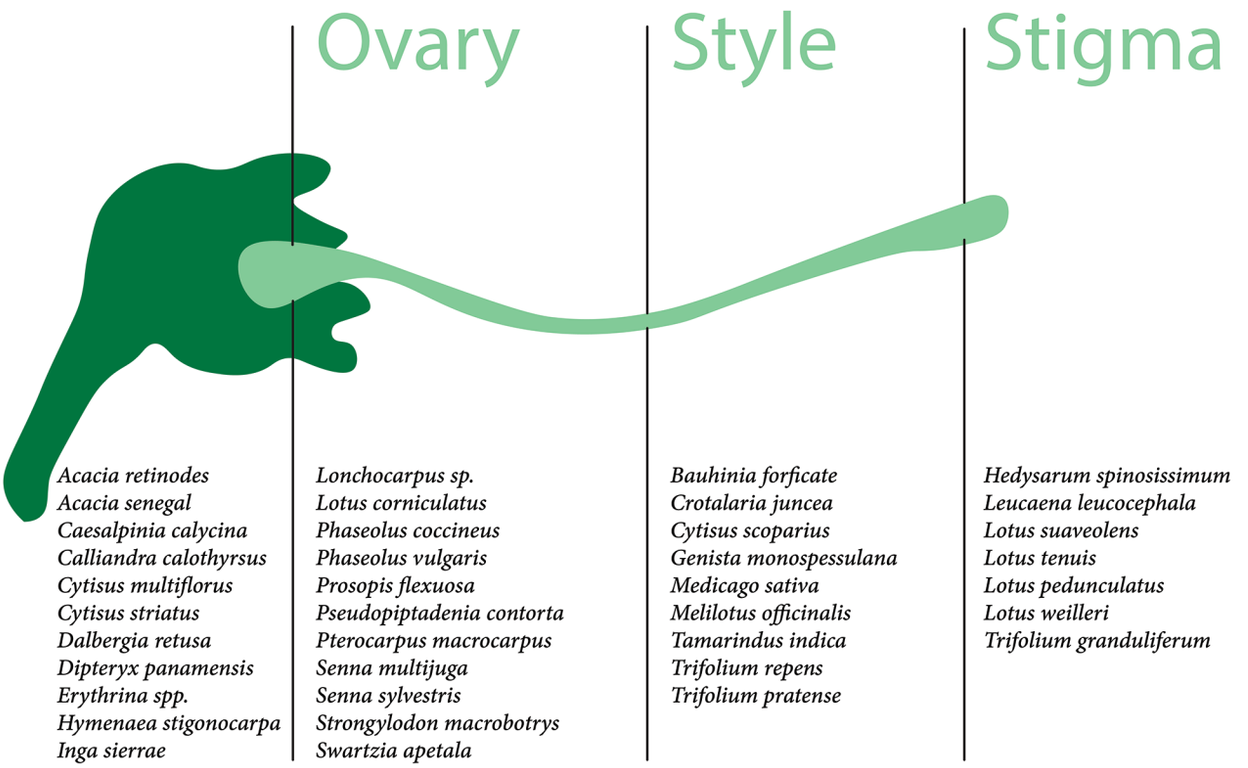

Fig. 3 Site of fertilization failure following self-pollination for the studied Fabaceae species. For stigma and style, the site of self-pollen tube arrest is indicated and likely involves self-incompatibility (SI). Species with arrest at the stigma additionally include those for which stigmatic membrane disruption can facilitate successful self-pollination (e.g., many *Lotus* species). For ovarian rejection, it is unclear without detailed study whether the failure mechanism is consistent with SI (late-acting SI) or whether the failure reflects early-acting inbreeding depression. We include reports of arrest from the nuclear layer of the ovule (*Acacia retinodes*) and any reported "incompatibility" reaction following fertilization (e.g., differential zygote and fruit abortion in selfed species). See table A1 (tables A1, A2 are available online) for a full list of references and study findings.

Dialioideae, Caesalpinioideae) and a lower frequency in the primarily herbaceous papilionoids (fig. A2).

Arroyo also pointed out that the lower proportion of species with SI in Papilionoideae possibly reflects their trend toward herbaceous stem composition. While SI is found in all subfamilies with breeding system data (table 1), its occurrence is generally concentrated in some of the less species-rich ("early diverging" or "basal") lineages—such as the tribes Bauhineae and Cercideae, as well as some mimosoids. It also appears elsewhere recently derived within narrowly circumscribed groups such as the genus *Trifolium*, potentially supporting the claim that some form of SI evolved independently in this family (Arroyo 1981). Without a clear understanding of the molecular genetic components and a broad sampling, however, it is presently impossible to confidently code traits, infer much detail regarding breeding system trait transitions, and/or estimate the confounded trait-dependent diversification rates.

Major differences seem to exist in the described phenomenon of SI, both within and between groups. Nevertheless, we recover a bimodal distribution of ISI values similar to the one described across many distantly related angiosperm families (e.g., Raduski et al. 2012; Grossenbacher et al. 2017). This widely observed bimodality of ISI may be effected by a combination of population genetic (Porcher and Lande 2005) and macroevolutionary processes. While SC does not ensure high selfing rates, breeding system (including ISI, its imperfect estimation metric) is strongly associated with mating system (Raduski et al. 2012). Therefore, some of the long-term equilibrium dynamics may be explained by the causal interaction of breeding and mating systems—breakdown of SI allows selfing rates to depart from zero. Species with high ISI values generally have high outcrossing rates, while those with low ISI values have lower ones, albeit with a wider range of outcrossing rates, as species with SC may employ other characters that promote outcrossing. The current theoretical approaches fail to accommodate higher-level processes, despite the possibility that transitions may be cladogenetic and highly asymmetric, and differential speciation and extinction rates may be ever present (Goldberg et al. 2010). The expected ISI distribution can then take on a number of shapes, including the observed bimodality, even in the absence of stabilizing lower-level selection (Igić and Busch 2013). It should be noted, however, that intraspecific distributions of ISI also tend to adhere to extreme values, casting considerable doubt on any proposition of persistent stabilizing selection on breeding systems (Raduski et al. 2012).

A challenging feature in determining the distribution and frequency of SI is the wide range in the quality of the reported data.

Table 2

Summary of Results Evaluating Trait Correlations

Model	Significant term(s)	AIC	Coefficient(s)	α
StemComp + Longevity + Ploidy	Longevity***	140	.43, 3.79, −.78	.81
StemComp + Longevity	Longevity***	142	.35, 3.53	.98
Longevity	Longevity***	142	3.65	1.03
StemComp	StemComp*	171	1.56	.65
Ploidy	None	183	.22	2.03
StemCompLong + StemCompLong:Ploidy	StemCompLong,*** StemCompLong:Ploidy*	140	2.07, −1.85 (2.10, ploidy)	1.48
StemCompLong + Ploidy + StemCompLong:Ploidy	StemCompLong,*** StemCompLong:Ploidy*	141	2.07, 2.10, −1.85	1.47
StemCompLong	StemCompLong***	145	2.05	1.62
StemCompLong + Ploidy	StemCompLong***	147	2.09, −.57	1.68

Note. We performed phylogenetic logistic regression in phyloglm to assess possible correlated evolution between breeding system and three traits: stem composition (StemComp), longevity (Longevity), and ploidy (Ploidy). We also evaluated a collapsed term combining stem composition with longevity (StemCompLong). We report the models that evaluated stem composition and longevity as separate terms, followed by the models that evaluated the jointly coded stem composition and longevity term (StemCompLong). Coefficients from each independent term are listed in the order that these terms appear in the model. The phylogenetic correlation parameter (α) is also reported. AIC = Akaike information criterion.

* $P < 0.05$.

*** $P < 0.001$.

Data reporting varies in part because of the varying definitions of SI employed and the methods employed to establish whether a plant has SI. We subscribe to the century-old reasoning that broadly limits SI to prefertilization phenomena and considers postzygotic differences in the results of self- and cross-pollinations distinct (perhaps most frequently inbreeding depression, selective abortion, etc.). This definition is consistent with those of East and Park (1917), Brewbaker (1958), Arasu (1968), and de Nettancourt (2001), who all suppose SI to be the inability of a cosexual plant producing fertile gametes to set seed when self-pollinated.

The use of a QS provides a means of quantifying confidence in the reported data. Manual pollination experiments that allow for the calculation of ISI are of the highest quality because these methods distinguish SI from the myriad of other reasons that a plant may not set seed from its own pollen (e.g., resource limitation, unreceptive stigma, immature pollen, or experimental errors). Moreover, when manual self-pollination is performed without manual outcrossing in tandem, this may provide information on the ability to self-fertilize but is uninformative with respect to SI. Yet restricting the survey to only species for which high-quality data exist excludes a high proportion of available data. Many informative sources include the incompatibility status of a species without experimental data because that aspect either was not appropriate in that venue (e.g., floras) or was not the purpose of the paper. It is likely that such sources also contain accurate information on SI.

Weak Evidence regarding Mechanism(s) of SI in Fabaceae

Several correlated characteristics plausibly implicate RSI in self-rejection within Fabaceae, particularly in *Trifolium* (Casey et al. 2010). These characteristics include the GSI rejection of pollen established through reciprocal crossing experiments (Brewbaker 1955), pollen tube arrest in the style following self-pollination (although some evidence implicates stigmatic location; Casey et al. 2010), and the presence of wet stigmas and binucleate pollen grains (Brewbaker 1967). The wide occurrence of RSI across eudicots—especially the close rosid relatives Rosaceae—along with its ancestral condition in core eudicots (Igić and Kohn 2001) adds up to a bantamweight of evidence supporting the possible occurrence of RSI in Fabaceae.

Yet a closer inspection yields a complicated picture. While most species indeed possess wet styles and binucleate pollen, there is almost no clear shared pattern of the SI phenomenon beyond this broad characterization. We found reports of evaluation of the genetic pattern of SI response in eight papilionoid species, half of which are from *Trifolium*. While Casey et al. (2010) located a single S-locus in *Trifolium repens*, their study did not report evidence regarding the presence or absence of candidate T2/S-RNases known to cause the pistil-part response or SLF genes, which cause the pollen-part response. Aguiar et al. (2015) also screened five papilionoid species for candidate S-RNases and did not find likely actors in the SI response, although the majority of their sample was composed of species with unambiguous SC. Finally, the RSI mechanism is commonly subject to "competitive interactions," whereby chromosome doubling results in the breakdown of SI (Lewis 1947; Brewbaker 1954). *Trifolium repens* is a recent tetraploid, which lowers the likelihood that it employs the RSI mechanism (Casey et al. 2010).

Systematic study of the documented site of pollen tube inhibition is informative because it provides hints about the possible underlying causes of SI. If pollen rejection occurs on or within the stigmatic depth, it is generally associated with a dry stigmatic surface and is characterized by sporophytic SI (SSI) genetic mechanisms. Otherwise, if pollen rejection occurs on or within the style, it is generally associated with a wet stigmatic surface and is characterized by GSI genetic mechanisms (Brewbaker 1967), which include RSI. Wet stigmatic surfaces are characteristic of Fabaceae (Heslop-Harrison and Shivanna 1977). Yet the reported site of inhibition in SI across the family appears unusually varied, with the arrest of self-pollen tubes not limited to the style. Pollen tube arrest is noted both on the stigma and in the style, while the majority

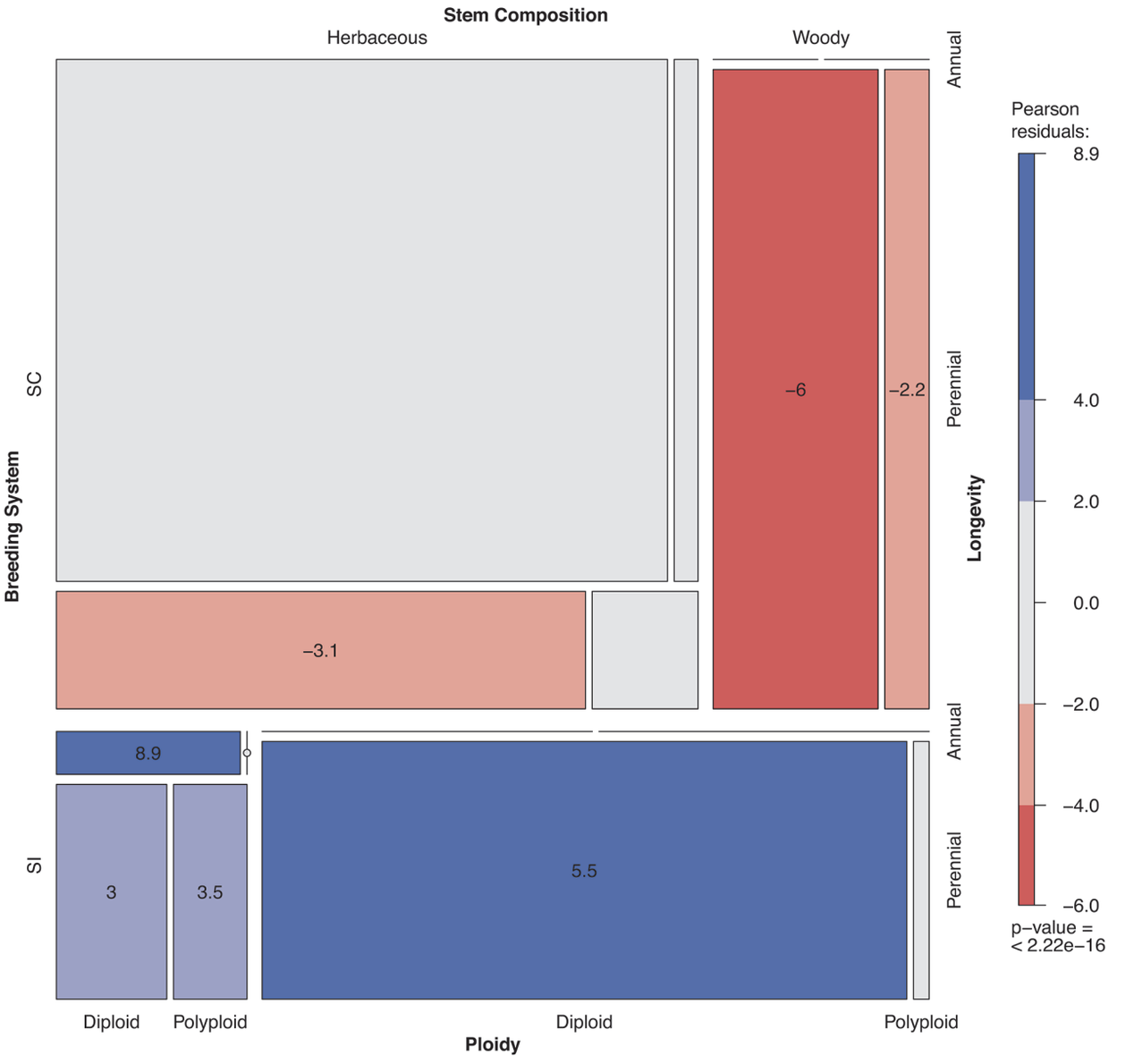

Fig. 4 Mosaic plot for the best-performing trait correlation model. We selected one of the best-performing trait correlation models (see table 2 for full model results): breeding system as a response variable of stem composition, longevity, and ploidy as separate terms (StemComp + Longevity + Ploidy). Individual boxes are sized relative to the number of observations for each group, and numbers in each box display Pearson residuals. Shading reflects the residuals (i.e., lack of fit) for this particular model: for example, we observed far fewer diploid woody perennials than would be expected if these traits were independent. Note that trait evolution of all evaluated models likely violates implicit model assumptions (e.g., equal diversification rates, homogeneity of rates), and these results are highly preliminary. SC = self-compatibility; SI = self-incompatibility.

of reports find that pollen is rejected at the ovule or after fertilization, which indicates phenomena unrelated to SI (fig. 3).

Angiosperm families were long believed to be generally characterized by a single SI system (Stebbins 1957), although there is little firm biological basis for such expectations and they were based on sparse empirical data. While the evolution of novel self-recognition mechanisms is relatively rare (Igić et al. 2008), the taxonomic rank of "family" implies no specific age or number of lineages, both of which affect the expected number of novel mechanisms. Both Primulaceae and Rubiaceae, for example, harbor both homomorphic GSI and heteromorphic SSI systems (Talavera et al. 2001), and Polemoniaceae may express both GSI and SSI (Goodwillie 1997). The available data do not support the action of a conserved "system" operating in Fabaceae or possibly within any one of its genera. It remains to be seen whether any species in Fabaceae express RSI, and our preliminary findings indicate that, should such species exist, they are most likely to be found in the woody species outside Papilionoideae.

Finally, species with so-called late-acting (LSI) mechanisms—that is, ovarian (ovule penetration) or zygotic (zygote development)

arrest—are dramatically understudied across angiosperms, as well as in Fabaceae (Gibbs 2014). We are not aware of a single functional genetic study uncovering the underlying causes. The task is made more difficult because many studies that encounter a failure of plants to set fruit or seed often casually ascribe such phenomena to LSI, which conflates SI mechanisms of self-rejection with inbreeding depression, resource limitation, and maternal effects. Therefore, the distinction between pre- and postzygotic arrest may appear merely semantic, but it can bring clarity to functional and mechanistic studies in establishing causality in both applied and basic domains of research (de Nettancourt 2001). The wide acceptance of a distinction in the evolutionary literature likely traces to a rather straightforward conjecture: a chronic wastage of fertilized ovules would be quickly stemmed by natural selection, given any variation in a slightly earlier-acting barrier to fertilization. The lack of such a distinction between pre- and postzygotic phenomena also makes the pursuit of the molecular underpinnings, their evolutionary origins, and mechanisms by which they are maintained (or lost) much more difficult. The disparate nature of the various causes of self-sterility and incompatibility, as well as their inconsistent reporting across the family, all but ensures that they do not describe a single underlying genetic system.

Trait Associations

Arroyo (1981) noted a correlation between woodiness and SI within Fabaceae. Previous studies have commonly found an association between perennials and SI (Hamrick and Godt 1996), as such species may predominantly outcross as a result of the substantial reduction in fitness among selfed progeny (Morgan et al. 1997). Selection for SC may also be strong in annual species as a means of reproductive assurance. We conducted analyses that found that woodiness and SI may be correlated only through perennial duration, inasmuch as woody species are obligate perennials. In our best-performing models, we did not find evidence of a correlation between stem composition (alone) and SI. Instead, our analyses indicated that longevity may be significantly correlated with SI, while stem composition is not.

GSI mechanisms are often causally associated with the breakdown of SC following polyploidization as a result of S-locus duplication and so-called competitive interaction (Lewis 1947). Several studies have examined the potential correlation between SC and polyploidy. Mable (2004), when considering only groups with GSI, found some evidence of a correlation between polyploidy and SC in a comparative study across angiosperms. Other studies have similarly reported a correlation between SC and polyploidy in families with GSI (Miller and Venable 2000; Robertson et al. 2011). As such, a lack of correlation between ploidy and breeding system in some legumes discounts the probability of the occurrence of RSI. Recent work also supports a long-ago whole-genome duplication in the common ancestor of the papilionoids (Cannon et al. 2015), additional evidence that the presence of RSI is unlikely in this subfamily. Although we found no significant correlation between SC and polyploidy, our analysis was restricted to a small portion of Fabaceae species (approximately 1%), the majority of which are papilionoids. It is as yet unclear whether there is a single homologous SI mechanism operating in the family or several independently evolved mechanisms, so any interpretations of comparative analyses of this kind are deeply uncertain.

A relatively low frequency of SI and the observed trait associations may have multifarious causes or a single major underlying cause. For example, they may each be caused by selection imposed by frequent colonization and the dispersal mode of speciation. A common parallel outcome of colonization is a filtering effect for uniparental reproduction (Baker 1955; Grossenbacher et al. 2017). Such a filtering effect may also be amplified in short-lived species (e.g., annuals) given their need for immediate reproduction and may likewise be alleviated in long-lived perennials, which would allow more than one immigrant of the same species to accumulate over an extended period of time. Interestingly, although there is considerable literature on the topic of establishment following long-distance dispersal, especially the role of breeding systems (termed "Baker's rule"; reviewed in Pannell et al. 2015) and rhizobial partner nodulation (e.g., Parker 2001), virtually none of it examines their interaction. While the data on the frequency of SI are not particularly reliable because of low reporting for most angiosperm families, Igić et al. (2008) surmised it to be ca. 40%–50% for flowering plants overall. The eye-catchingly low frequency of ca. 20% SI for Fabaceae is also well under the best estimates for Asteraceae (60%), Brassicaceae (30%–40%), and Solanaceae (40%; Grossenbacher et al. 2017).

Phylogenetic comparative analyses such as the ones we conducted have some fairly basic limitations. First, even though the best evidence strongly links the traits we examined with differential diversification and asymmetric transition rates (Goldberg et al. 2010; Mayrose et al. 2011), because our data are both sparse and unevenly collected, we did not attempt to fit state-dependent diversification models, sometimes referred to as *SSE family models (Maddison et al. 2007). This is very likely to affect key results. Trait-dependent diversification of the kind recorded elsewhere with breeding systems violates the basic assumptions of the widely deployed comparative phylogenetic models (including those we implement in phyloglm). This family of models assumes that traits essentially passively evolved into a tree of phylogenetic relationships (Maddison 2006). Violations of state-independent diversification, in turn, interact particularly poorly in cases of asymmetric trait changes (Goldberg and Igić 2008), such as the strongly biased transitions from SI to SC, the diploid to polyploid state, woody to herbaceous stem composition, and the perennial to annual life history. Second, the haphazard reporting of SI and SC, in particular, and the subsequent replication of reporting errors in databases can clearly affect all downstream comparative analyses, as well as any experimental designs that rely on fidelity in trait reporting. In many ways, the principal value of our article is the associated database, which attempts to curate the data to both improve accuracy and provide some idea regarding quality assessment for individual data points (using QSs). Nevertheless, we believe the exploratory analyses to be useful, in the sense that they provide information about the distribution of SI, raw correlations, and summary statistics, each of which can minimally highlight the value of reporting and accounting for breeding system, enabling future assessment of its effect on diversification.

Analyses such as ours are, however, only as good as the data used to conduct them. Recent computational advances enable the application of increasingly complex large-scale models, providing the potential to tackle fundamental questions related to the interplay between traits and the mechanisms that shape them and

how this interplay affects the disparity and diversity of organisms. As computational power increases, so must the quality and availability of trait data sets. In Fabaceae, data on SI are absent for more than 400 genera and are generally disproportionately concentrated in groups with agricultural significance. Our (immediately available and unrestricted) comprehensive and curated data set of breeding systems across Fabaceae can be used to address a variety of linked basic questions in ecology and evolution. The key upshot for work on the molecular genetic basis of breeding systems is that the older caesalpinioid species with documented pollen tube arrest in the style may be a good starting point for a candidate-based RNA-Seq approach in any future exploration of SI in legumes (as in Ramanauskas and Igić 2021). SI has extensive and profound consequences in the evolutionary history of angiosperms. Without the continued difficult work of recording and reporting the basic phenomenon of SI, as well as examining its molecular components, it remains an open question whether SI in Fabaceae represents a single system with many losses across groups or many separate independently evolved systems. Meaningful data on SI and its distribution across flowering plants enable the determination of the selective pressures influencing both the origin and the maintenance of SI systems across macroevolutionary timescales and help determine how these systems influence the diversity and ecological success of angiosperms. Such data enable agricultural and conservation efforts and provide improved power for tackling evolutionary problems concerning how reproductive processes shape organisms and their distributions.

Acknowledgments

We were inspired to pursue this work by a paper by M. T. Kalin-Arroyo published in 1981 without a global electronic data network. We are grateful to the staff at the Richard J. Daley Library, University of Illinois at Chicago, for their help with collecting many difficult-to-find books, papers, and theses. Andrew Raduski and Karolis Ramanauskas provided help and advice throughout the project. The article was vastly improved by critiques and suggestions from two anonymous reviewers and H. Sauquet. This work was supported by National Science Foundation grant NSF-DEB-1655692 to B. Igić.

Literature Cited

Aguiar B, J Vieira, AE Cunha, CP Vieira 2015 No evidence for Fabaceae gametophytic self-incompatibility being determined by Rosaceae, Solanaceae, and Plantaginaceae *S-RNase* lineage genes. BMC Plant Biol 15:129.

Allen A, S Hiscock 2008 Evolution and phylogeny of self-incompatibility systems in angiosperms. Pages 73–101 *in* VE Franklin-Tong, ed. Self-incompatibility in flowering plants. Springer, Berlin.

Arasu N 1968 Self-incompatibility in angiosperms: a review. Genetica 39:1–24.

Arroyo MK 1981 Breeding systems and pollination biology in Leguminosae. Pages 723–769 *in* R Polhill, P Raven, eds. Advances in legume systematics. Pt 2. Royal Botanic Gardens, Kew, UK.

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

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

Barrett SCH 1998 The evolution of mating strategies in flowering plants. Trends Plant Sci 3:335–341.

——— 2013 The evolution of plant reproductive systems: how often are transitions irreversible? Proc R Soc B 280:20130913.

Barrett SCH, LD Harder, AC Worley 1996 The comparative biology of pollination and mating in flowering plants. Philos Trans R Soc B 351:1271–1280.

Barringer BC 2007 Polyploidy and self-fertilization in flowering plants. Am J Bot 94:1527–1533.

Bawa KS 1974 Breeding systems of tree species of a lowland tropical community. Evolution 28:85–92.

Boyle BL, N Matasci, D Mozzherin, T Rees, GC Barbosa, R Kumar Sajja, BJ Enquist 2021 Taxonomic name resolution service, version 5.0. Botanical Information and Ecology Network. https://tnrs.biendata.org/.

Brewbaker JL 1954 Incompatibility in autotetraploid *Trifolium repens* L. I. Competition and self-compatibility. Genetics 39:307.

——— 1955 Studies of oppositional allelism in *Trifolium nigrescens*. Hereditas 41:367–375.

——— 1958 Self-compatibility in tetraploid strains of *Trifolium hybridum*. Hereditas 44:547–553.

——— 1967 The distribution and phylogenetic significance of binucleate and trinucleate pollen grains in the angiosperms. Am J Bot 54:1069–1083.

Cannon SB, MR McKain, A Harkess, MN Nelson, S Dash, MK Deyholos, Y Peng, et al 2015 Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Mol Biol Evol 32:193–210.

Casey NM, D Milbourne, S Barth, M Febrer, G Jenkins, MT Abberton, C Jones, D Thorogood 2010 The genetic location of the self-incompatibility locus in white clover (*Trifolium repens* L.). Theor Appl Genet 121:567–576.

Charlesworth D 1985 Distribution of dioecy and self-incompatibility in angiosperms. Pages 237–268 *in* PG Greenwood, PH Harvey, M Slatkin, JM Smith, eds. Evolution: essays in honour of John Maynard Smith. Cambridge University Press, Cambridge.

de Nettancourt D 2001 Incompatibility and incongruity in wild and cultivated plants. 2nd ed. Springer, Berlin.

East EM, JB Park 1917 Studies on self-sterility. I. The behavior of self-sterile plants. Genetics 2:505.

Franklin-Tong VE, F Franklin 2003 The different mechanisms of gametophytic self-incompatibility. Philos Trans R Soc B 358:1025–1032.

Friedman J 2020 The evolution of annual and perennial plant life histories: ecological correlates and genetic mechanisms. Annu Rev Ecol Evol Syst 51:461–481.

Friendly M 2021 vcdExtra. https://CRAN.R-project.org/package=vcdExtra.

Gibbs PE 2014 Late-acting self-incompatibility: the pariah breeding system in flowering plants. New Phytol 203:717–734.

Goldberg EE, B Igić 2008 On phylogenetic tests of irreversible evolution. Evolution 62:2727–2741.

Goldberg EE, JR Kohn, R Lande, KA Robertson, SA Smith, B Igić 2010 Species selection maintains self-incompatibility. Science 330:493–495.

Goodwillie C 1997 The genetic control of self-incompatibility in *Linanthus parviflorus* (Polemoniaceae). Heredity 79:424–432.

Grant V 1956 The influence of breeding habit on the outcome of natural hybridization in plants. Am Nat 90:319–322.

Grossenbacher DL, Y Brandvain, JR Auld, M Burd, P-O Cheptou, JK Conner, AG Grant, et al 2017 Self-compatibility is over-represented on islands. New Phytol 215:469–478.

Hamrick J, MW Godt 1996 Effects of life history traits on genetic diversity in plant species. Philos Trans R Soc B 351:1291–1298.
Heslop-Harrison Y, K Shivanna 1977 The receptive surface of the angiosperm stigma. Ann Bot 41:1233–1258.
Ho LST, C Ane, R Lachlan, K Tarpinian, R Feldman, Q Yu, W van der Bijl, J Maspons, R Vos, MLST Ho 2016 Package phylolm. http://cran.r-project.org/web/packages/phylolm/index.html.
Igić B, JW Busch 2013 Is self-fertilization an evolutionary dead end? New Phytol 198:386–397.
Igić B, JR Kohn 2001 Evolutionary relationships among self-incompatibility RNases. Proc Natl Acad Sci USA 98:13167–13171.
Igić B, R Lande, JR Kohn 2008 Loss of self-incompatibility and its evolutionary consequences. Int J Plant Sci 169:93–104.
Ives AR, T Garland Jr 2010 Phylogenetic logistic regression for binary dependent variables. Syst Biol 59:9–26.
Lande R, DW Schemske 1985 The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. Evolution 39:24–40.
Lavin M, PS Herendeen, MF Wojciechowski 2005 Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. Syst Biol 54:575–594.
Lewis D 1947 Competition and dominance of incompatibility alleles in diploid pollen. Heredity 1:85–108.
Liang M, Z Cao, A Zhu, Y Liu, M Tao, H Yang, Q Xu, et al 2020 Evolution of self-compatibility by a mutant S_m-*RNase* in citrus. Nat Plants 6:131–142.
Lloyd DG 1968 Pollen tube growth and seed set in self-incompatible and self-compatible *Leavenworthia* (Cruciferae) populations. New Phytol 67:179–195.
Lloyd DG, DJ Schoen 1992 Self-and cross-fertilization in plants. I. Functional dimensions. Int J Plant Sci 153:358–369.
Mable BK 2004 Polyploidy and self-compatibility: is there an association? New Phytol 162:803–811.
Maddison WP 2006 Confounding asymmetries in evolutionary diversification and character change. Evolution 60:1743–1746.
Maddison WP, PE Midford, SP Otto 2007 Estimating a binary character's effect on speciation and extinction. Syst Biol 56:701–710.
Mayrose I, SH Zhan, CJ Rothfels, K Magnuson-Ford, MS Barker, LH Rieseberg, SP Otto 2011 Recently formed polyploid plants diversify at lower rates. Science 333:1257.
McClure BA, V Haring, PR Ebert, MA Anderson, RJ Simpson, F Sakiyama, AE Clarke 1989 Style self-incompatibility gene products of *Nicotlana alata* are ribonucleases. Nature 342:955–957.
Miller J, DL Venable 2000 Polyploidy and the evolution of gender dimorphism in plants. Science 289:2335–2338.
Morgan MT, DJ Schoen, TM Bataillon 1997 The evolution of self-fertilization in perennials. Am Nat 150:618–638.
National Plant Data Team 2020 USDA, NRCS: the PLANTS database. http://plants.usda.gov.
Neal P, G Anderson 2005 Are "mating systems" "breeding systems" of inconsistent and confusing terminology in plant reproductive biology? or is it the other way around? Plant Syst Evol 250:173–185.
Pannell JR, JR Auld, Y Brandvain, M Burd, JW Busch, P-O Cheptou, JK Conner, et al 2015 The scope of Baker's law. New Phytol 208:656–667.
Paradis E, J Claude, K Strimmer 2004 APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.
Parker MA 2001 Mutualism as a constraint on invasion success for legumes and rhizobia. Divers Distrib 7:125–136.
Porcher E, R Lande 2005 Loss of gametophytic self-incompatibility with evolution of inbreeding depression. Evolution 59:46–60.
Raduski AR, EB Haney, B Igić 2012 The expression of self-incompatibility in angiosperms is bimodal. Evolution 66:1275–1283.
Ramanauskas K, B Igić 2017 The evolutionary history of plant T2/S-type ribonucleases. PeerJ 5:e3790.
——— 2021 RNase-based self-incompatibility in cacti. New Phytol 231:2039–2049.
Ramsey J, DW Schemske 1998 Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst 29:467–501.
R Core Team 2020 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.
Rice A, L Glick, S Abadi, M Einhorn, NM Kopelman, A Salman-Minkov, J Mayzel, O Chay, I Mayrose 2015 The chromosome counts database (CCDB): a community resource of plant chromosome numbers. New Phytol 206:19–26.
Rivero R, EB Sessa, R Zenil-Ferguson 2019 EyeChrom and CCDBcurator: visualizing chromosome count data from plants. Appl Plant Sci 7: e01207.
Robertson K, EE Goldberg, B Igić 2011 Comparative evidence for the correlated evolution of polyploidy and self-compatibility in Solanaceae. Evolution 65:139–155.
Román-Palacios C, YF Molina-Henao, MS Barker 2019 Polyploidy increases overall diversity despite higher turnover than diploids in the Brassicaceae. bioRxiv P.717306.
Sanderson MJ 2002 Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol 19:101–109.
Scofield DG, ST Schultz 2006 Mitosis, stature and evolution of plant mating systems: low-φ and high-φ plants. Proc R Soc B 273:275–282.
Silva N, D Goring 2001 Mechanisms of self-incompatibility in flowering plants. Cell Mol Life Sci 58:1988–2007.
Sims T, M Ordanic, R Fjellstrom, P Collins 1999 Molecular genetics of gametophytic self-incompatibility in *Petunia hybrida* and *Trifolium pratense*. Page 52 *in* Pollen-stigma interactions. International Conference Dedicated to the Memory of Jack Heslop-Harrison, University of Oxford. (Abstract.)
Smith SA, JW Brown 2018 Constructing a broadly inclusive seed plant phylogeny. Am J Bot 105:302–314.
Stebbins GL 1938 Cytological characteristics associated with the different growth habits in the dicotyledons. Am J Bot 25:189–198.
——— 1957 Self fertilization and population variability in the higher plants. Am Nat 91:337–354.
Steinbachs J, K Holsinger 2002 *S*-RNase–mediated gametophytic self-incompatibility is ancestral in eudicots. Mol Biol Evol 19:825–829.
Stone J 2002 Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. Q Rev Biol 77:17–32.
Stout AB, C Chandler 1942 Hereditary transmission of induced tetraploidy and compatibility in fertilization. Science 96:257–258.
Talavera S, PE Gibbs, MP Fernández-Piedra, MA Ortiz-Herrera 2001 Genetic control of self-incompatibility in *Anagallis monelli* (Primulaceae: Myrsinaceae). Heredity 87:589–597.
Wood TE, N Takebayashi, MS Barker, I Mayrose, PB Greenspoon, LH Rieseberg 2009 The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci USA 106:13875–13879.
Yahara T, F Javadi, Y Onoda, LP de Queiroz, DP Faith, DE Prado, M Akasaka, et al 2013 Global legume diversity assessment: concepts, key indicators, and strategies. Taxon 62:249–266.
Zavada MS, TN Taylor 1986 The role of self-incompatibility and sexual selection in the gymnosperm-angiosperm transition: a hypothesis. Am Nat 128:538–550.
Zenil-Ferguson R, JG Burleigh, WA Freyman, B Igić, I Mayrose, EE Goldberg 2019 Interaction among ploidy, breeding system and lineage diversification. New Phytol 224:1252–1265.