doi:10.1111/evo.13103

Inbreeding depression and drift load in small populations at demographic disequilibrium

Rachel B. Spigler,[1,2] Konstantinos Theodorou,[3] and Shu-Mei Chang[4]

[1]*Department of Biology, Temple University, 1900 N. 12th Street, Philadelphia, Pennsylvania 19122*

[2]*E-mail: rachel.spigler@temple.edu*

[3]*Biodiversity Conservation Laboratory, Department of Environment, University of the Aegean, University Hill 81100, Mytilene, Greece*

[4]*Department of Plant Biology, University of Georgia, 2502 Miller Plant Sciences, Athens, Georgia 30602–7271*

Received February 20, 2015
Accepted October 13, 2016

Inbreeding depression is a major driver of mating system evolution and has critical implications for population viability. Theoretical and empirical attention has been paid to predicting how inbreeding depression varies with population size. Lower inbreeding depression is predicted in small populations at equilibrium, primarily due to higher inbreeding rates facilitating purging and/or fixation of deleterious alleles (drift load), but predictions at demographic and genetic disequilibrium are less clear. In this study, we experimentally evaluate how lifetime inbreeding depression and drift load, estimated by heterosis, vary with census (N_c) and effective (estimated as genetic diversity, H_e) population size across six populations of the biennial *Sabatia angularis* as well as present novel models of inbreeding depression and heterosis under varying demographic scenarios at disequilibrium (fragmentation, bottlenecks, disturbances). Our experimental study reveals high average inbreeding depression and heterosis across populations. Across our small sample, heterosis declined with H_e, as predicted, whereas inbreeding depression did not vary with H_e and actually decreased with N_c. Our theoretical results demonstrate that inbreeding depression and heterosis levels can vary widely across populations at disequilibrium despite similar H_e and highlight that joint demographic and genetic dynamics are key to predicting patterns of genetic load in nonequilibrium systems.

KEY WORDS: **Bottlenecks, genetic load, heterosis, mating system evolution, mutation accumulation, population size.**

Inbreeding depression is the reduction in fitness of inbred relative to outcrossed individuals, caused primarily by the expression of (partially) recessive, deleterious mutations in homozygous inbred progeny (Charlesworth and Charlesworth 1987; Carr and Dudash 2003; Charlesworth and Willis 2009). Inbreeding depression can balance the automatic gene transmission advantage of selfing in favor of outcrossing (Fisher 1941) and, as such, is *the* key determining parameter in practically all models of mating system evolution (reviewed in Goodwillie et al. 2005) and likewise influential in the evolution of combined versus separate sexes (Charlesworth and Charlesworth 1978; Spigler and Ashman 2012). Furthermore, inbreeding depression is widely documented in both plants and animals (Crnokrak and Roff 1999; Keller and Waller 2002; Winn et al. 2011) and can threaten the viability of small populations (Frankham 2005; O'Grady et al. 2006). Precisely due to inbreeding depression's evolutionary and conservation significance, the problem of identifying the factors that influence inbreeding depression and predicting its magnitude in any given population has remained relevant for over a century (Darwin 1876).

One such factor, population size, has received both theoretical and experimental attention. Most of the current theory predicts that small populations should exhibit lower inbreeding depression (Wang et al. 1999; Bataillon and Kirkpatrick 2000; Kirkpatrick and Jarne 2000). This is fundamentally due to (1) their higher inbreeding rate, which eventually leads to purging of highly recessive, segregating lethal and semilethal deleterious alleles and/or

EVOLUTION INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

(2) genetic drift, causing fixation of slightly deleterious alleles or loss of variation at overdominant loci ("drift load," sensu Keller and Waller 2002; Kimura et al. 1963; Crow and Kimura 1970; Lynch et al. 1995; Whitlock 2000; Glémin 2003). In case of drift load, genetic diversity losses in small populations effectively reduce variation in fitness between outcrossed and inbred progeny and thus the *relative* difference in their performance, that is, inbreeding depression.

Drift may play a more pivotal role in shaping inbreeding depression when considering small population sizes because the efficiency of purging deleterious recessive alleles can be limited when inbreeding is due to random mating in small populations alone versus recurrent nonrandom mating and is always limited once populations fall below a critical size threshold (Glémin 2003). Comparisons among progeny resulting from crosses between and within populations are useful for identifying this role (Keller and Waller 2002). If deleterious alleles have fixed within a population due to drift, lowering fitness of all individuals, crosses made between populations will result in the masking of these deleterious alleles in heterozygous offspring and thus increased fitness compared to offspring produced by local inbreeding (i.e., within-population crosses), a phenomenon referred to as heterosis (Whitlock et al. 2000; Charlesworth and Willis 2009). Evaluation of both inbreeding depression and heterosis is critical for understanding the nature of the genetic load in small populations. For example, low inbreeding depression accompanied by weak to no heterosis implies that purging has improved mean fitness of inbred individuals within a population, whereas low inbreeding depression accompanied by strong heterosis is consistent with drift decreasing the fitness of all individuals.

A notable problem in investigations of the relationship between inbreeding depression and population size is that census population size often diverges from effective population size (Allendorf et al. 2013; Palstra and Ruzzante 2008) and so may fail to predict the level of inbreeding depression. This could explain the equivocal results from studies in plants focused on census size (Hauser and Loeschcke 1994; Ouborg and Van Treuren 1994; Paland and Schmid 2003; Michaels et al. 2008; Oakley and Winn 2012). Instead, estimates of genetic diversity should provide a better proxy for effective population size. Indeed, a recent study showed a clear pattern of increasing inbreeding depression and decreasing heterosis with increasing genetic diversity in *Daphnia*, supporting theoretical predictions about the effect of population size on these important genetic parameters (Lohr and Haag 2015).

However, these predictions are based on the assumption that populations are at mutation–selection balance or, more generally, have stable dynamics and equivalent demographic histories. When considering demographic and genetic disequilibrium due to bottlenecks, fragmentation, and fluctuations due to anthropogenic disturbance or environmental stochasticity, predictions may not be a simple extension of those under equilibrium, even when considering genetic diversity. That is, while genetic diversity can reveal some differences in past history, for example, populations that have experienced a bottleneck should show lower genetic diversity than stable populations of the same census population size, it alone may not accurately predict inbreeding depression (Wang et al. 1999; Theodorou and Couvet 2006). There is a paucity of theory addressing the consequences of variation in population size for inbreeding depression at demographic disequilibrium (Kirkpatrick and Jarne 2000; Theodorou and Couvet 2006), but such considerations are becoming increasingly warranted in a rapidly changing world.

Here, we empirically and theoretically investigate the influence of census and effective population size on inbreeding depression and heterosis. First, we present results of experimental crosses from six populations of the self-compatible biennial *Sabatia angularis* (Gentianaceae) evaluating inbreeding depression and heterosis across the life cycle in a common field environment. We asked how the magnitudes of cumulative lifetime inbreeding depression and heterosis vary with census population size and effective population size, estimated as expected heterozygosity (H_e). Second, we develop novel models evaluating the evolution of inbreeding depression and heterosis under contrasting demographic scenarios at disequilibrium. Finally, we compare our empirical results with theoretical predictions.

Methods

STUDY SYSTEM

Native to eastern United States and southeastern Canada, *S. angularis* is found in a variety of habitats including prairies, glades, marshes, old fields, and roadsides. Seeds germinate in spring, and seedlings develop into rosettes that overwinter until the following spring when they begin to bolt. From July to August, plants produce displays of showy, pink flowers that are visited by a suite of generalist pollinators and develop into many seeded, dry dehiscent capsules.

Opportunities for expression of the inbreeding load in natural *S. angularis* populations—particularly small ones—may be frequent. The selfing rate is on average 22% (±0.12% SD, $N =$ 8 populations) but varies across populations from only 5 to 40%, with higher selfing in small populations (measured by census size; Spigler et al. 2010). Selfing likely occurs through a mixture of autonomous selfing within flowers (R. B. Spigler, unpubl. data) and pollinator-mediated selfing among flowers (geitonogamy). Previous estimates of inbreeding depression from a single, large *S. angularis* population revealed that it can be high ($\delta = 1 - w_{\text{selfed}}/w_{\text{outcrossed}} > 0.5$, Dudash 1990).

We conducted this study in North and South Carolina, United States, where many *S. angularis* populations are small and occur

in power line right of ways, patches of private property, or roadsides. We included six populations separated by ~0.40 to 88 km (Fig. S1). Genetic distance among populations in the region is fairly high ($F_{st} = 0.27$ based on AMOVA using seven polymorphic allozyme loci; Spigler et al. 2010). Study populations (BC1, BC2, JRM, LI1, LI2, and WF in Spigler and Chang 2008 and Spigler et al. 2010) were chosen to span a range of census sizes representative for the area based on initial censuses in 2004 ($N =$ 51–220 adults). Census population size fluctuated to some extent between years, and in 2005 the range was smaller ($N = 38$–91). Therefore, we use the harmonic mean of census sizes ($\overline{N_c}$) from the two sampling periods to best represent census population size (N_c; Gillespie 1998). This mean is strongly correlated with census size for both years and significantly so for 2004 ($r_{2004} = 0.95$, $P = 0.003$; $r_{2005} = 0.71$, $P = 0.11$). Although small, the range of census population sizes evaluated here is typical for the species in the area under study (Spigler and Chang 2008). Populations can reach thousands of plants, but only in actively protected and managed prairies and glades, which were not located in the area. As a surrogate for effective population size (N_e), here we use previous estimates of H_e for the study populations (0.14–0.34; Table 2 in Spigler et al. 2010). Briefly, average H_e was calculated across seven to eight polymorphic allozyme loci based on genotypes of eight to 15 seedlings from each of 20 maternal plants per population. However, given their patchy distribution in disturbed habitats across a fragmented landscape and prior work demonstrating independence among N_c, expected heterozygosity (H_e), and population inbreeding coefficients (F; Spigler et al. 2010), study populations were unlikely to be at demographic equilibrium.

EXPERIMENTAL POLLINATIONS

In 2005, we performed experimental crosses on 25 plants per population. We covered plants with pollinator-exclusion bags prior to flowering and emasculated the protandrous flowers upon opening, prior to stigmatic receptivity. Once receptive, we applied either self-pollen, outcross pollen from within the population (within-population), or outcross pollen from a different population (between-population) to stigmas. Donor pollen was always collected the same day it was used. We note that individual flowers for within and between crosses received pollen from only one donor of the correct type, comparable to self-pollinations, but donor genotypes were different across flowers, assigned haphazardly. In total, pollen for between-population crosses for each population came from at least four other populations. In a minority of cases this included pollen from four additional populations within the region when fresh pollen from the focal populations was limited, as well as to ensure a diversity of donors given the geographic distribution of focal populations (see Fig. S1, Table S1). All plants received all three treatments, and all flowers per plant were pollinated ($N = 3$–54 pollinations/plant, $N_{Total} =$ 1215). Unwelcome mowing and a deadly pathogen in the natural populations ultimately reduced our sample size ($N = 11$–24 maternal individuals/population). A final total of 101 maternal individuals and 955 experimental pollinations across the six populations were included in the study.

We performed experimental crosses in the field rather than the greenhouse to evaluate initial fruit and seed set under realistic field conditions, as inbreeding depression and heterosis can be masked in benign greenhouse conditions (reviewed in Armbruster and Reed 2005). One concern with this approach is that if census population size is correlated with habitat quality, maternal effects could carry over and confound any relationships with census size. However, neither mean plant height ($r =$ –0.36, $P = 0.49$) nor mean flower number per plant ($r = 0.05$, $P = 0.93$) was significantly related to $\overline{N_c}$, indicating that census population size and maternal quality are independent in our study.

ESTIMATING FITNESS COMPONENTS AND CUMULATIVE LIFETIME FITNESS

On all plants, we scored fruit set for each pollinated flower and determined the mass of up to 10 fruits per cross type, a surrogate for seed set (Spigler and Chang 2008). We evaluated germination in the University of Georgia Plant Biology Greenhouse; extremely small seed size (<0.4 mm diameter) makes germination in the field intractable. We planted a random subset of seeds for each cross type per plant (family; 25–50 seeds/cross type/plant; $N_{total} = 8945$) into trays containing an enriched 60:40 pine bark/vermiculite soil mixture in a randomized design, cold stratified them, and examined germination weekly. Seedlings remained in the greenhouse until transplanted into a common garden site; we recorded survival to transplanting (early survival).

In Fall 2006, we transplanted up to five rosettes/cross type/family into a common garden plot at the University of Georgia Plant Sciences Farm using a randomized complete block design. A common garden setting provides more realistic measures of inbreeding depression relative to a greenhouse, particularly in this species (Dudash 1990), while at the same time allowing relative comparisons of inbreeding depression and heterosis across populations. We scored survival to flower and flower number/plant and left flowers open to pollinators. We estimated fruit set as fruit relative to bud number and estimated seed set from fruit mass as above (~6 fruits/plant).

For each family, we calculated average cumulative lifetime fitness of a given cross type as the product of the probability of germinating, probability of early survival, probability of survival to flower, mean flower number, mean fruit set, and mean seed set converted to seed number (Spigler and Chang 2008). This formulation of lifetime fitness can be interpreted intuitively as the expected number of offspring produced by a seed resulting from

a given cross type. A single mean lifetime fitness estimate was calculated for each cross type per family. A few individuals ($N = 16$) survived to make fruits, but seed number data were missing (e.g., because fruits dehisced prior to collection). This can lead to bias in mean lifetime fitness estimates because nonsurviving individuals never have such missing data; their zero values would always be included in the family-cross mean. Therefore, to avoid bias against surviving individuals, we assigned those individuals their population's mean seed set for the appropriate cross type when calculating lifetime fitness. This approach did not affect our conclusions; significance of results did not change using lifetime fitness through fruit set (data not shown).

STATISTICAL ANALYSES

All data analysis was performed using SAS software version 9.3 (SAS Institute, Inc., Cary, NC). We evaluated which components of lifetime fitness exhibit inbreeding depression and heterosis using hierarchical mixed models. For each model, cross type (self, within- and between-population), population, and their interaction were included as fixed effects, and family was included as a random effect. We also included a random family-by-cross interaction term where replication allowed and block effects where relevant. We treated fruit set (following initial hand-pollinations), early survival, and survival to flower as binary variables; germination and open-pollinated fruit set were treated as following a binomial distribution (GLIMMIX procedure). Data for seed set from initial pollinations and total flower number were log-transformed to fit model assumptions (MIXED procedure). We used preplanned contrasts to evaluate the presence of inbreeding depression (self vs. within population) and heterosis (within vs. between population) for each trait. No selfed offspring from population BC2 survived to flower, creating the classic "empty cell" problem for analysis of population-by-cross interactions for traits beyond survival to flower (flower number, fruit set, seed set). For those postsurvival traits, we separated the data into two analyses (Quinn and Keough 2002): one for inbreeding depression without population BC2 and another for heterosis containing all populations. For combined analyses, we rejected the null hypothesis at $P \leq 0.05$; for preplanned contrasts and where data were split into two analyses using within-population data, we used $P \leq 0.025$.

Distributions of lifetime fitness notoriously violate assumptions of normality (e.g., Shaw and Geyer 2010), and our data were no exception. To test for the presence of inbreeding depression and heterosis in lifetime fitness, we used a zero-inflated negative binomial (ZINB) model (Zurr 2009), which includes two submodels representing different processes that can generate the pattern seen in our data: zero inflation and overdispersion of counts. We included cross type, population, and their interaction as predictor variables in both submodels (negative binomial regression and logit [zero inflated] submodel; PROC GENMOD). We ran separate analyses to evaluate inbreeding depression (self vs. within population) and heterosis (within vs. between population) because original models including all data and all possible interactions failed to converge. Random maternal effects were also left out to reduce model complexity (Bolker et al. 2009). We note that results of paired t tests evaluating overall presence of lifetime inbreeding depression and heterosis based on population mean lifetime fitness values per cross type are consistent with those from the ZINB analysis (results not shown). Significance was evaluated as above.

The overall *magnitude* of population-level inbreeding depression was estimated as $\delta = 1 - w_{\text{self}}/w_{\text{within}}$, where w_{self} and w_{within} are mean observed fitness of selfed and within-population crosses per population, respectively. However, because estimated inbreeding coefficients were above zero for two of the populations (BC1 and JRM; Spigler et al. 2010), selfed progeny of plants in these populations would have an inbreeding coefficient (F_{self}) higher than 0.5, the expected value for selfed progeny of noninbred parents. Therefore, we adjusted our estimates of δ for these two populations as follows. The inbreeding coefficient for these selfed progeny will be $F_{\text{self}} = F + 1/2(1 - F)$, where F is the population inbreeding coefficient estimated for the maternal generation. The inbreeding coefficient for outcrossed progeny will be zero ($F_{\text{outcrossed}} = 0$; Falconer 1960). If we assume that the relationship between fitness and F is linear, we can calculate for each population the expected fitness for the standardized selfed progeny ($F = 0.5$) as $w_{\text{self},F=0.5} = w_{\text{within}} + 0.5 \times (w_{\text{self}} - w_{\text{within}})/(F_{\text{self}})$. By imposing this adjustment to the populations with plants that were already inbred to a certain degree, our comparison of inbreeding depression values is not biased by differences in inbreeding level among our study populations prior to this study. However, we note that inbreeding coefficients in those populations were extremely low ($F = 0.03$–0.05). For the four populations with inbreeding coefficients equal to or lower than zero, the corresponding parameters are $F_{\text{self}} = 0.5$, $F_{\text{outcrossed}} = 0$ and $w_{\text{self},F=0.5} = w_{\text{self}}$. Consequently, inbreeding depression estimates based on $w_{\text{self},F=0.5}$ were nearly identical with inbreeding depression based on w_{self} ($r = 1$, $P < 0.001$).

Heterosis generally refers to hybrid vigor and can be defined in a number of ways (e.g., Lynch and Walsh 1998; Whitlock et al. 2000; Theodorou and Couvet 2002; Glémin et al. 2003; Roze and Rousset 2004). Here, we define heterosis as the reduction in fitness of progeny produced by within-population crosses relative to between-population crosses: $1 - w_{\text{within}}/w_{\text{between}}$, where w_{within} and w_{between} are mean observed fitness of within-population and between-population crosses, respectively (e.g., Paland and Schmid 2003; Coutellec and Caquet 2011; Oakley and Winn 2012). This formulation does not provide insight into the underlying genetic mechanism but rather emphasizes the relative change in fitness when comparing local inbreeding to outbreeding

among populations (e.g., Busch 2006) and is comparable to our calculation of inbreeding depression as both are bounded between 0 and 1 when inbred fitness is lower than outbred (see Fig. S2 for alternate calculation of heterosis).

We investigated whether inbreeding depression and heterosis vary with $\overline{N_c}$ and H_e using linear regression (PROC REG). Significant relationships were further evaluated using permutation tests (lmPerm in R; Wheeler 2010, R Core Team. 2015), which revealed similar *P* values (data not shown). To explore the extent to which inbreeding depression and heterosis at different stages in the life cycle may be driving these patterns, we also examined the signs of correlations between stage-specific values and $\overline{N_c}$ and H_e. We rejected the null hypothesis at $P \leq 0.05$.

MODELING INBREEDING DEPRESSION AND HETEROSIS CONSIDERING DISEQUILIBRIUM

We modeled the evolution of inbreeding depression and heterosis in populations at demographic disequilibrium following different demographic trajectories using individual-based stochastic simulations. We envisaged three plausible demographic scenarios: (1) a fragmentation scenario, wherein a previously large and continuous population is subdivided to smaller populations of stable census population size; (2) a bottleneck scenario, wherein populations recover slowly from a single bottleneck event experienced at some point in their past; and (3) a disturbance scenario, wherein populations are subject to periodic disturbances that repeatedly and drastically reduce its census size. We further allowed for variable selfing rates predicted by population size based on observations of Spigler et al. (2010) and consistent with findings in other systems (e.g., Robledo-Arnuncio et al. 2004; Delmas et al. 2014).

Genetic model

Our model follows closely common assumptions of several theoretical models of mutation accumulation (e.g., Wang et al. 1999; García-Dorado 2003; Jaquiery et al. 2009; Theodorou and Couvet 2015). We considered a two allele per locus model, *A* being the wild-type allele and *a*, a deleterious and partially recessive allele. The relative fitness of the *AA*, *Aa*, and *aa* genotypes at locus *i* are 1, $1 - h_i s_i$, and $1 - s_i$, respectively, where s_i is the selection coefficient and h_i the dominance coefficient of the deleterious allele at each locus. The genetic fitness of an individual is assumed to be multiplicative across loci and is determined by:

$$w = \prod_{nhet} (1 - h_i s_i) \prod_{nhom} (1 - s_i), \quad (1)$$

where the product is calculated over the number of loci that are in the heterozyote (*nhet*) and homozygote (*nhom*) state for the deleterious alleles. The number of loci under selection is $L = 5000$. Free recombination between loci is considered.

Two kinds of mutations were considered together: mildly or moderately deleterious and lethals. Mildly deleterious mutations have a mean selection coefficient $s_d = 0.05$ and mean dominance coefficient $h_d = 0.36$, with mutations occurring at a rate $U_d = 1$ per diploid genome and generation (Lynch et al. 1999; Haag-Liautard et al. 2007). The selection coefficients were sampled from a gamma distribution with shape parameter $\beta = 1$. This corresponds to an exponential distribution of gene effects (see, e.g., Schultz and Lynch 1997). The dominance coefficient was obtained from an exponential function of selection coefficients. The model is that proposed by Caballero and Keightley (1994), for which the dominance coefficient of a mutant is taken from a uniform distribution between 0 and exp(–*ks*), where *k* is a constant allowing the mean dominance coefficient to be the desired one (0.36 in this case). Lethal mutations ($s_l = 1$) are highly recessive ($h_l = 0.02$). According to data from *Drosophila melanogaster*, lethals arise each generation with rate $U_l = 0.02$ per diploid genome (Crow and Simmons 1983). However, it is been suggested that the mutation rate of lethal mutations can be much higher in plants (Lande et al. 1994; Winn et al. 2011). We therefore ran simulations with $U_l = 0.2$ (see also Porcher and Lande 2005). Although a higher U_l results in higher ID (Kirkpatrick and Jarne 2000), the two mutation rates give an almost identical *relative* decrease in ID with time (results not shown). In each generation, the number of new mutations is Poisson distributed with mean U_d for mild and U_l for lethal mutations. The loci where mutations occur were chosen at random.

The initial population was formed by sampling randomly from an infinitely large (ancestral) population at mutation–selection balance and linkage equilibrium. At each locus, a wild-type and a mutant allele were assumed, and the equilibrium mutant allele frequency in the ancestral population was obtained from the equation:

$$s_i\,(1-2h_i)\,(1-F_{IS})\,q_{eq}^2 + s_i\,[h_i + (1-h_i)\,F_{IS}]\,q_{eq} - u = 0, \quad (2)$$

with

$$F_{IS} = \frac{(1 - s_i h_i)\,Self}{2 - Self\,(1 - 2s_i + s_i h_i)}\,. \quad (3)$$

Equations are adapted from Glémin (2003; see eqs. 1, A4, and A6 therein). For each locus, s_i and h_i take values from the distributions for the selection and dominance coefficient described previously, and u is the mutation rate per locus per generation, $u = U/2L$. *Self* stands for selfing rate and is set to $Self = 0.05$ as observed in large populations of *S. angularis* (Spigler et al. 2010). To decide the genotype of each individual at a given locus, we compared a random number generated from a uniform distribution to the genotype frequencies of the ancestral population, which are $(1 - q_{eq})^2 + q_{eq}(1 - q_{eq})F_{IS}$, $2q_{eq}(1 - q_{eq})(1 - F_{IS})$, ${q_{eq}}^2 + q_{eq}(1 - q_{eq})F_{IS}$ for the *AA*, *Aa* and *aa* genotype, respectively.

For a comprehensive analysis of allele frequencies at mutation–selection equilibrium under partial selfing, see Ohta and Cockerham (1974) and Holsinger (1986).

To assess genetic diversity, we also simulated 200 freely recombining neutral loci, which have initially two alleles in approximately equal frequencies. The expected and observed heterozygosity and number of alleles per locus, averaged over loci and replicate lines, were monitored over generations.

Life cycle

The order of operations in each generation is mutation, reproduction, and viability selection. For each of the demographic scenarios considered (described below), we simulated the reproduction of a hermaphroditic population with a mixed mating system; the probability of self-fertilization ranges from $Self = 0$ (complete outcrossing) to $Self = 1$ (complete selfing). The production of an offspring within each population is simulated as follows. A mother plant is chosen at random from the population. To decide whether self-fertilization occurs, we draw a random number from a uniform distribution in the range [0, 1]. If *Self* is greater than the random number, the ovule is self-fertilized. Otherwise, we choose a father plant at random from the population. The expected selfing rate is a function of population size in the parental generation, N_t, and is given by the equation $Self = 0.313 - 0.0007 \times N_t$ in accordance with the observations of Spigler et al. (2010).

The fitness of the offspring produced is calculated from the fitness function of equation (1) and compared to a random number generated from a uniform law between 0 and 1. If fitness is higher than this number, the offspring is retained otherwise, two parents are again chosen. This process is repeated until N offspring are produced. One hundred replications were run for six populations (as in the experimental study) that evolve in parallel under each scenario; results show the mean of all the replications.

Demographic considerations

We considered three different demographic scenarios:

1. Fragmented populations of stable population size ($N = 50$, 100, 150; "fragmentation scenario"). A previously large and continuous population is subdivided to six small populations of $N = 50$, 100, or 150 individuals each. For simplicity, population size is considered stable and equal to the fragment's carrying capacity. We present the mean results for each population size during early (1st to 10th) and late (91st to 100th) generations.
2. Populations undergo a single bottleneck (bottleneck scenario). We assume that the initial generation coincides with the bottleneck event and we follow the populations during their recovery. The populations recover gradually following a density-dependent growth. To model population growth, we used the classic discrete time Beverton–Holt model, according to which population size in generation $t + 1$ is given by:

$$N_{t+1} = f(N_t) = \left\| \frac{e^r N_t}{1 + (e^r - 1)N_t / K} \right\|, \qquad (4)$$

where N_t is the population size in generation t (rounded to the nearest integer), r is the intrinsic rate of increase, and K the carrying capacity (e.g., Brannstrom and Sumpter 2005). The following parameter values were used: $N_0 = 20$, $r = 0.1$, and $K = 300$. This growth rate is comparable to that of other fragmented forb populations (e.g., Oostermeijer 2000; Brys et al. 2004; Ramula et al. 2007). In this scenario, differences in population sizes reflect differences in time since the bottleneck occurred, that is, larger populations were reduced in size in a more distant past.

3. Populations subject to frequent disturbances resulting in successive bottlenecks (disturbance scenario). The populations under study occur in a patchy and disturbed habitat. Bottlenecks should, therefore, be frequent. To simulate this situation, we worked out a scenario with disturbances occurring with probability f_d each generation. A perturbation occurs if f_d is greater than a random number between 0 and 1 drawn at random from a uniform distribution. Disturbances result in a drastic decrease in population size; we refer to the multiplicative effect of a disturbance on population size as disturbance effect, e_d. After a disturbance event, recovery follows the density-dependent growth of equation (4). This is an ascending pattern of disturbances (sensu Wichmann et al. 2003) according to which the decrease in size is sudden but the increase may be gradual. In the results shown, we used the following parameter values: $r = 0.5$, $K = 300$, $f_d = 0.1$ (i.e., disturbances are expected to occur every 10 generations), $e_d = 0.1$ (for instance, a population of 100 flowering plants will be reduced to 10 after a disturbance). All populations have an initial size of 150 individuals. However, to facilitate comparisons with the stable fragment scenario, we present estimates of mean inbreeding depression or heterosis over all replicate populations that show census sizes approximately equal ($\pm$ 10%) to $N = 50$, 100, or 150 during early (first to 10th) and late (91st to 100th) generations. In this way, data presented are analogous to snapshots of the fluctuating replicate populations at different population sizes and points in time. With increasing time, different populations will, thus, build different demographic trajectories according to the number of perturbations. The joint choice of intrinsic rate of increase, r, and carrying capacity, K, produce populations that are viable and show mean population sizes in the range of the experimental populations ($N = 50$–150). Although similar results can be obtained with different combinations of r and K, it is important

to note that slow growing populations are not viable under the disturbance regime. More rapid growth might also be expected after disturbance if it also opens microsites and/or promotes recruitment from the seed bank (Eriksson and Eriksson 1997; Sletvold and Rydgren 2007).

Inbreeding depression and heterosis

Inbreeding depression and heterosis are defined as described above for the empirical data (see Section "Statistical Analyses"). Mean fitnesses of progeny resulting from selfing and within-population crosses were computed, before selection, by producing 50 progeny through complete selfing and complete outcrossing, respectively, for each population. The mean fitness of progeny resulting from between-population crosses was computed by producing 50 progeny for each population with pollen coming in random from the other five populations in approximate accordance with the experimental protocol. We used expected heterozygosity, H_e, as a proxy for effective population size and examined patterns of inbreeding depression and heterosis in relation to both census size and H_e.

Results

EMPIRICAL PATTERNS

Neither fruit set ($F_{2,298.4} = 0.04$, $P = 0.96$) nor seed set ($F_{2,662} = 1.32$, $P = 0.27$) resulting from initial experimental hand pollinations was affected by cross type (Fig. 1). However, inbreeding depression was expressed at virtually all other postdispersal life stages (Fig. 1). Selfed offspring had significantly lower germination rates ($F_{1,138.3} = 15.95$, $P = 0.0001$), lower early survival ($F_{1,182.5} = 7.27$, $P = 0.008$), lower survival to flower ($F_{1,765} = 5.13$, $P = 0.02$), fewer flowers ($F_{1,47.5} = 5.63$, $P = 0.02$), and lower seed set ($F_{1,24.5} = 5.08$, $P = 0.03$), although for seed set this only approaches significance when considering correction for multiple tests. For all traits, the population-by-cross interaction was not significant, revealing ubiquitous inbreeding depression across populations at these life stages. In contrast to pervasive inbreeding depression throughout the lifespan, heterosis was significant only for survival to flower (Fig. 3; $F_{1,765} = 20.5$, $P < 0.0001$). Table S2 contains detailed model results.

Cumulative disadvantages to selfed offspring throughout the life cycle led to significant inbreeding depression for lifetime fitness on average across populations (full ZINB model S-WP contrast $\chi^2_{df=2} = 14.28$, $P = 0.0008$; Fig. 2). The magnitude of lifetime inbreeding depression was substantial; selfed individuals were 82% less fit compared to outcrossed individuals, based on mean self and within-population cumulative lifetime fitness across populations. Average heterosis too was significant and high (mean heterosis = 0.53; full ZINB model WP-BP contrast $\chi^2_{df=2} = 16.91$, $P = 0.0002$; Fig. 2). The presence of inbreeding depression for lifetime fitness was consistent across all populations, as indicated by nonsignificant population-by-cross interaction terms ($P > 0.05$) in both submodels comparing self- versus within-population crosses. However, the presence of heterosis was not; there was a significant population-by-cross interaction in the negative binomial submodel, which predicts counts of lifetime fitness (i.e., fitness of those that survived to reproduce), comparing within- versus between-population crosses ($\chi^2_{df=5} = 13.85$, $P = 0.02$). Table S3 in Supporting Information contains detailed model results.

There was a significant negative relationship between the magnitude of population-level inbreeding depression and $\overline{N_c}$ ($F = 10.7$, $P = 0.03$; Fig. 3A). Inbreeding depression was *greater* in *smaller* populations and declined in the larger populations. Inbreeding depression ranged from 0.28 in population BC1 to 1.0 in BC2, and census population size explained 73% of this variation. In contrast, inbreeding depression did not vary with H_e ($F = 0.97$, $P = 0.38$; Fig. 3B). The magnitude of heterosis tended to increase with $\overline{N_c}$ ($F = 3.22$, $P = 0.15$; Fig. 3C), although power to detect significance is limited. Heterosis, however, declined significantly with increasing H_e ($F = 7.77$, $P = 0.05$; Fig. 3D). $\overline{N_c}$ and H_e were not correlated ($r = -0.45$, $P = 0.37$).

Relationships between stage-specific inbreeding depression and heterosis and either $\overline{N_c}$ or H_e generally mirrored overall trends for lifetime fitness (Table S4). In particular, $\overline{N_c}$ was consistently negatively correlated with inbreeding depression at all stages ($-0.64 \leq r \leq -0.25$, $P > 0.10$) and positively correlated with stage-specific heterosis ($0.13 \leq r \leq 0.52$, $P > 0.3$), though not statistically significant. Relationships between H_e and stage-specific heterosis were consistently negative ($-0.98 \leq r \leq -0.24$), significantly so for offspring fruit set ($r = -0.98$, $P < 0.001$) and marginally so for early survival and seed set of offspring ($r = -0.80$ and -0.79, respectively, $P = 0.06$).

THEORETICAL SIMULATION RESULTS

We find that the relationship of N_c with both inbreeding depression and heterosis depends jointly on (1) the number of generations since the initial disturbance event (fragmentation, bottleneck, onset of repeated disturbances) and (2) the specific demographic scenario (Fig. 4A and C). Later generations are associated with lower inbreeding depression and higher heterosis (for fragmentation and disturbance scenarios, compare triangle points with circles in Fig. 4). In the bottleneck scenario, this translates into a negative relationship between inbreeding depression and N_c (i.e., as populations recover). However, we note that the slope of this relationship depends considerably on the intrinsic rate of increase, r; the decline is steep when r is low, but only moderate to slight as r increases (see Fig. S3). Among demographic

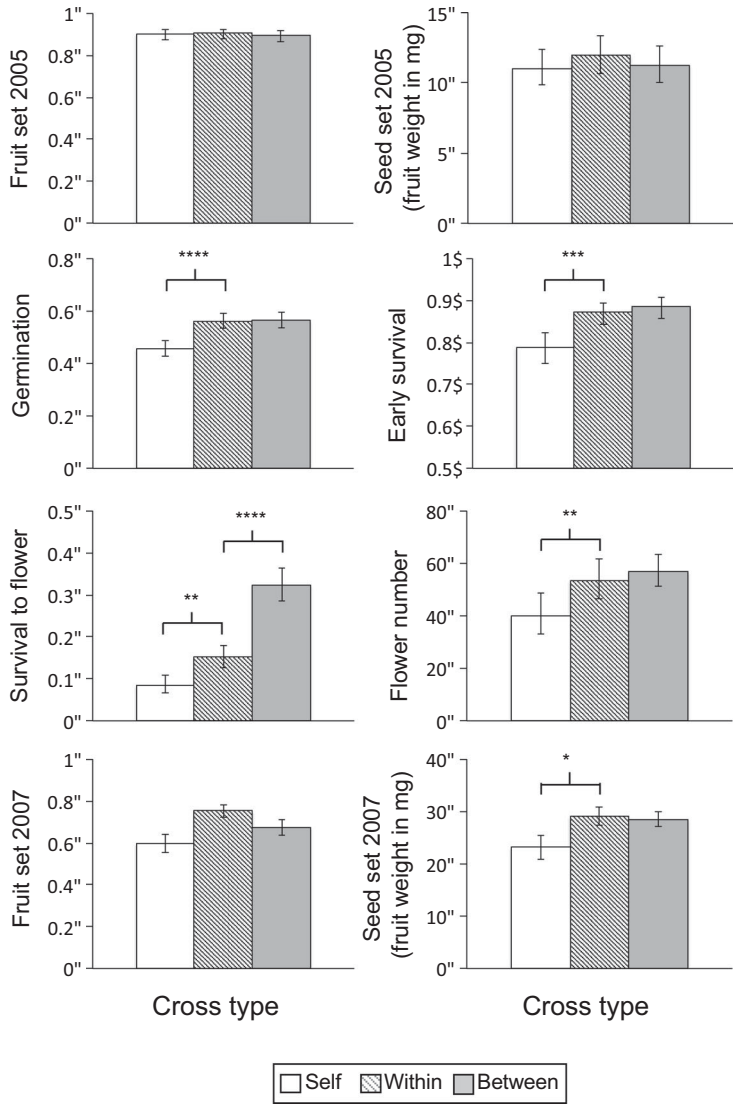

Figure 1. **Fitness components for three cross types (self, within-population, and between-population) averaged across six *Sabatia angularis* populations. (A) Initial fruit set in 2005, (B) initial seed set in 2005, (C) germination in 2006, (D) early survival in 2006, (E) survival to flower in 2007, (F) flower number in 2007, (G) fruit set in 2007, and (H) seed set in 2007. Least squares means ± SE are presented in each panel. Where appropriate, means and SEs were back-transformed (see Methods). Significant differences are based on analyses as described in methods. *P* values for pairwise comparisons based on preplanned contrasts are indicated as follows: *≤0.05, **≤0.025, ***≤0.01, ****≤0.001.**

scenarios, disturbed populations show a more prominent change in both inbreeding depression and heterosis with time (generations) at disequilibrium. This not only implies that populations at early stages after fragmentation would have a substantially higher inbreeding depression than those in later generations, but also that disturbed populations would have, *ceteris paribus*, lower inbreeding depression and higher heterosis than stable ones of the same size (compare hatched triangles with filled ones in Fig. 4A and C). Inbreeding depression and heterosis are not predicted by N_c under disturbance.

Our model results further reveal that inbreeding depression and heterosis are not well predicted by H_e for populations at demographic (and genetic) disequilibrium (Fig. 4B and D). Populations can show very different levels of inbreeding depression for the same H_e, depending on their population size or demographic history. With respect to the latter, populations under a history of repeated disturbances diverge from stable ones in the magnitude of inbreeding depression, and the direction can depend upon the level of heterozygosity. For example, populations under disturbance show similar or even slightly greater levels of inbreeding depression compared to larger stable populations when H_e is relatively high but can show lower inbreeding depression when H_e is low (compare the disturbance scenario with the stable population of $N = 150$ in Fig. 4B). The relationship between H_e and inbreeding

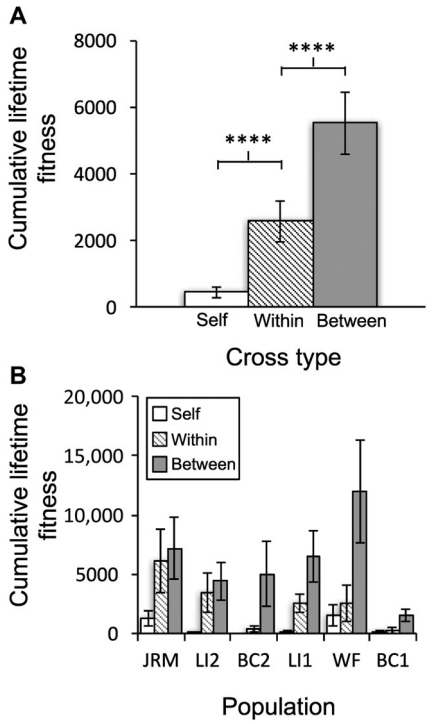

Figure 2. Lifetime cumulative fitness for three cross types across six *Sabatia angularis* populations. (A) Mean fitness of self, within-population, and between-population cross types across populations (± SE). *P* values for pairwise comparisons are indicated as follows: *≤0.05, **≤0.025, *≤0.01, ****≤0.001. (B) Mean lifetime fitness of cross types separated by population (± SE). Populations are presented in order of small to large from left to right. Presented means and SEs are based on raw data.**

depression becomes even more complicated when populations are recovering from a bottleneck. In early stages of recovery (i.e., earlier generations when H_e is highest), initial declines in genetic diversity are accompanied by a decrease in inbreeding depression, creating a positive relationship between inbreeding depression and heterozygosity (Fig. 4B). However, this is transformed to a negative one in later stages of recovery from the bottleneck and then plateaus. These single bottleneck dynamics as modeled here are governed primarily by changes in the number of segregating loci rather than the fixation rate (Fig. S4).

Demography determines strongly the relationship of H_e and heterosis as well. We see that for the same level of genetic diversity (1) smaller populations would show higher heterosis than larger ones and (2) populations under strong fluctuations in size, as in the disturbance scenario, would show substantially higher heterosis than stable populations (Fig. 4D).

Discussion

We reveal a negative relationship between inbreeding depression and $\overline{N_c}$ in *S. angularis*. In addition, heterosis decreased with H_e. The latter pattern meets predictions from existing theory, suggesting negative fitness impacts of genetic drift, and is consistent with recent empirical studies on the relationship between heterosis and H_e (Willi et al. 2013; Lohr and Haag. 2015). However, a pattern of decreased inbreeding depression with increasing $\overline{N_c}$ is largely unexpected. One generally predicts that inbreeding depression and population size (or at least H_e) should positively covary (see Introduction). Next, we discuss these results, theoretical predictions for populations at demographic disequilibrium, and the extent to which the theoretical predictions can explain our empirical patterns.

HIGHER INBREEDING DEPRESSION IN THE SMALLEST *S. angularis* Populations?

Relatively lower inbreeding depression in our relatively larger populations is unlikely due to purging of deleterious recessive alleles for three reasons. First, if purging had reduced inbreeding depression, we would expect to see significantly higher absolute fitness of selfed individuals in these populations, but absolute fitness of selfed individuals does not vary with mean census size (Fig. 2, $r = 0.13$, $P = 0.81$). Second, purging of deleterious recessive alleles is not consistent with substantially reduced survival and reproduction of selfed offspring in this system, which should result in selective interference among loci contributing to inbreeding depression (Lande et al. 1994; Winn et al. 2011). Third, even the size of our relatively larger populations may be too small for efficient purging (Glémin 2003). These lines of reasoning instead point toward the role of drift. Certainly, the two largest populations with the lowest inbreeding depression also experienced high heterosis. One caveat with our estimates of heterosis is that, because our formulation does not allow us to determine the underlying genetic factors, we cannot say with certainty that levels seen here are due solely to masking of deleterious recessive alleles, in which case we would expect absolute fitness of progeny resulting from between-population crosses to be equivalent across populations. Instead, we observed variation in absolute fitness of these progeny among populations, as was also seen in a study comparing heterosis across *Leavenworthia* populations (Busch 2006). However, we note that absolute fitness of progeny derived from within-population crosses tended to decline with $\overline{N_c}$ ($r = -0.63$, $P = 0.18$), significantly so based on rank correlation ($r_s = -0.83$, $P = 0.04$), and strongly declined with heterosis ($r = -0.89$, $P = 0.02$), as would be expected if genetic drift were shaping the patterns of inbreeding depression and heterosis found here. Additional among-population variation seen in fitness of within- and between-population progeny might be attributable to the degree of difference in environmental factors between the novel common garden and each population's site of origin.

Although the small number of populations and range of population sizes under study limited our statistical power, we find our

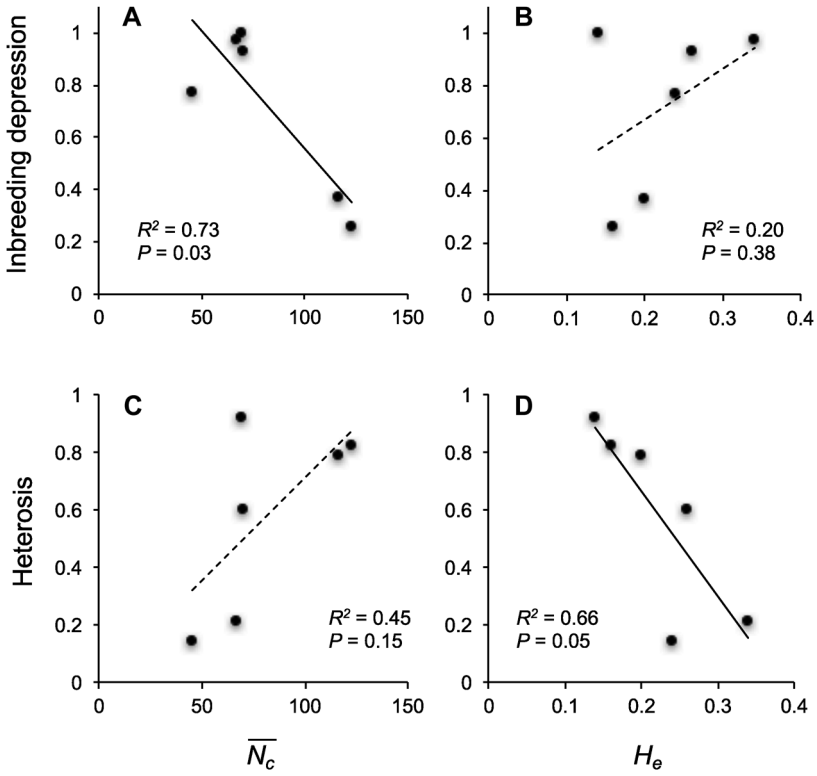

Figure 3. Relationships between (A) inbreeding depression and mean census population size ($\overline{N_c}$), (B) inbreeding depression and genetic diversity (H_e), (C) heterosis and $\overline{N_c}$, and (D) heterosis and H_e across size *Sabatia angularis* populations. Regression lines and corresponding statistics are shown.

results compelling. First, our finding of reduced inbreeding depression with increasing census size is consistent with data from a previous study including the same populations (Spigler et al. 2010). In that study, the magnitude of the change in F between maternal and offspring generations within *S. angularis* populations (delta-F)—one index for the strength of selection against inbred progeny given certain assumptions (Brown 1979)—decreased with census population size. Remarkably, those delta-F estimates are highly correlated with our experimentally derived inbreeding depression estimates ($r_s = 0.94$, $P = 0.005$). Second, we found consistent patterns between inbreeding depression and census size across the life cycle, such that no single stage was driving the results. Third, the range of population sizes evaluated here, though small, is typical for the species in the area under study. Although several studies investigating the influence of census size on inbreeding depression have included populations of >1000 plants (e.g., Ouborg and Van Treunen 1994; Paland and Schmid 2003), for some species >100 may be considered large (e.g., Oakley and Winn 2012). Certainly that is the case in *S. angularis*, wherein 75% of wild populations have ≤ ~200 adults ($N = 54$ populations, R. B. Spigler, unpubl. data), and we concede that the small range may in fact may be precisely why we detected such patterns. That is, any of our populations may have recently declined in size or are in the process of rebounding and thus unlikely to be at demographic equilibrium. Independence among N_c, H_e, and F (Spigler et al. 2010) and the populations' patchy distribution across disturbed habitats (road sides, power line right of ways, patches of private property) support this assertion. Moreover, mowing seen during flowering establishes that at least some populations are subject to disturbances. Consequently, assumptions of theoretical models predicting the relationship between inbreeding depression and population size, census or effective, at equilibrium would not be met and populations may vary in their departure from equilibrium.

THEORETIC PREDICTIONS AT DISEQUILIBRIUM

What are the expected relationships between inbreeding depression, heterosis, census population size, and genetic diversity if populations are in demographic disequilibrium? Our model results reveal surprising and complex dynamics of inbreeding depression and heterosis that depart from previous predictions and depend strongly on population dynamics. For one, genetic diversity alone cannot adequately predict inbreeding depression in nonequilibrium conditions. We see populations showing dissimilar levels of inbreeding depression despite similar H_e. In addition, we find several scenarios that could account for greater inbreeding depression in small populations in our empirical data. First, this could arise if smaller populations are under fragmentation

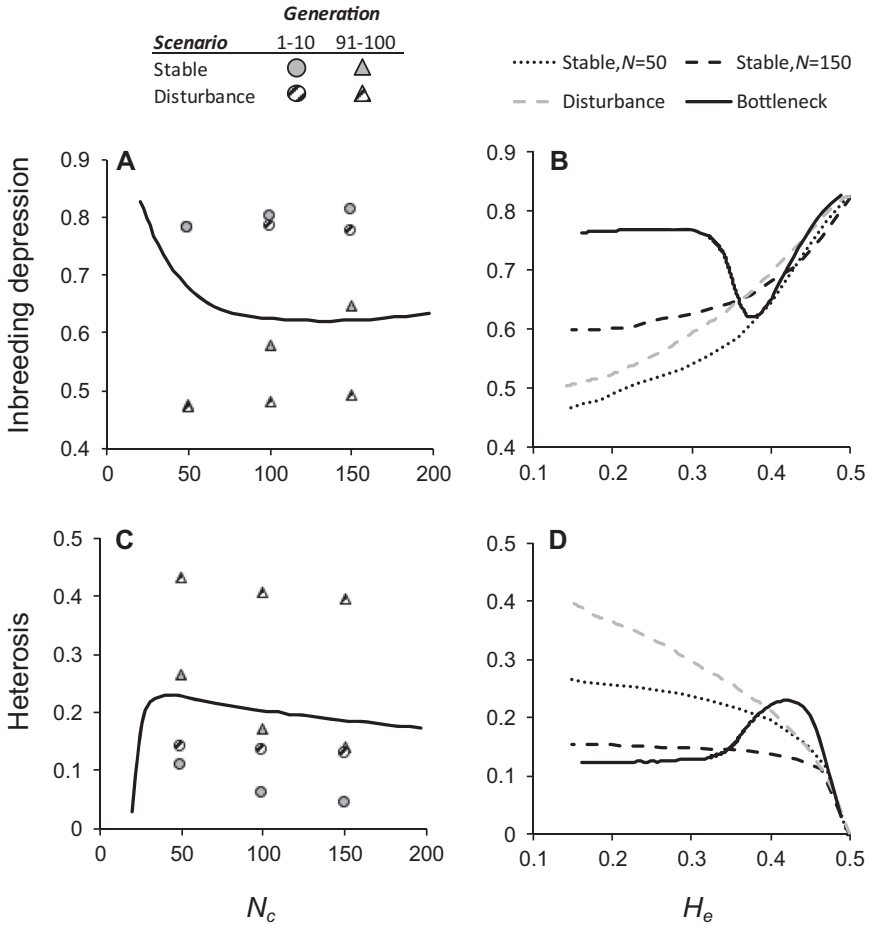

Figure 4. A summary of the expected relationship between: (A) inbreeding depression and census population size, N_c; (B) inbreeding depression and genetic diversity, H_e; (C) heterosis and N_c; and (D) heterosis and H_e. Results from three demographic scenarios are shown (see inset legends): (1) fragmentation ($N_c = 50$, 100, and 150, "stable"), (2) a single bottleneck event of $N_0 = 20$ individuals that is allowed to recover to carrying capacity ($K = 300$) with an intrinsic rate of increase of $r = 0.1$, and (3) repeated bottlenecks ("disturbance') that occur with a frequency of $f_d = 0.1$ and multiplicative effect on its size, $e_d = 0.1$. For the fragmentation and disturbance scenarios, values shown are means either in early (first to 10th) or late (91st to 100th) generations. Note that all populations in the disturbance scenario come from populations with the same overall mean size ($N \approx 150$); data shown for $N = 50$, 100, and 150 were obtained by finding the mean inbreeding depression or heterosis over all replicate populations that show census sizes approximately equal (± 10%) to $N = 50$, 100, or 150 during early (1st to 10th) and late (91st to 100th) generations. In this way, data presented for the disturbance scenario are analogous to snapshots of the fluctuating replicate populations at different population sizes and points in time. Note that the points in panel A representing early disturbance (hatched circle) and late stable (gray triangle) for $N_c = 50$ are not visible because they are obscured by the other points at $N_c = 50$.

for fewer generations, which is the same interpretation as given in Wang et al. (1999) and Theodorou and Couvet (2006) for their models demonstrating higher inbreeding depression in small populations. Second, under slow recovery from a bottleneck, larger populations represent a longer recovery time and so show lower inbreeding depression and higher heterosis. This finding is in agreement with results of Kirkpatrick and Jarne (2000), considering slow recovery. Third, in the disturbance scenario, populations that have experienced more fluctuations should show lower inbreeding depression (and higher heterosis) than stable populations. Importantly, all populations in this scenario were modeled with the same initial (and average) size, so that it is the temporal variance in size that shapes inbreeding depression. Thus, decreased inbreeding depression in the relatively larger *S. angularis* populations might indicate that they have experienced greater fluctuations in size, underscoring the need to measure variance in population size across time rather than a static snapshot. Although our larger *S. angularis* populations did show greater variance in size across years, two years of census data are insufficient to generalize this theoretical finding to our empirical one at this time. Generally speaking across scenarios, decreases in inbreeding depression in smaller populations and/or fluctuating

ones seen from our simulations were due to a higher rate of fixation of mildly deleterious alleles lowering offspring fitness from within-population crosses because of less efficient selection and stronger genetic drift (Fig. S4). However, we would also see an increase in absolute fitness of selfed offspring due to purging of strongly detrimental alleles (Fig. S4), although the higher fixation rate is more important in the disturbance scenario.

Which scenario most likely explains the relationships seen in our empirical data? The relationship between genetic diversity and N_c can help us discriminate. The absence of a relationship between H_e and N_c across *S. angularis* populations rules out the fragmentation scenario, which implies a strong negative relationship between H_e and N_c if larger populations represent those that have been fragmented for a longer period of time. Under the bottleneck scenario, only a weak negative relationship between H_e and N_c is expected (Fig. S5), which could be undetectable with our small sample size or because of limited variation in allozyme markers. Finally, under the disturbance scenario, fluctuating dynamics eliminate any correlation between N_c and H_e (Fig. S5) in accordance with our data. We note that no single scenario may apply to all populations. But broadly speaking, our other empirical results align with predictions from the disturbance scenario: (1) high heterosis and (2) a stronger negative relationship between heterosis and H_e compared to a relatively weaker relationship between inbreeding depression and H_e. These occur because exposure to recurrent disturbances leads to a steep, continuous increase in loci fixed for slightly deleterious alleles as generations go by and H_e decreases due to drift, in contrast to transient effects from a single bottleneck (Fig. S4). The relationship between inbreeding depression and H_e should be weaker because the number of loci segregating for strongly detrimental alleles, responsible for a great part of inbreeding depression, changes little after initial decreases in homozygosity in early generations (Fig. S4). These conclusions also highlight differences between inbreeding depression and heterosis dynamics: inbreeding depression should respond more rapidly to population changes because the number of loci segregating for strongly detrimental alleles changes faster with population size than the number of loci underlying heterosis. Instead, genetic diversity is expected to be the decisive parameter for heterosis.

Conclusions

Our work highlights the importance of considering the interaction between demography and genetic diversity when predicting genetic load across populations at disequilibrium, demonstrating that very different inbreeding depression and heterosis levels can arise across populations even when they show similar H_e. Such considerations could offer a possible explanation for the inconsistency across studies that have empirically tested the relationship between inbreeding depression and population size (Paland and Schmid 2003; Angeloni et al. 2011; Hauser and Loeschcke 1994; Ouborg and Van Treuren 1994; Michaels et al. 2008; Coutellec and Caquet 2011; Willi et al. 2013). Although the mix of results could reflect variation in mutation rate and dominance levels across taxa (Lynch et al. 1999) or the stochastic nature of purging in natural populations (Byers and Waller 1999; Crnokrak and Barrett 2002), our models show that different patterns between genetic load and population size could also reflect differences in population demography and environment. We recognize that our empirical sample size is small, limiting our conclusions, but hope that this work stimulates research on nonequilibrium dynamics, the need for which is becoming more acute in the face of increasing habitat loss and fragmentation.

ACKNOWLEDGMENTS

We thank Plant Biology Greenhouse staff, especially M. Boyd, and E. Fyfe for field and greenhouse assistance. S. Hubbell, J. Hamrick, R. Pulliam, C. Heckel, and S. Glémin provided helpful comments and discussion. This article was improved by additional comments by R. Shaw, A. Case, L. Fishman, R. Gomulkiewicz, and anonymous reviewers. Support and funding for this work was awarded to RBS from Adkins Arboretum, Georgia Botanical Society, Georgia Native Plant Society, Georgia Museum of Natural History, Highlands Biological Station, National Science Foundation Graduate Research Fellowship, North Carolina Native Plant Society, Plant Biology Small Grant Awards, UGA Graduate Student Assistantship, and Temple University.

LITERATURE CITED

Allendorf, F. W., G. H. Luikart, and S. N. Aitken. 2013. Conservation and the genetics of populations. 2nd ed. Wiley-Blackwell, Chichester, West Sussex, U.K.

Angeloni, F., N. J. Ouborg, and R. Leimu. 2011. Meta-analysis on the association of population size and life history with inbreeding depression in plants. Biol. Conserv. 144:3–43.

Armbruster, P., and D. H. Reed. 2005. Inbreeding depression in benign and stressful environments. Heredity 95:235–242.

Bataillon, T., and M. Kirkpatrick. 2000. Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. Genet. Res. 75:75–81.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. Stevens, H Henry, and J.-S. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. 24:127–135.

Brannstrom, A., and D. J. T. Sumpter. 2005. The role of competition and clustering in population dynamics. Proc. R. Soc. B Biol. Sci. 272:2065–2072.

Brown, A. H. D. 1979. Enzyme polymorphism in plant-populations. Theor. Popul. Biol. 15:1–42.

Bush, J. W. 2006. Heterosis in an isolated, effectively small, and self-fertilizing population of the flowering plant *Leavenworthia alabamica*. Evolution 60:184–191.

Brys, R., Jacquemyn, H., Endels, P., Blust, G. D., and Hermy, M., 2004. The effects of grassland management on plant performance and demography in the perennial herb *Primula veris*. J. Appl. Ecol. 41:1080–1091.

Byers, D. L., and D. M. Waller. 1999. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. Annu. Rev. Ecol. Syst. 30:479–513.

Caballero, A., and P. D. Keightley. 1994. A pleiotropic nonadditive model of variation in quantitative traits. Genetics 138:883–900.

Carr, D. E., and M. R. Dudash. 2003. Recent approaches into the genetic basis of inbreeding depression in plants. Philos. Trans. R. Soc. B 358:1071–1084.

Charlesworth, B., and D. Charlesworth. 1978. A model for the evolution of dioecy and gynodioecy. Am. Nat. 112:975–997.

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. 18:237–268.

Charlesworth, D., and J. H. Willis. 2009. The genetics of inbreeding depression. Nat. Rev. Genet. 10:783–796.

Coutellec, M. A., and T. Caquet. 2011. Heterosis and inbreeding depression in bottlenecked populations: a test in the hermaphroditic freshwater snail *Lymnaea stagnalis*. J. Evol. Biol. 24:2248–2257.

Crnokrak, P., and S. C. H. Barrett. 2002. Purging the genetic load: a review of the experimental evidence. Evolution 56:2347–2358.

Crnokrak, P., and D. A. Roff. 1999. Inbreeding depression in the wild. Heredity 83:260–270.

Crow, J., and M. Kimura. 1970. An introduction to population genetics theory. Harper & Row, New York.

Crow, J. F., and M. J. Simmons. 1983. The mutation load in *Drosophila*. Pp. 2–35 in C. Ashburner, and J. N. Thompson Jr., eds. The genetics and biology of *Drosophila*. Academic Press, New York.

Darwin, C. R. 1876. The effects of cross- and self-fertilization in the vegetable kingdom. John Murray, London.

Delmas, C. E., P.-O. Cheptou, N. Escaravage, and A. Pornon. 2014. High lifetime inbreeding depression counteracts the reproductive assurance benefit of selfing in a mass-flowering shrub. BMC Evol. Biol. 14: 1–12.

Dudash, M. R. 1990. Relative fitness of selfed and outcrossed progeny in a self-compatible, protandrous species, *Sabatia angularis* L. (Gentianaceae)—a comparison in 3 environments. Evolution 44:1129–1139.

Eriksson, Å., and O. Eriksson. 1997. Seedling recruitment in semi-natural pastures: the effects of disturbance, seed size, phenology and seed bank. Nord. J. Bot. 17:469–482.

Falconer, S. 1960. Introduction to quantitative genetics. Oliver & Boyd, Edinburgh, London.

Fisher, R. A. 1941. Average excess and average effect of a gene substitution. Ann. Eugenic. 11:53–63.

Frankham, R. 2005. Genetics and extinction. Biol. Conserv. 126:131–140.

García-Dorado, A. 2003. Tolerant versus sensitivity genomes: the impact of deleterious mutations on fitness and conservation. Conserv. Genet. 4:311–324.

Gillespie, J. H. 1998. Population genetics: a concise guide. John Hopkins Univ. Press, Baltimore, MD.

Glémin, S. 2003. How are deleterious mutations purged? Drift versus nonrandom mating. Evolution 57:2678–2687.

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

Haag-Liautard, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan, B. Charlesworth, and P. D. Keightley. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. Nature 445:82–85.

Hauser, T. P., and V. Loeschcke. 1994. Inbreeding depression and mating-distance dependent offspring fitness in large and small populations of *Lychnis flos-cuculi* (Caryophyllaceae). J. Evol. Biol. 7:609–622.

Holsinger, K. E. 1986. Dispersal and plant mating systems: the evolution of self-fertilization in subdivided populations. Evolution 40:405–413.

Jaquiery, J., F. Guillaume, and N. Perrin. 2009. Predicting the deleterious effects of mutation load in fragmented populations. Conserv. Biol. 23:207–218.

Keller, L. F. and D. M. Waller. 2002. Inbreeding effects in wild populations. Trends Ecol. Evol. 17:230–241.

Kimura, M., T. Maruyama, and J. F. Crow. 1963. The mutation load in small populations. Genetics 48:1303–1312.

Kirkpatrick, M., and P. Jarne. 2000. The effects of a bottleneck on inbreeding depression and the genetic load. Am. Nat. 155:154–167.

Lande, R., D. W. Schemske, and S. T. Schultz. 1994. High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. Evolution 48:965–978.

Lohr, J. N., and C. R. Haag. 2015. Genetic load, inbreeding depression, and hybrid vigor covary with population size: an empirical evaluation of theoretical predictions. Evolution 69:3109–3122.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA.

Lynch, M., J. Conery, and R. Burger. 1995. Mutation accumulation and the extinction of small populations. Am. Nat. 146:489–518.

Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz, L. Vassilieva, and J. Willis. 1999. Spontaneous deleterious mutation. Evolution 53:645–663.

Michaels, H. J., X. J. Shi, and R. J. Mitchell. 2008. Effects of population size on performance and inbreeding depression in Lupinus perennis. Oecologia 154:651–661.

O'Grady, J. J., B. W. Brook, D. H. Reed, J. D. Ballou, D. W. Tonkyn, and R. Frankham. 2006. Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. Biol. Conserv. 133:42–51.

Oakley, C. G., and A. A. Winn. 2012. Effects of population size and isolation on heterosis, mean fitness, and inbreeding depression in a perennial plant. New Phytol. 196:261–270.

Ohta T., and C. C. Cockerham. 1974. Detrimental genes with partial selfing and effects on a neutral locus. Genet. Res. 23:191–200.

Oostermeijer, J. G. B. 2000. Population viability analysis of the rare Gentiana pneumonanthe: the importance of genetics, demography, and reproductive biology. Pp. 313–334 *in* A. G. Young, and G. M. Clarke, eds. Genetics, demography, and viability of fragmented populations. Cambridge Univ. Press, Cambridge, U.K.

Ouborg, N. J., and R. Van Treuren. 1994. The significance of genetic erosion in the process of extinction. 4. Inbreeding load and heterosis in relation to population size in the mint *Salvia pratensis*. Evolution 48:996–1008.

Paland, S., and B. Schmid. 2003. Population size and the nature of genetic load in *Gentianella germanica*. Evolution 57:2242–2251.

Palstra, F. P., and D. E. Ruzzante. 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? Mol. Ecol. 15:3428–3447.

Porcher, E., and R. Lande. 2005. The evolution of self-fertilization and inbreeding depression under pollen discounting and pollen limitation. J. Evol. Biol. 18:497–508.

Quinn, G. P., and M. J. Keough. 2002. Experimental design and data analysis for biologists. Cambridge Univ. Press, Cambridge, U.K.

R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/.

Ramula, S., E. Toivonen, and P. Mutikainen. 2007. Demographic consequences of pollen limitation and inbreeding depression in a gynodioecious herb. Int. J. Plant Sci. 168:443–453.

Robledo-Arnuncio, J. J., R. AlÍA, and L. Gil. 2004. Increased selfing and correlated paternity in a small population of a predominantly outcrossing conifer, *Pinus sylvestris*. Mol. Ecol. 13:2567–2577.

Roze, D., and F. Roussett. 2004. Joing effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. Genetics 167:1001–1015.

Schultz, S. T., and M. Lynch. 1997. Mutation and extinction: the role of variable mutational effects, synergistic epistasis, beneficial mutations, and degree of outcrossing. Evolution 51:1363–1371.

Shaw, R. G., and C. J. Geyer. 2010. Inferring fitness landscapes. Evolution 64:2510–2520.

Sletvold, N., and Rydgren, K., 2007. Population dynamics in *Digitalis purpurea*: the interaction of disturbance and seed bank dynamics. J. Ecol. 95:1346–1359.

Spigler, R. B., and T.-L. Ashman. 2012. Gynodioecy to dioecy: are we there yet? Ann. Bot. London. 109:531–543.

Spigler, R. B., and S.-M. Chang. 2008. Effects of plant abundance on reproductive success in the biennial *Sabatia angularis* (Gentianaceae): spatial scale matters. J. Ecol. 96:323–333.

Spigler, R. B., J. L. Hamrick, and S.-M. Chang. 2010. Increased inbreeding but not homozygosity in small populations of *Sabatia angularis* (Gentianaceae). Plant Syst. Evol. 284:131–140.

Theodorou, K., and D. Couvet. 2002. Inbreeding depression and heterosis in a subdivided population: influence of the mating system. Genet. Res. 80:107–116.

———. 2006. On the expected relationship between inbreeding, fitness, and extinction. Genet. Sel. Evol. 38:371–387.

———. 2015. The efficiency of close inbreedng to reduce genetic adaptation to captivity. Heredity 114:38–47.

Wang, J. L., W. G. Hill, D. Charlesworth, and B. Charlesworth. 1999. Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. Genet. Res. 74:165–178.

Wheeler, R. E. 2010. multResp() lmPerm. The R project for statistical computing. Available at http://www.r-project.org/.

Whitlock, M. C. 2000. Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. Evolution 54:1855–1861.

Whitlock, M. C., P. K. Ingvarsson, and T. Hatfield. 2000. Local drift load and the heterosis of interconnected populations. Heredity 84:452–457.

Wichmann, M. C., K. Johst, K. A. Maloney, C. Wissel, and F. Jeltsch. 2003. Extinction risk in periodically fluctuating environments. Ecol. Model. 167:221–231.

Willi, Y., P. Griffin, and J. Van Buskirk. 2013. Drift load in populations of small size and low density. Heredity 110:296–302.

Winn, A. A., E. Elle, S. Kalisz, P. O. Cheptou, C. G. Eckert, C. Goodwillie, M. O. Johnston, D. A. Moeller, R. H. Ree, R. D. Sargent, et al. 2011. Analysis of inbreeding depression in mixed-mating plants provides evidence for selective interference and stable mixed mating. Evolution 65:3339–3359.

Zurr, A. F. 2009. Zero-truncated and zero-inflated models for count data. Pp. 261–293 *in* A. F. Zurr, E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith, eds. Mixed effects models and extensions in ecology with R. Springer Science+Business Media, LLC, NY.

Associate Editor: A. Case
Handling Editor: R. Shaw

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Map of study population locations at three scales.

Figure S2. Relationship between two estimators of heterosis and a table of their values for each study population.

Figure S3. Inbreeding depression (upper panel) and heterosis (lower panel) plotted against population size for populations recovering from a bottleneck ($N_0 = 20$ individuals).

Figure S4. The predicted number of (A) segregating and (B) fixed loci as a function of expected heterozygosity, H_e, for four demographic scenarios.

Figure S5. Scatterplot showing the relationship between expected heterozygosity, H_e, and census population size, N_c, for the bottleneck and the disturbance scenarios.

Table S1. Locations of all populations included in the study including focal study populations where crosses to estimate inbreeding depression and heterosis were performed and additional populations used as occasional pollen sources.

Table S2. *F*-statistics for fixed effects from general and generalized linear mixed models testing the effect of cross type, population, and their interaction on (A) fruit and seed set resulting from initial experimental cross pollinations and (B) components of offspring lifetime fitness.

Table S3. Likelihood ratio chi-square statistics from a zero-inflated negative binomial model testing the effect of cross type, population, and their interaction on cumulative lifetime fitness.

Table S4. Pearson correlation coefficients exploring the relationships (A) inbreeding depression and (B) heterosis for fitness components throughout the life history and either mean census population size or expected heterozygosity (H_e).