Estimating inbreeding and its effects in a long-term study of snapdragons (*Antirrhinum majus*)

Louise Arathoon,*[1,4] Parvathy Surendranadh,*[1] Nicholas Barton,[1] David L. Field,[2] Melinda Pickup,[1,3] Carina A. Baskett[1]

*Joint first authors
[1]IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria
[2]School of Science, Edith Cowan University, 270 Joondalup Drive, Joondalup WA 6027 Australia
[3]Greening Australia, 1 Underwood Avenue, Shenton Park, WA 6008, Australia
[4]Corresponding author

Abstract

Inbreeding depression can be estimated by correlating heterozygosity with fitness components, but such heterozygosity-fitness correlations are typically weak. For over ten years, we studied a population of the self-incompatible plant, *Antirrhinum majus*, measuring heterozygosity and fitness proxies from 22,353 plants. Using a panel of 91 SNPs, we find that relatedness declines rapidly over short spatial scales. Individual heterozygosity varies more between individuals than expected, reflecting identity disequilibrium (g_2) due to variation in inbreeding – a prerequisite for detecting inbreeding depression. We use two types of simulations to ask whether the heterozygosity distribution is consistent with spatially structured mating. First, we simulate offspring from matings with fathers at different distances and find that the distribution of heterozygosity in the field data is consistent with the measured pollen dispersal kernel. Second, we simulate a 1000-generation pedigree using the known spatial distribution, and find that identity disequilibrium, though highly variable between simulations, is consistent with that observed. Finally, we estimate inbreeding depression through the relationships between heterozygosity and six fitness proxies. Only the number of flowering stems is predicted by heterozygosity. Our approach provides a novel example of how long-term studies can elucidate population structure and fitness variation in the wild.

Keywords

inbreeding depression, heterozygosity, identity disequilibrium, population structure

Introduction

Inbreeding depression is one of the most universal and long-standing observations in biology: fitness is reduced in offspring of matings between relatives, leading to the evolution of extraordinary mechanisms for the avoidance of inbreeding [1]. With the advent of molecular markers, it became possible to estimate inbreeding depression from

contemporary samples, without requiring pedigrees. Thus, both the effects of inbreeding on fitness and its genetic underpinnings can be studied in nature.

The discovery of extensive allozyme variation stimulated a vigorous debate over whether that variation is maintained by balancing selection [2, 3]. Positive correlations between heterozygosity and fitness proxies were interpreted as being directly due to the allozyme alleles, and therefore, as evidence for overdominance (i.e., superiority of heterozygotes at single loci) [4]. However, when such heterozygosity-fitness correlations (HFC) were also observed with DNA-based markers, such as microsatellites, it became clear that they reflect an association with inbreeding depression: individuals more heterozygous at marker loci tend to be more heterozygous for the recessive deleterious alleles that are responsible for inbreeding depression [5, 6]. Overdominance is now thought to make a relatively small contribution to inbreeding depression [7].

Given the ubiquity of inbreeding depression due to recessive deleterious alleles, HFCs should be widespread and easily detected in nature, yet meta-analyses have not consistently found significant HFCs (e.g., [8]). A possible explanation is that the study populations lack sufficient variation in inbreeding: if all individuals are equally inbred, then there is no correlation between heterozygosity at unlinked loci, and HFC can only be due to the small fraction of inbreeding depression that is linked to the markers [6]. Miller and Coltman [8] calculated identity disequilibrium (a measure of covariance in heterozygosity between loci) for 50 HFC studies and found that it was rarely significant, indicating that HFC studies usually do not detect sufficient variation in inbreeding. Their analysis included only studies of vertebrates, which avoid inbreeding by having separate sexes; plants, on the other hand, have a greater range of adaptations to avoid self-fertilisation or mating among relatives, and thus may show different patterns of inbreeding.

In addition to its importance for detecting HFCs, variation in inbreeding is interesting in itself because it contains information about demographic history. Traditionally, effective population size and dispersal rate are inferred from the pattern of isolation-by-distance – that is, by how the probability of identity by descent (IBD) between two genes decays with spatial separation [9]. Isolation by distance leads to a reduction in *mean* heterozygosity, relative to that in a well-mixed population. Spatial variation in inbreeding, which generates excess *variance* in heterozygosity, gives a different measure of association, between two pairs of genes, and gives us additional information about population structure. There is a plethora of different statistics that aim to reveal different aspects of population structure [10]. An important open question is how far we can combine these measures to make detailed inferences about demographic history [11].

Here, we use a detailed long-term study of a plant population to bring together measures of isolation by distance with variance in inbreeding and to ask whether distributions of heterozygosity can be explained by what we know about spatial population structure. For over a decade, we have sampled fitness-related traits and SNPs in a hybrid zone between two *Antirrhinum majus* subspecies with different flower colours. Our population-wide sampling permits a powerful test of whether heterozygosity is correlated with fitness. First, we verify if there is excess variation in heterozygosity, which reflects an underlying

variance in inbreeding. Second, we compare the empirical distribution of heterozygosity to that of offspring from simulated matings with nearby individuals at different dispersal scales. Third, we ask whether the distribution of heterozygosity is consistent with population structure generated by simulating a pedigree over 1000 generations, conditional on observed locations of plants. Finally, we ask whether fitness proxies are correlated with heterozygosity, to examine limits on the strength of inbreeding depression.

Methods

Study System and Field Sampling

Antirrhinum majus is a self-incompatible, hermaphroditic, short-lived perennial herb native to the Iberian Peninsula. It has a seed bank; most individuals' parents are from 3-4 years before they are sampled (D. Field, unpublished data). It grows in a variety of microhabitats with relatively bare soil or frequent disturbance, including rail embankments, rocky cliffs, and regularly mowed roadsides. Our study includes two subspecies defined by flower colour: *A. majus pseudomajus* has magenta flowers and occurs in northern Spain and south-western France, including the Pyrenees. *A. majus striatum* has yellow flowers and a smaller range, encircled by *A. m. pseudomajus*. The subspecies are parapatric; where they meet, narrow hybrid zones are readily formed, and there is no evidence for post-zygotic reproductive barriers [12]. We focus on such a hybrid zone in the Vall de Ribès, Spain [13], where we have collected demographic data on the majority of the population annually since 2009. There is little divergence between the subspecies in this population across nearly all of the genome, except for regions involved in floral pigmentation [14].

Genetic samples and trait measurements were obtained annually from 2009-2019 from every flowering individual that we could access in the well-mixed "core" of the hybrid zone (~1 km) and the surrounding ~2 km of "flanks" on either side where parental flower colours dominate. Most of the sampling was conducted along two parallel roads that cross the Vall de Ribès, dubbed the "lower road" (LR; ~1150 m elevation; GIV-4016) and "upper road" (UR; ~1350 m; N-260). We also sampled along small side roads, railroad embankments, rivers, and hiking trails; density was very low away from these disturbed areas. The total sample size across this area, summed over all years is 22,353 plants, ranging from ~750 plants in the smallest year (2018), to ~5500 plants in the largest year (2014). Each year there were fewer than 100 visible but inaccessible plants, so we estimate that we sample the majority of individuals in the study area. Sampling was conducted during the peak of flowering (early June to mid-July), some years including late May and a second visit in August. Areas adjacent to the roads are mowed multiple times in the flowering and fruiting season by the municipality.

For each plant, we collected leaf samples for genotyping, recorded locations, and measured several traits (described under Heterozygosity-Fitness Correlations). Leaf samples were refrigerated upon return to the field station, and dried in silica gel within three days. Locations were recorded with GeoXT handheld GPS units (Trimble,

Sunnydale, CA, USA) and are accurate to within 3.7 m, which we determined by comparing accidentally duplicated samples within and between years.

SNP panel and isolation by distance

Previously, a panel of 248 SNPs spread throughout the genome was designed for the focal hybrid zone (see methods in [15]). We follow these methods but include an additional five years of data (2015-2019) and a subset of 91 SNPs; the mean sample size per SNP was 21,212, or ~95% of the sample. (See Supplemental Material (SM) for SNP filtering methods). We checked for isolation by distance by first randomly imputing the missing genotypes based on the allele frequencies at respective markers and then calculating $F_{ST}(r)$ (defined in SM) for distance classes of 20m (Fig. S1). This was calculated for each year and then averaged over years to obtain an F_{ST} of 0.0244.

Variation in inbreeding

To describe the variation in heterozygosity in the *A. majus* population, we calculated multilocus heterozygosity, denoted here by H, pooling across all years. (Since there is seed dormancy, and plants may live several years, we cannot clearly define "generations." We do not expect pooling to have much effect.) This is defined as the fraction of heterozygous loci in an individual: $= \frac{1}{n}\sum_{i=1}^{n} h_i$, where h_i = 0 or 1, according to whether locus *i* is heterozygous, and there are *n* loci. Offspring from related individuals tend to have lower heterozygosity.

We observed an excess of individuals around half the mean heterozygosity (see Results). To check whether the pattern was consistent with rare selfing mutants, we compared the likelihood of a single Gaussian to a mixture of two Gaussian distributions, one with the observed mean and variance and the other with half its mean and variance.

The variance in individual heterozygosity consists of two components. The first is due to the variance in whether or not an individual locus is heterozygous, and decreases in proportion to the number of SNP, n: it equals $(1-F)^2\langle 2pq(1-2pq)\rangle/n$, where < > is an average over loci. The second is due to covariance in heterozygosity between loci, which is termed the identity disequilibrium (ID). For a given pedigree, heterozygosity is independent across unlinked loci, and so this second component is proportional to the variance in inbreeding across individuals. The first component can be estimated from the allele frequencies, or simply by shuffling the data across loci, so as to eliminate ID. The excess variance is proportional to the variance in F across individuals, and is measured by the statistic g_2:

$$g_2 = \frac{\sum_{i \neq j} \mathrm{cov}[h_i, h_j]}{E[H]^2} = \frac{\mathrm{var}(F)}{(1 - E[F])^2}$$

(from Eq. 1 in [6]). Here, cov[h_i,h_j] is the ID between loci i and j, and the sum over all distinct i,j is the excess variance in H due to ID. Dividing by the square of the mean

heterozygosity $E[H]^2$ eliminates dependence on allele frequency, such that g_2 estimates the variance in F across individuals.

Significant ID is a fundamental condition for the existence of heterozygosity-fitness correlations [6]. To describe the variance of inbreeding across individuals, we first check if the variance in the distribution of individual heterozygosity is significantly greater than the average variance obtained from 100 replicates by shuffling data across loci. We then computed g_2 using the g2_snps function from the R package InbreedR (in R version 3.6.1 [16]), which implements a modified formula for large data sets to estimate g_2, tests whether the value is significantly different from zero via permutation tests, and provides confidence intervals via bootstrapping [17]. We decomposed ID into components due to linked and unlinked SNPs by comparing correlations of H for all individuals to those likely to share recent ancestors, at several scales: across all pairs, within linkage groups, and between adjacent SNPs (Table S1).

Effects of pollen dispersal on heterozygosity

With isolation by distance, the distribution of heterozygosity is expected to depend on the distance between parents: heterozygosity of offspring from nearby parents will have a lower mean and higher variance compared to offspring from distant parents. We show this in the SM, by simulating offspring using all field individuals as mothers and choosing fathers from a given distance away (Fig. S2). Then, we compared the distribution of H between the field data and offspring simulated from matings with three models of pollen dispersal. Fathers were chosen from: the nearest neighbour to the mother, a Gaussian distribution ($\sigma = 300$ m; see explanation in "Heterozygosity in a spatially structured pedigree"), and a leptokurtic dispersal kernel sampled from 1463 empirical measurements of pollen dispersal, estimated as the distances between assigned parents (SM; D. Field, unpublished data). The genotype of the offspring was assigned using Mendelian inheritance. The simulations were implemented both with and without linkage between markers, using the known linkage map (SM; D. Field, unpublished data). Including linkage did not substantially change results, so we mainly show results for unlinked markers. We compared distributions, means, and variance of H using Kolmogorov-Smirnov tests, t-tests, and F-tests, respectively. For the leptokurtic pollen dispersal simulation, we checked for an excess of low-heterozygosity individuals generated by mating between close relatives by asking whether a mixture of two Gaussian distributions is more likely than a single Gaussian distribution.

Heterozygosity in a simulated spatial pedigree

In order to compare actual variation in heterozygosity with a null model, we simulated a continuous population in two-dimensional space, conditioned on the known locations of the individuals, using Mathematica 12.0 [18]. This differs from commonly used models (e.g., island [19], stepping stone [20] and continuous Wright-Malécot models [9, 21]) in that we include spatial patchiness by specifying actual locations to determine relationships in the pedigree. Here, we only provide a proof-of-principle, by asking whether a plausible model of spatial structure can explain observed heterozygosity. We do not include all features of the actual population – in particular, we repeatedly sample

ten years of spatial distributions; we ignore linkage; we assume Gaussian pollen dispersal; we simplify the self-incompatibility system; and we assume an annual life cycle (no perenniality or seed bank). Thus, our simulation parameters should be seen as "effective" values, analogous to the traditional N_e.

The simulation was adapted to resemble the hybrid zone of *A. majus*, for which individual locations are available for the years 2009 to 2019; there were fewer individuals from 2017-2018, so these were merged, giving distribution data for 10 time points. We simulate a region of ~ 1.1 x 1.8 km, which was sampled consistently. We subsample from these locations, so as to maintain a constant population size (N). We randomly sample from the ten consecutive time points, and repeat for 100 cycles, thus iterating for 1000 generations. This naïve approach allows us to simulate a spatial structure that is realistic over at least small scales (Fig. S4, S5). Genetic diversity is shaped over the coalescent timescale ($2N_e$, ~170,000 generations in *A. majus* [14]), which is far longer than the 1000 generations that we simulate. However, we are concerned here with the *local* population structure that determines the variation in inbreeding amongst individuals within an area of a few km^2, which will equilibrate rapidly.

First, we generate the pedigree, by choosing parents for each individual according to a backwards dispersal distribution (see methods for choosing parents in SM). Once the spatial pedigree is generated, genotypes are assigned by dropping genes down the pedigree, starting with the known genotypes from 2009. In fact, one could start with any initial frequencies, since F_{ST}-like measures are independent of them. Here, we distinguish two definitions of identity by descent (IBD), each appropriate for different questions. Given some pedigree, we can calculate the probability of IBD of any pair of genes; we denote this by F_p [22]. Whether a pair of genes is in fact IBD depends on the random outcome of meiosis. If we sample across the genome, using a finite number of loci, we will find that a fraction F_g of genes are IBD. This has expectation F_p, but even if we sample the whole genome, there may be considerable variation, since large blocks of the genetic map will share ancestry. We take the view that the evolutionary process generates a random pedigree, with some F_p, and that random meiosis then generates F_g [11]. If we are concerned with inbreeding depression, then F_g is the relevant measure, whereas if we want to understand how population structure has shaped the pedigree, then we need to know F_p. With enough SNPs, we can estimate F_g more accurately than we could from the pedigree [23]. Conversely, F_p is only imperfectly estimated by F_g, which can only tell us about genetic ancestry: beyond a few generations, most pedigree ancestors do not pass on any genetic material at all [24]. Because we generate the pedigree and genotypes in separate steps, we can distinguish variation in F_p from that in F_g.

F_{ST} is obtained by calculating F_p from three replicate pedigrees, and the pattern of isolation by distance was compared to the data (Fig. S6A). The simulated population size and dispersal were adjusted in order to match the observed F_{ST} from the data, calculated from the small simulated area. Three replicate sets of genotypes were run for each of three replicate pedigrees. F_{ST} was also obtained by calculating F_g from replicate genotype sets assigned for one pedigree (Fig. S6B). Values of g_2 were calculated for each replicate from each pedigree using InbreedR. A direct estimate of g_2 was also obtained from the

pedigree (see g_2 definition in "Variation in inbreeding"). We compare both estimates of g_2 to that obtained from the field data in the same sampling area that was simulated.

Heterozygosity-fitness correlations

We recorded traits related to size and fecundity for every sampled individual in the field surveys from select years: number of flowering stems (2009-2019), number of open flowers (2010-2019), height of the tallest stem (cm; 2014-2019), and number of leaf nodes on the tallest stem (2014-2019). Depending on accessibility, areas were visited 1-3 times each year. Thus, trait measurements are a snapshot taken at a random point in time, rather than a maximum. We address the potential added noise from this constraint by including measurement date as a covariate in HFC analyses. For perennial plants, which have trait measurements in multiple years, the year with the highest measured value for each trait was included in pooled year analyses.

We assessed whether the traits measured on every individual can be used as fitness proxies by asking whether they are related to more direct measures of fecundity. In 2014-2017, 1047 individuals in un-mowed areas were sampled at the end of the season. The total number of fruit and flowers borne throughout the season can be inferred from inflorescence nodes and remains of dehisced fruits. We call the snapshot flower counts performed on all individuals "estimated flowers" to distinguish them from the "total flower" counts performed for this subset. Impossible values were filtered out for reach trait (e.g., height = 10m). We asked whether fitness proxies (flowering stems, estimated flowers, total flowers, height, leaf nodes) correlate with female fecundity as assessed by total fruit counts, using Spearman's ρ.

To test whether fecundity and fitness proxies were correlated with heterozygosity, we used generalized linear models (GLMs) in R, using the MASS package for negative binomial models [25]. Response variables were fitness proxies: total fruit, total flowers, number of flowering stems, stem height, and leaf nodes. Because not all markers are typed across individuals, the predictor was standardised multilocus heterozygosity (H^*), defined as the proportion of heterozygous loci per individual, divided by the population mean heterozygosity for those typed loci [26], calculated with InbreedR [17]. Each model also contained fixed-effect covariates of year, road (upper vs. lower, designated according to latitude), and perennial status ("annual" for individuals collected only once; "perennial" if a genotype or tag was collected in multiple years). Julian day of measurement was an additional covariate for stem height, estimated flowers, and number of flowering stems. The family and link function for the GLM was assigned according to the distribution of each trait: negative binomial (to account for overdispersion) for total flowers, total fruit, estimated flowers, and number of flowering stems; and gamma with a log link for stem height.

The residuals of the models indicate spatial autocorrelation (significantly positive Moran's I), so variances are likely underestimated. However, due to a lack of significant regression coefficients (see Results), we did not correct model variance for spatial autocorrelation in most models. For the significant result, number of flowering stems, we

accounted for spatial autocorrelation with a generalised linear mixed model (GLMM) with an exponential spatial correlation structure nested within road and year, using the MASS and nlme packages [25, 28, 29].

Correlations between heterozygosity and fitness are due to the product of the correlation of inbreeding with both heterozygosity and fitness (Eq. 2 in [6]). Using this relationship, the results of the HFC analysis can be used to calculate the inbreeding load, $\beta_{f,w}$, for each trait [6]. Although most traits showed non-significant relationships, the potential magnitude of inbreeding load was calculated for each trait (formulas in SM).

Results

Variation in inbreeding

Variance in the distribution of individual heterozygosity (H) in the field data suggests that there is variance in inbreeding in the population (Fig. 1). Furthermore, there is an excess of individuals with around half the mean heterozygosity (i.e., with H~0.22, rather than 0.44; Fig. 1, blue, lower left). These might be due to a low rate of selfing. Indeed, a mixture between two Gaussian with means ~ 0.22 and 0.44, and variances in the same ratio, fits significantly better than a single Gaussian (Fig. 1, compare red and black to blue) with an increase of likelihood of 11.3. However, we shall see in the next section that this excess is also consistent with matings between close relatives, without the need to invoke a breakdown in self-incompatibility.

The observed distribution of heterozygosity can be compared with the distribution when data are shuffled across loci (Fig. 1, grey), eliminating identity disequilibrium (ID). The mean (0.44602) necessarily remains the same, but the observed variance (var(H)=0.00336) is significantly higher than the average variance in 100 shuffled replicates (mean var(H) = 0.00282, s.d. 0.000029). This excess variance implies that the mean standardised ID is g_2 = 0.0029 (95% CI: 0.0026-0.0036), a significant variance in inbreeding between individuals.

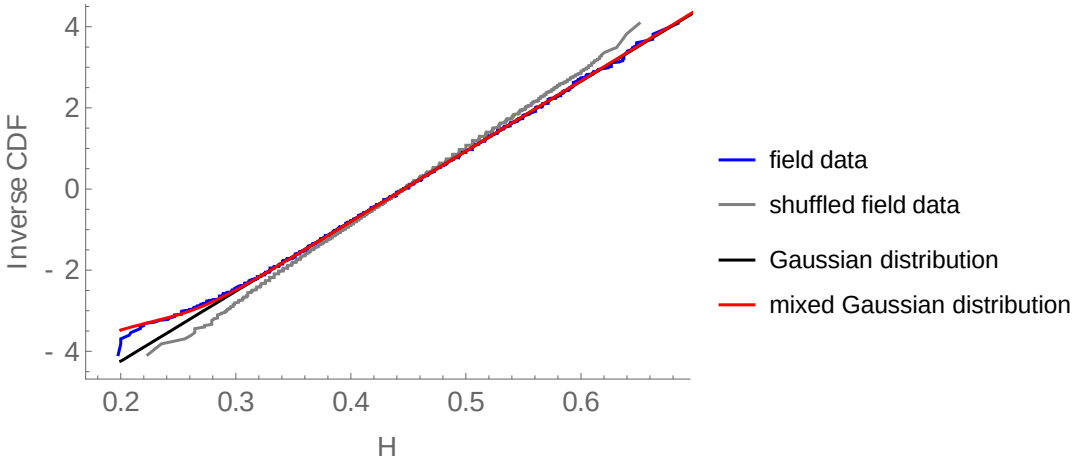

Figure 1. Probit transform of the distribution of individual heterozygosity (H) from the field data (blue), one replicate of the marker-shuffled field data (grey) compared to a single Gaussian (black) and better-fitting mixed Gaussian (red) distribution.

Effects of pollen dispersal on heterozygosity

In the SM, we show that heterozygosity of offspring depends on distance between parents (Fig. S2). Next, we checked whether offspring between parents separated by the actual pollen dispersal kernel have a distribution of heterozygosity consistent with three simulated pollen dispersal scenarios. The mean and variance of heterozygosity do not significantly differ between the field data and simulated offspring from leptokurtic dispersal, but do differ from simulated matings with nearest neighbours and with Gaussian dispersal (Fig. 2A, Table S2, Table S3). On the other hand, all three dispersal schemes differ in the tail as judged by Kolmogorov-Smirnov tests (Table S3). These comparisons were made for a single replicate, but because each involves 22,353 individuals, there is little variation in mean and variance between replicates.

We next examined deviations in the left tail of the distribution, where an excess of low heterozygosity individuals might arise from selfing, or matings between close relatives. We focussed on the leptokurtic dispersal, which gave a bulk distribution consistent with the field data. We estimated the increase in likelihood between fitting a single versus mixed Gaussian distribution (see "Variation in inbreeding") for 100 replicate simulations. The mixed Gaussian is a better fit than a single Gaussian, as we find a significant increase in likelihood for 69 of 100 replicates; the estimated fraction of putatively "selfed" individuals was 0.00043, averaged over replicates, which is about half the estimate from the actual data, 0.00086; only 4/100 replicates gave higher estimates than that observed (Fig. S3). This suggests that the suspected selfed offspring (see "Variation in inbreeding") in the field data, indicated by an excess of individuals with low heterozygosity, can largely be explained by matings between relatives under leptokurtic pollen dispersal, though there is a marginally significant excess of such individuals. There is considerable variation in fit between replicates, simply because deviations in the tail involve few individuals.

The coefficient g_2 reflects excess variation due to identity disequilibrium, and showed similar patterns as the variance in H: g_2 from field data and offspring from simulated matings with leptokurtic pollen dispersal did not differ significantly, but g_2 from Gaussian and neighbour matings were 80% higher than g_2 from field data and leptokurtic matings - a significant difference, as 95% confidence intervals do not overlap (Fig. 2B). However, as we discuss below, these confidence intervals only include sampling error, and not the larger variance between evolutionary realisations.

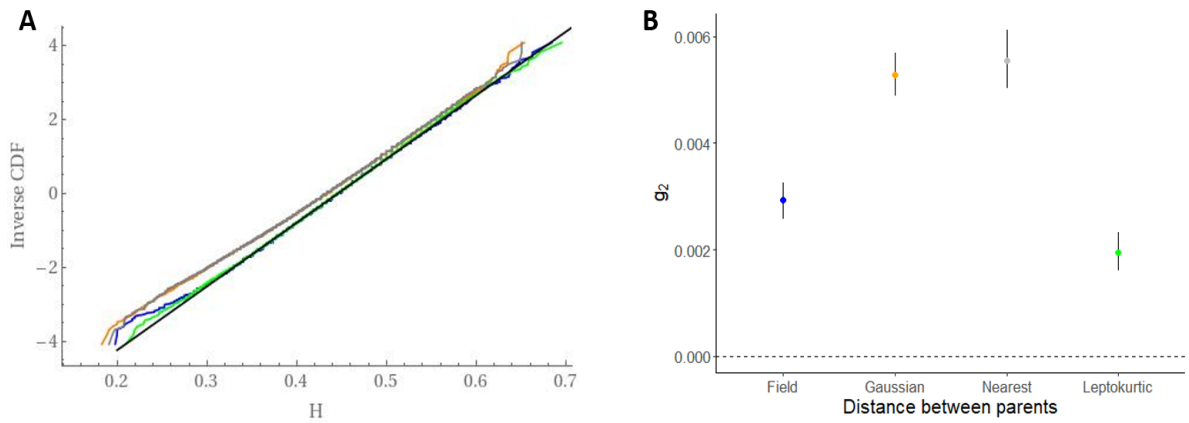

Figure 2. A: Probit transform of multilocus heterozygosity, H, for the field data (blue) versus simulated offspring from a single replicate with leptokurtic pollen dispersal kernel (green), a Gaussian pollen dispersal kernel (orange), and pollen from the nearest neighbour (grey). A normal distribution with mean and standard deviation as that of the field data (black) is included for comparison **B:** Plot of g_2 values for the same data as above indicating mean and 95% CI calculated via bootstrapping.

<u>Heterozygosity in a simulated spatial pedigree</u>

To examine if the observed g_2 is consistent with a simple null model, we simulated a pedigree over 1000 generations, conditioning on the locations of individuals observed over ten years, repeated over 100 cycles. We specified a Gaussian dispersal function, with σ=300m, and chose N=700 individuals, in order to match the observed F_{ST}~0.022 that we re-calculated from the field data for the simulated sub-area of the field site. For example, F_{ST} of 0.0228, 0.0231, 0.0217 was obtained from three replicates of a single pedigree. Figure S8 shows the decline in pairwise identity with distance from the field data for the simulated region, the three replicate pedigrees, and the replicate genotypes of one pedigree. High variation among genotypes suggest that we need many more SNPs to match the pattern from the pedigree; linkage further increases the variance of F_g.

Identity disequilibrium (g_2) estimates from the genotypes from simulated pedigree overlapped with zero in all nine cases (Fig. 3); there was substantial variation between the three simulated pedigrees, and between the three draws of 91 SNPs from each pedigree (Fig. 3). The mean g_2 calculated directly from the three pedigrees is 0.0007, much lower than the lower bound of the observed mean annual g_2 from field data of 0.0019 (calculated for the simulated sub-area). The average g_2 estimated from the 3*3 replicates is -0.0005, which is also lower than the observed g_2 from the field; however, this was not a significant difference, given the wide CI on the simulation due to a small population size (N=700). However, confidence limits for the field data only include error due to sampling a limited number of individuals - and not that due to sampling a limited number of SNP, or the random variation between evolutionary realisations. We return to this issue in the Discussion.

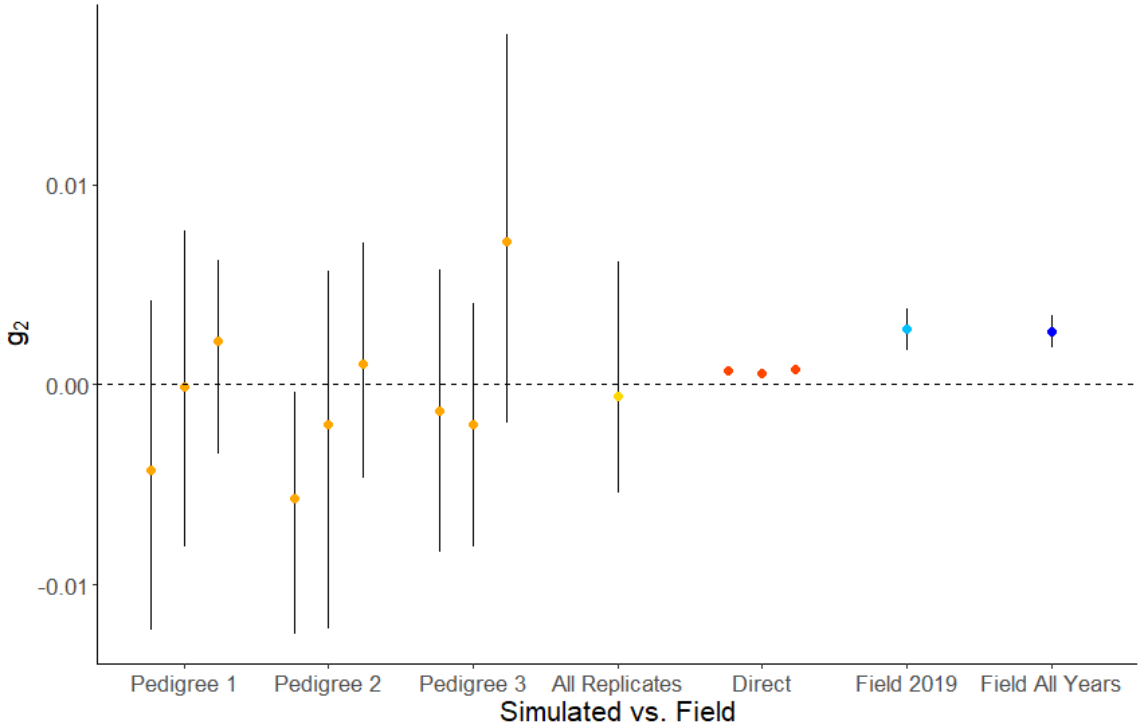

Figure 3. Identity disequilibrium (g_2) calculated from genotypes of the simulated pedigrees (three replicates per pedigree in light orange, mean across replicates in yellow), g_2 calculated directly from simulated pedigrees (dark orange) and from field data (2019 in light blue, mean across all years 2009-2019 in dark blue). The direct pedigree estimates (red) do not have confidence intervals as they contain no error due to sampling of individuals: they are derived from the pedigree of the full (simulated) population.

Heterozygosity-fitness correlations

All size- and fecundity-related traits (stem height, leaf nodes, estimated flowers, number of flowering stems, and total flowers) were significantly positively correlated with total fruits (Spearman's ρ; Fig. S9), justifying use of these traits as fitness proxies.

Even without considering spatial auto-correlation, the standardised heterozygosity H^* was only a significant predictor for one of the six fitness proxies: plants with higher heterozygosity had more flowering stems (Fig. 4, Table S4). This result was still significant after accounting for spatial autocorrelation (data not shown), although it should be noted that model assumptions for normality of residuals was not met. In contrast to heterozygosity, the covariates typically were highly significant predictors of fitness proxies, indicating that it is unlikely that the lack of strongly significant HFCs is due to low power (Table S4).

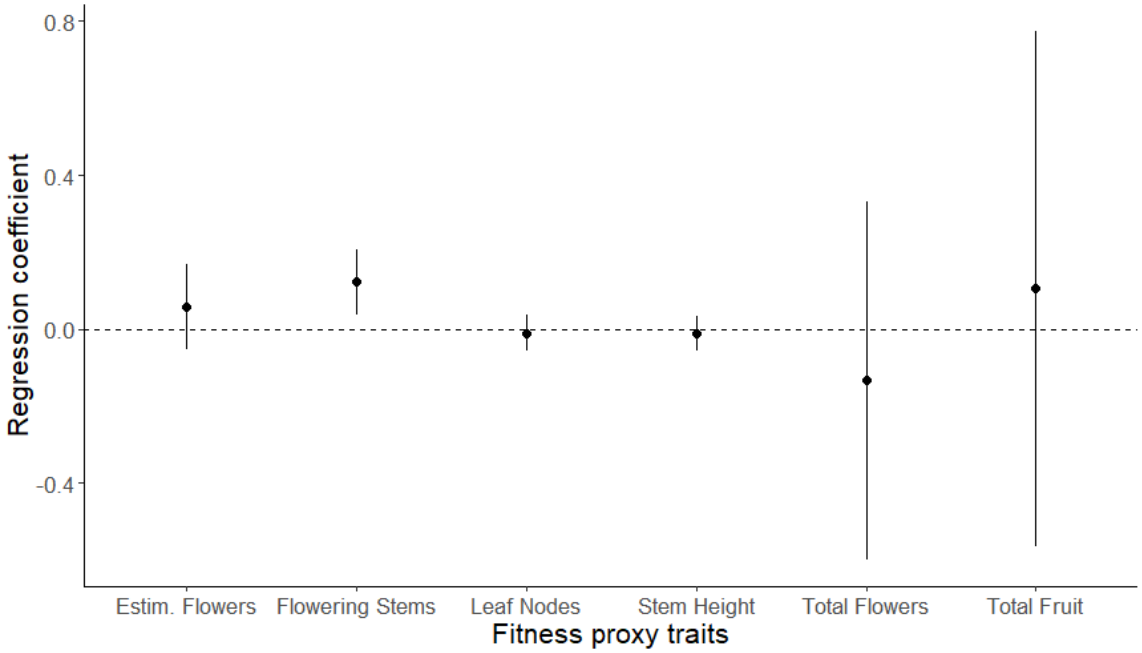

Figure 4. Regression coefficients with 95% confidence intervals for H^* as a predictor of various fitness proxies. Note that coefficients show the percentage change in each trait as H* increases, due to analysis with a log link function.

The regression coefficient of inbreeding with heterozygosity, $\beta_{f,H}$, was calculated to be -0.173. The maximum magnitude of the inbreeding load, $\beta_{f,W}$, for each trait is shown in Table S5, representing the percentage change in each trait value with increasing inbreeding. For the one trait with a significant HFC – number of flowering stems – individuals have approximately 70% less flowering stems for each unit increase in inbreeding (mean $\beta_{f,W}$ = -0.7020, upper CI = -0.2043, lower CI = -1.2000; s.d. of flowering stems = 3.47).

Discussion

Inbreeding increases homozygosity across the genome, leading to inbreeding depression when deleterious recessive mutations are expressed [7]. Inbreeding depression is expected to be strong, especially in self-incompatible (SI) species such as *Antirrhinum majus* [30]; this prediction is borne out for many SI species when inbreeding depression is measured through experimental crosses [31-35]. It also follows that more heterozygous individuals should have higher fitness, allowing the detection of inbreeding depression by quantifying the correlation between heterozygosity and fitness [6]. We tested whether *A. majus* shows evidence of the strong inbreeding depression that would be predicted for a SI species, using a dataset of heterozygosity and fitness proxies measured on over 20,000 plants, sampled from a natural population over 11 years. We first confirmed that there is significant correlation in heterozygosity across markers (g_2, a measure of identity disequilibrium), suggesting variation in inbreeding, such that inbreeding depression has the potential to cause heterozygosity-fitness correlations (HFCs). By simulating offspring from matings between actual parents, we show that the mean heterozygosity increases,

and the variance of heterozygosity decreases, with the distance between parents; strikingly, these changes occur over very short scales (~10m, Fig. S2). We found that the observed distribution of heterozygosity is consistent with a leptokurtic distribution of pollen dispersal. We also simulate the population over 1000 generations, including information on spatial patchiness, and show that this simple null model can generate the observed distribution of heterozygosity. Yet despite significant variation in inbreeding, only one of six fitness proxies (number of flowering stems) increased significantly with heterozygosity. We find little evidence for inbreeding depression in nature, and can set a small upper bound for the four traits for which we have large samples (Fig. 4). Over the range of inbreeding in the population, inbreeding depression makes little to no contribution to fitness variance, and is hard to detect even with our very large sample.

Plant mating systems play an important role in structuring genetic variation within populations. Our simulations showed that individuals are as heterozygous as expected with random mating and no selfing. Self-incompatibility systems are more complex than the simple simulation rule to forbid selfing. *Antirrhinum majus* has a gametophytic system (GSI [36]), whereby the pollen detoxifies secretions from the style, unless the pollen and style genotypes share alleles at the S-locus (Fuji et al. 2016). This system prevents selfing, but given that related plants are more likely to share S-alleles, also reduces mating among relatives (biparental inbreeding [37, 38]). Thus, we might expect that our simulated matings would have lower mean heterozygosity than the empirical measurement; yet, we found no evidence for such an effect. This result agrees with the conclusion of Leducq et al. [39], which found in theoretical models of SI systems that GSI had little effect on biparental inbreeding. Indeed, we found that a small amount of biparental inbreeding with leptokurtic pollen dispersal may explain an excess of individuals with low heterozygosity of about half the mean (Fig. 2, Fig. S3), although we have low statistical power to distinguish this scenario from rare selfing mutants.

Many studies that test for HFCs find that the excess variation in heterozygosity, g_2, which arises from identity disequilibrium, is low [8]. In our study, we estimate that g_2=0.0029 (95% CI: 0.0026-0.0036). This corresponds to a coefficient of variation of non-identity ($1-F_g$) of only $\sqrt{0.0029}$ =0.054. Thus, it is not surprising that it is difficult to detect inbreeding depression via HFCs: in our study, we can only set an upper limit on inbreeding depression, despite our exceptionally large sample size. The empirical identity disequilibrium was consistent with that from simulated pedigrees conditioned on the actual spatial distribution of plants, indicating that patchiness may contribute to patterns of heterozygosity. Our data were consistent with simulations, even though (for technical reasons) we could not explore smaller σ, which would require much larger N. Thus, in simulations the identity did not fall as sharply over very short distances as observed (Fig. S8). Comparing simulations with data highlights the complexities of demographic and ecological processes such that fitting IBD and introducing a seed bank remains a challenge for understanding population structure in this system.

However, comparing ID from our data to the simulations illustrates a general problem with statistical comparisons in evolutionary biology. There are three sources of error in estimating g_2: from sampling a limited number of individuals, from sampling a limited

number of SNP, and from random variation between evolutionary realisations. The first source (a limited number of individuals) is shown by the confidence intervals in Fig. 3, obtained by bootstrapping across individuals [17]. These CI are very tight for the field data, because of our large sample size, and wider for the simulations because they are based on only 700 individuals. Yet these 700 individuals are the entire (simulated) population, so arguably, there is no error at all due to sampling individuals. Thus, we also show values of g_2 calculated directly from the pedigree – and therefore without error, because F_p is calculated exactly for the whole (simulated) population. The second source of error (a limited number of SNPs) is shown by the substantial variation in g_2 of the 3*3 genetic replicates around the pedigree values (even without taking into account partial linkage). This source of error is not reflected in the confidence intervals for the field data; even though we find significant identity disequilibrium (implying significant variation in F_g), given more SNPs we might well see lower variation in the underlying F_p, as in the simulations. Finally, the wide variation in estimates of g_2 due to random meiosis, and the random generation of the pedigree (Fig. 3) should remind us that typically our estimates of parameters are limited by the randomness of evolution (the third source of error), far more than simply by the number of individuals or SNPs that we sample.

Although identity disequilibrium makes it possible in principle to detect HFCs, we only found a significant positive correlation between heterozygosity and the number of flowering stems. This result is surprising because self-incompatible plants should have high inbreeding depression [30, 38]. The lack of HFCs could be due to several factors. First, the excess variation in inbreeding should make it easier to purge deleterious alleles [40, 41], and reduces the mutation load. However, as noted above, the coefficient of variation in non-identity is small (~0.054), making this a weak effect. Second, identity disequilibrium, while significant, may still be too low to allow detection of weak HFCs, especially with a small set of markers [8]. Third, the fitness proxies available in our complete dataset are correlated with fecundity (total fruit set) measured on a subset of plants, but these relationships are noisy. Environmental and stochastic effects on traits related to size and fecundity (e.g., the strong covariate effects in Fig. S10) may swamp fitness effects due to heterozygosity. Furthermore, we simply used the largest trait value available for perennial plants, as we lack a reasonable calculation for total lifetime fitness. A final explanation for the lack of HFCs is the "invisible fraction," the many individuals that die before their fitness is evaluated [42]. Our methods would not detect early-acting inbreeding depression that would cause fruit abortion, germination failure, seedling death, or failure to flower; we can only detect variation in heterozygosity and fitness among flowering individuals, which are typically a very small portion of the total number of individuals counted from seeds. For example, in the short-lived perennial *Olearia flocktoniae*, < 1% of viable seeds survive to germination and only 2.5% of those survive the earliest seedling stage [43]. Contributions to inbreeding depression from these early life-stages have not yet been evaluated using experimental crosses in wild *A. majus*, so it is unclear whether the low inbreeding depression we measured is realistic.

In conclusion, our study fits the broader pattern that HFCs are difficult to detect, even with an exceptionally large dataset and with significant identity disequilibrium. However, we found identity disequilibrium to be an interesting measure of population structure in

its own right, which complements the more widely used pairwise identity, F_{ST}. Our future work will focus on understanding fitness variation, both due to flower colour and more generally, and on understanding how population structure influences genetic diversity. Here, we have only begun to investigate how the distribution of heterozygosity is shaped by a highly heterogeneous population structure. How a patchy and transient distribution, leptokurtic dispersal, a seed bank, and variance in fitness together shape the interplay between selection and drift poses a challenging problem for understanding population structure.

Acknowledgements

We thank the many volunteers and friends who have contributed to data collection in the hybrid zone over the years, in particular those who have managed field seasons: Barbora Trubenova, Maria Clara Melo, Tom Ellis, Eva Cereghetti, Lenka Matejovicova, Beatriz Pablo Carmona. Frederic Ferrer and Eva Salmerón Mateu have been immensely helpful with logistics at our informal field station, El Serrat de Planoles.

References

1. Darwin C. The Effects of Cross and Self fertilisation in the Vegetable Kingdom. Londom: John Murray; 1876.
2. Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D, Charlesworth B. The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. Evolution. 2019;73(1):111-4.
3. Lewontin RC, Hubby JL. A moleuclar approach to study of genic heterozygosity in natural populations 2. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics. 1966;54(2):595-609.
4. Zouros E, Singh SM, Miles HE. Growth rate in oysters: An overdominant phenotype and its possible explanations. Evolution. 1980;34(5):856-67.
5. Houle D. Allozyme-associated heterosis in *Drosophila melanogaster*. Genetics. 1989;123(4):789-801.
6. Szulkin M, Bierne N, David P. Heterozygosity-fitness correlations: A time for reappraisal. Evolution. 2010;64(5):1202-17.
7. Charlesworth D, Willis JH. The genetics of inbreeding depression. Nature Reviews Genetics. 2009;10(11):783-96.
8. Miller JM, Coltman DW. Assessment of identity disequilibrium and its relation to empirical heterozygosity fitness correlations: A meta-analysis. Molecular Ecology. 2014;23(8):1899-909.
9. Wright S. Isolation by distance. Genetics. 1943;28(2):114-38.
10. Milligan BG, Archer FI, Ferchaud AL, Hand BK, Kierepka EM, Waples RS. Disentangling genetic structure for genetic monitoring of complex populations. Evolutionary Applications. 2018;11(7):1149-61.
11. Bradburd GS, Ralph PL. Spatial population genetics: It's about time. Annu Rev Ecol Evol S. 2019;50:427-49.
12. Andalo C, Cruzan MB, Cazettes C, Pujol B, Burrus M, Thebaud C. Post-pollination barriers do not explain the persistence of two distinct *Antirrhinum* subspecies with parapatric distribution. Plant Systematics and Evolution. 2010;286(3-4):223-34.

13. Whibley AC, Langlade NB, Andalo C, Hanna AI, Bangham A, Thebaud C, Coen E. Evolutionary paths underlying flower color variation in *Antirrhinum*. Science. 2006;313(5789):963-6.
14. Tavares H, Whibley A, Field DL, Bradley D, Couchman M, Copsey L, et al. Selection and gene flow shape genomic islands that control floral guides. Proceedings of the National Academy of Sciences. 2018;115(43):11006-11.
15. Ringbauer H, Kolesnikov A, Field DL, Barton NH. Estimating barriers to gene flow from distorted isolation-by-distance patterns. Genetics. 2018;208(3):1231-45.
16. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
17. Stoffel MA, Esser M, Kardos M, Humble E, Nichols H, David P, Hoffman JI. inbreedR: an R package for the analysis of inbreeding based on genetic markers. Methods in Ecology and Evolution. 2016;7(11):1331-9.
18. Wolfram Research Inc. Mathematica. Champaign, Illinois: Wolfram Research, Inc.; 2019.
19. Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97-159.
20. Kimura M, Weiss GH. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics. 1964;49(4):561-76.
21. Malécot G. The Mathematics of Heredity (English translation, 1969). San Francisco: WF Freeman; 1948.
22. Charlesworth B, Charlesworth D. Elements of Evolutionary Genetics. Greenwood Village, Colorado: Roberts and Company Publishers; 2012.
23. Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. Evolutionary Applications. 2016;9(10):1205-18.
24. Thompson EA. Identity by descent: Variation in meiosis, across genomes, and in populations. Genetics. 2013;194(2):301-26.
25. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. New York: Springer; 2002.
26. Coltman DW, Pilkington JG, Smith JA, Pemberton JM. Parasite-mediated selection against inbred Soay sheep in a free-living, island population. Evolution. 1999;53(4):1259-67.
27. Hughes J. reghelper: Helper functions for regression analysis. R package version 0.3.6. https://CRAN.R-project.org/package=reghelper2020.
28. Dormann CF, McPherson JM, Araujo MB, Bivand R, Bolliger J, Carl G, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. Ecography. 2007;30(5):609-28.
29. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. nlme: Linear and nonlinear mixed effects models. R package version 3.1-131. https://CRAN.R-project.org/package=nlme2017.
30. Lande R, Schemske DW. The evolution of self-fertilization and inbreeding depression in plants. 1. Genetic models. Evolution. 1985;39(1):24-40.
31. Nason JD, Ellstrand NC. Lifetime estimates of biparental inbreeding depression in the self-incompatible annual plant *Raphanus sativus*. Evolution. 1995;49(2):307-16.
32. Busch JW. Inbreeding depression in self-incompatible and self-compatible populations of *Leavenworthia alabamica*. Heredity. 2005;94(2):159-65.

33. Husband BC, Schemske DW. Evolution of the magnitude and timing of inbreeding depression in plants. Evolution. 1996;50(1):54-70.
34. Stift M, Hunter BD, Shaw B, Adam A, Hoebe PN, Mable BK. Inbreeding depression in self-incompatible North American *Arabidopsis lyrata*: Disentangling genomic and S-locus-specific genetic load. Heredity. 2013;110(1):19-28.
35. Winn AA, Elle E, Kalisz S, Cheptou PO, Eckert CG, Goodwillie C, et al. Analysis of inbreeding depression in mixed-mating plants provides evidence for selective interference and stable mixed mating. Evolution. 2011;65(12):3339-59.
36. McCubbin A, Carpenter R, Coen E, Dickinson H. Self-incompatibility in *Antirrhinum*. In: Ottaviano E, Gorla MS, Mulcahy DL, Mulcahy GB, editors. Angiosperm Pollen and Ovules. New York: Springer; 1992.
37. Cartwright RA. Antagonism between local dispersal and self-incompatibility systems in a continuous plant population. Molecular Ecology. 2009;18(11):2327-36.
38. Charlesworth D, Charlesworth B. Inbreeding depression and its evolutionary consequences. Annual Review of Ecology and Systematics. 1987;18:237-68.
39. Leducq JB, Llaurens V, Castric V, Saumitou-Laprade P, Hardy OJ, Vekemans X. Effect of balancing selection on spatial genetic structure within populations: Theoretical investigations on the self-incompatibility locus and empirical studies in *Arabidopsis halleri*. Heredity. 2011;106(2):319-29.
40. Byers DL, Waller DM. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. Annual Review of Ecology and Systematics. 1999;30:479-513.
41. Carr DE, Dudash MR. Recent approaches into the genetic basis of inbreeding depression in plants. Philosophical Transactions of the Royal Society B-Biological Sciences. 2003;358(1434):1071-84.
42. Grafen A. On the uses of data on lifetime reproductive success. In: Clutton-Brock TH, editor. Reproductive Success Studies of Individual Variation in Contrasting Breeding Systems. Chicago: University of Chicago Press; 1988. p. 454-71.
43. Gross CL, Mackay D. Two decades of demography reveals that seed and seedling transitions limit population persistence in a translocated shrub. Ann Bot. 2014;114(1):85-96.