Effects of fine-scale population structure on inbreeding 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

The distribution of heterozygosity in a population is commonly used to quantify inbreeding depression through the use of heterozygosity-fitness correlations (HFCs), but the demographic processes shaping variability in inbreeding are not well understood. For 11 years, we measured heterozygosity and six fitness proxies in a population of the self-incompatible plant *Antirrhinum majus*. Using a panel of 91 SNPs in 22,353 individuals, we find that relatedness declines rapidly over short spatial scales. Excess variance in heterozygosity between individuals (identity disequilibrium, g_2) reflects significant variation in inbreeding. We use two types of simulation to ask whether variation in heterozygosity is consistent with fine-scale spatial patterns of mating and population structure. First, we simulate offspring from matings with fathers at different distances, showing that the pollen dispersal kernel affects expected g_2. Second, we simulate a 1000-generation pedigree using the known dispersal and spatial distribution and find that g_2 is consistent with that observed. In contrast, a simulated population with uniform density underestimates g_2, indicating that heterogeneous density promotes identity disequilibrium. Finally, we estimate inbreeding depression through HFCs. Only flowering stem count increases with heterozygosity. Our study shows that heterogeneous density and leptokurtic dispersal can together explain pairwise F_{ST} and the distribution of heterozygosity.

Keywords

inbreeding depression, heterozygosity, identity disequilibrium, population structure, isolation by distance

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]. The avoidance of self-fertilization through outcrossing would be a sufficient mechanism to avoid inbreeding in a large, randomly mating population. But for many organisms, both

gene dispersal and relatedness are spatially structured, such that individuals closer in space are more likely to mate and more likely to be closely related than individuals further apart. Such spatial population structure should reduce population *mean* heterozygosity, relative to a well-mixed population. Furthermore, real populations are typically patchy rather than uniformly distributed, which should cause spatial variation in inbreeding and hence excess *variance* in heterozygosity. Such variation in heterozygosity has often been used to estimate inbreeding and its fitness effects through heterozygosity-fitness correlations (HFCs) [2, 3]. However, we lack an understanding of what factors (e.g., life history, demography, population structure) are important for shaping the distribution of heterozygosity in a population.

Variation in inbreeding makes it possible, in principle, to estimate inbreeding depression from contemporary samples, without requiring pedigrees. In early studies using allozyme markers, patterns of positive correlations between heterozygosity and fitness proxies were interpreted as being directly due to the allozyme alleles themselves, and therefore, as evidence for superiority of heterozygotes at single loci (overdominance) [4]. However, when such heterozygosity-fitness correlations were also observed with DNA-based markers, 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 [2]. Overdominance is now thought to make a relatively small contribution to inbreeding depression [5].

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., [3]). 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 [2]. Miller and Coltman [3] 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. However, their analysis included only studies of vertebrates. Plants, with their sessile nature and often limited dispersal [6], may be likely to show high identity disequilibrium if relatedness declines rapidly with distance and density is heterogeneous, promoting variation in the frequency of biparental inbreeding.

In addition to its importance for detecting HFCs, variation in inbreeding *per se* is interesting because it depends on demographic history. Yet potential sources of variation in identity disequilibrium have not, to our knowledge, been explored. Mating system can influence variation in inbreeding [7]. Outcrossing species, with generally low levels of inbreeding, provide an opportunity to examine other factors that may affect identity disequilibrium. Here, for an outcrossing organism with fine-scale population structure, spatial patterns of density and mating could have strong effects on the degree of mating with related individuals, and thus affect identity disequilibrium. If the sources of variation in inbreeding are better understood, we may be able to combine identity disequilibrium

with other statistics of population structure to improve inferences about demographic history [8, 9].

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 spatial population structure. For over a decade, we have sampled fitness-related traits and SNPs in a population of the self-incompatible plant *Antirrhinum majus*. Our population-wide sampling permits a powerful test of whether the observed density and dispersal in this population can account for both the decay of pairwise relatedness with distance, and for the distribution of heterozygosity across individuals. First, we verify that there is excess variation in heterozygosity, which reflects an underlying variance in inbreeding. Second, to understand the role of spatial patterns of dispersal in generating variation in heterozygosity, we compare the empirical distribution of heterozygosity with that of offspring from simulated matings with nearby individuals at different dispersal scales. Third, we ask whether heterogeneous population density promotes variation in inbreeding, by comparing simulated pedigrees conditioned on uniform density versus observed locations of plants. Finally, we ask whether fitness proxies are correlated with heterozygosity, to set limits to the strength of inbreeding depression. Taken together, addressing these questions provides insight into how mating patterns and demography contribute to inbreeding and fitness outcomes in natural populations.

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. *Antirrhinum majus striatum* has yellow flowers and a smaller range, encircled by *A. m. pseudomajus*. The subspecies are parapatric; narrow hybrid zones form wherever they meet, and there is no evidence for post-zygotic reproductive barriers [10]. We focus on such a hybrid zone in the Vall de Ribès, Spain [11], where we have collected demographic data annually since 2009. Across nearly all of the genome, there is little divergence between the subspecies in this population, except for limited regions involved in floral pigmentation [12].

Genetic samples and trait measurements were obtained annually from 2009-2019 from every flowering individual that we could access in ~5 km stretches of two parallel roads that cross the Vall de Ribès, dubbed the "lower road" (GIV-4016; ~1150 m elevation) and "upper road" (N-260; ~1350 m). We also sampled along small side roads, railroad embankments, rivers, and hiking trails; density was very low away from these disturbed

areas. Some years we were limited to genotyping only in the core area, ~1 km along each road. The total genotyped sample 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). Eighteen percent of individuals were sampled in more than one year. 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 peak flowering (early June to mid-July). 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, determined by comparing samples that had been inadvertently recorded twice in the field (individuals with similar geographic location and near-identical genotypes, accounting for SNP errors).

SNP panel and isolation by distance

Previously, a panel of 248 SNPs spread throughout the genome was designed for the focal population (see methods in [13]). 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 total. (See Supplemental Material 1.1 (SM1.1) for SNP filtering methods). We imputed the ~5% missing genotypes for each SNP by randomly assigning genotypes according to the population-wide allele frequencies at each marker. The average F_{ST} over all years is 0.0244 (definition and calculation are detailed in SM1.1). The average pairwise F_{ST} is calculated each year for individuals separated by different distance classes and is averaged across years (details in SM1.1). Isolation by distance is shown by plotting pairwise F_{ST} against pairwise geographic distance (SM1.1: Fig. S1).

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. (Because of seed dormancy and perenniality, 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: $\mathrm{H} = \frac{1}{n}\sum_{i=1}^{n} h_i$, where h_i = 0 or 1, according to whether locus *i* is heterozygous, and there are *n* genotyped loci. Offspring from related individuals tend to have lower heterozygosity.

We observed an excess of individuals with around half the mean heterozygosity (see Results). To check whether the pattern was consistent with rare selfing, 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 $(1-F)^2\overline{2pq(1-2pq)}/n$. The second is due to covariance in heterozygosity between loci, which is termed the identity disequilibrium (ID). For a given pedigree, unlinked genes flow independently, thus 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 then 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 [2]). 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 [2]. 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 reassigning heterozygous status randomly across loci within individuals, which would eliminate correlations between loci generated by ID. We then computed g_2 using the g2_snps function from the R package InbreedR (in R version 3.6.1 [14]), which implements a modified formula for large data sets to estimate g_2, and provides confidence intervals via bootstrapping to account for the finite number of individuals sampled [15]. We decomposed ID into components due to linked and unlinked SNPs by comparing correlations of H for all individuals to those with low H, at several scales: across all pairs, within linkage groups, and between adjacent SNPs (SM1.2: Table S1).

<u>Effects of pollen dispersal on heterozygosity</u>

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. To test this prediction, we simulated offspring using all field individuals as mothers and choosing fathers from a given distance away (more detail in SM1.3). Then, we compared the distribution of H between the field data and offspring simulated from matings with three models of pollen dispersal: the nearest neighbour to the mother, a Gaussian distribution (σ = 300 m), and a leptokurtic dispersal kernel sampled from 1463 empirical measurements of pollen dispersal, estimated as the distances between assigned parents (electronic supplementary material; D. Field, unpublished data). The genotype of the offspring was assigned using Mendelian inheritance, either without linkage between

markers, or using the known linkage map (electronic supplementary material; courtesy of Yongbiao Xue, Beijing Institute of Genomics). 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 that expected for a spatially structured population, we simulated a continuous two-dimensional population, conditioned on the known locations of the individuals and the empirically measured seed and pollen dispersal distances (D. Field, unpublished data), using Mathematica 12.0 [16]. Our simulation differs from commonly used models (e.g., island [17], stepping stone [18] and continuous Wright-Malécot models [19, 20]) in that we include heterogeneity in density by specifying actual locations to determine relationships in the pedigree. 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 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.

First, however, we simulated a population with uniform density (the continuous Wright-Malécot model) as a null model, in order to compare expected heterozygosity with and without heterogeneous spatial structure. We simulate a region of ~1.1 x 1.8 km that was sampled consistently in the *A. majus* focal population (SM1.4: Fig. S4-S5). Locations were assigned by randomly sampling N points from a uniform distribution each generation, for 1000 generations. Genetic diversity is shaped over the coalescent timescale ($2N_e$, ~170,000 generations in *A. majus* [12]), 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. The spatial pedigree was generated by choosing parents for each individual according to a backwards dispersal distribution measured empirically (see methods for choosing parents in SM1.4). Once the spatial pedigree is generated, 10 replicate sets of genotypes are assigned by dropping genes down the pedigree, starting with equal expected frequencies of both alleles at each of 91 loci. In fact, one could start with any initial frequencies, since F_{ST}-like measures are independent of them. Population size was adjusted so that F_{ST} matched the empirical data for the simulated sampling area.

Next, we simulated a population with realistic heterogeneous spatial structure by using the individual locations available for the years 2009 to 2019 in the *A. majus* focal population. There were fewer individuals from 2017-2018, so these were merged, giving

distribution data for 10 time points. We randomly sample from the ten consecutive time points, and repeat for 100 cycles, thus iterating for 1000 generations. We subsample from these locations, so as to maintain a constant population size (N). If N is greater than the number of plants available in a given time point, say k, all k plants are first included and the remaining N-k locations were resampled from the same time point, displaced at a random angle on a circle of radius 3m to avoid having plants in the same location. This naïve approach allows us to simulate a spatial structure that is realistic over at least small scales. We then generated a pedigree following the procedure used for the uniform population, again adjusting population size to match the empirically observed F_{ST}. Ten replicate sets of genotypes were run for each of five replicate pedigrees.

Patterns of isolation by distance, heterozygote deficit (F_{IS}) and identity disequilibrium were compared between the two simulation types and the field data (calculated from the simulated sub-area of the field site). As the fitted population sizes were large (see Results), obtaining direct estimates of identity by descent and thus F_{ST} from the pedigrees was not feasible. Instead, F_{ST} was obtained for a pedigree as the average of replicate genotype sets generated from that pedigree. F_{IS} was calculated from the observed and expected heterozygosity. Values of g_2 were calculated for each replicate from each pedigree using InbreedR.

Heterozygosity-fitness correlations

We recorded traits related to size and reproduction from every individual in select years, and assessed their utility as fitness proxies by correlating against more complete measures of fecundity measured on a subset. The traits measured on all individuals were: (i) number of flowering stems (2009-2019), (ii) number of open flowers (2010-2019), (iii) height of the tallest stem (cm; 2014-2019), and (iv) 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 during the flowering season. To address the potential added noise from this constraint, we included measurement date as a covariate in the HFC analyses. For perennial plants, which have trait measurements in multiple years (18% of individuals), the year with the highest measured value for each trait was included in the pooled year analyses.

We assessed whether traits measured on every individual can be used as fitness proxies by asking whether they are related to 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. 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 fitness was correlated with heterozygosity, we used Generalized Linear Models (GLMs) in R, using the MASS package for negative binomial models [21].

Response variables were fitness proxies: total fruit, total flowers, estimated fruits, 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 [22], calculated with InbreedR [15]. Each model also contained fixed-effect covariates of year, road (upper vs. lower, designated by 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, leaf nodes, and flowering stems. The family and link function for the GLM was assigned according to the distribution of each trait: gamma with a log link for stem height, and negative binomial (to account for overdispersion) for all other traits.

The residuals of the models indicate spatial autocorrelation (significantly positive Moran's *I*), suggesting that the 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 in R [21, 23, 24].

Correlations between heterozygosity and fitness are due to the product of the correlation of inbreeding with both heterozygosity and fitness (Eq. 2 in [2]). Using this relationship, the results of the HFC analysis can be used to calculate the inbreeding load, $\beta_{f,W}$, for each trait [2]. Although most traits showed non-significant relationships, the potential magnitude of inbreeding load was calculated for each trait (formulas in SM1.5).

Results

<u>Variation in inbreeding</u>

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 increased 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 heterozygous status is shuffled across loci by random reassignment within individuals (Fig. 1, grey). Shuffling removes correlations between loci and thus eliminates identity disequilibrium (ID). The mean heterozygosity (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 between the observed and shuffled data implies that the mean standardised ID is g_2 = 0.0029 (95% CI: 0.0026-0.0033), representing a significant variance in inbreeding between individuals.

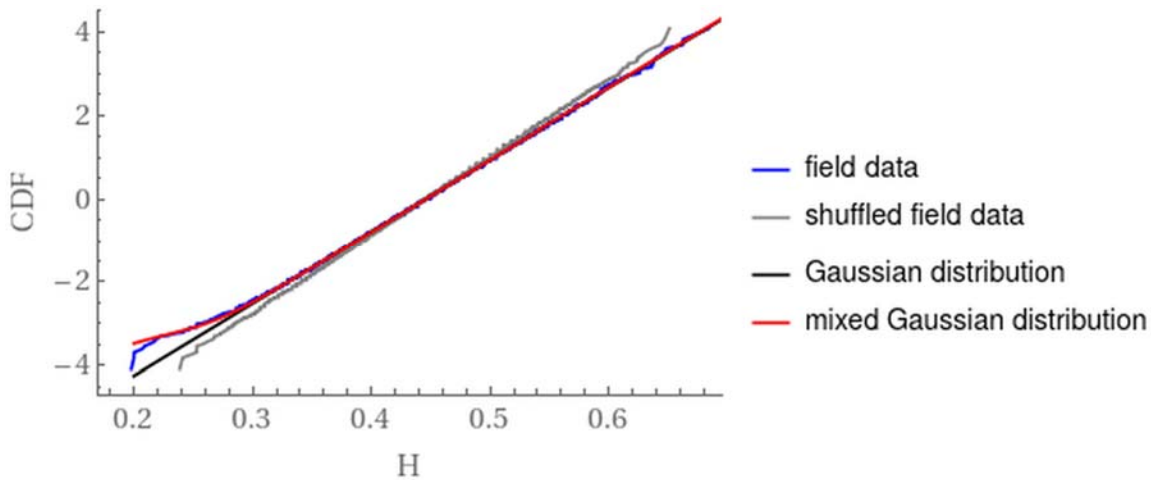

Figure 1. Probit transform of the cumulative distribution function (CDF) of the distribution of individual heterozygosity (H). The y-axis is the number of standard deviations of the standard normal distribution.

<u>Effects of pollen dispersal on heterozygosity</u>

Figure S2 shows that the heterozygosity of simulated offspring depends on distance between parents (SM1.3). Next, we checked whether real offspring 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, SM1.3: Tables S2-S3). On the other hand, all three dispersal schemes differ in the tail as assessed by Kolmogorov-Smirnov tests (SM1.3: 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 from matings between close relatives. We focused 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. In comparison, only 4/100 replicates gave

higher estimates than that observed (SM1.3: 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. Here, 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 – nominally this represents a significant difference, as the 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.

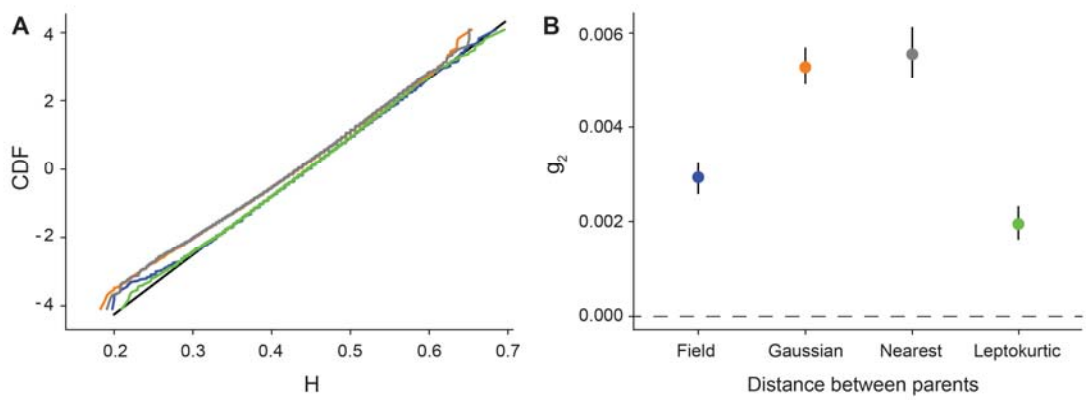

Figure 2. A: Probit transform of the CDF of multilocus heterozygosity, H, for the field data (blue) versus a single replicate of offspring simulated from Gaussian pollen dispersal (orange), nearest neighbor matings (grey), and leptokurtic pollen dispersal (green). A normal distribution (black) with the same mean and standard deviation as the field data is included for comparison **B:** Identity disequilibrium for the same data as above indicating mean and 95% CI.

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

To examine if the observed g_2 is consistent with a spatially structured model, we simulated a pedigree over 1000 generations, conditioned on the locations of individuals observed over ten years, repeated over 100 cycles. The realized seed and pollen dispersal matched the empirical seed and pollen dispersal distribution for both density types (SM1.4: Fig. S6, Table S4). We required N = 15500 individuals for the heterogeneous density model and 40000 individuals for the uniform density, in order to match the observed F_{ST}~0.022 calculated from the simulated sub-area of the field site (SM1.4: Table S5). Up to distances of 1km, the decline in pairwise identity with distance matched between the field data and the five replicate pedigrees simulated with heterogeneous

density (SM1.4: Fig. S8A, B). High variation among genotypes suggests that many more SNPs would be needed to match the pattern from the pedigree (SM1.4: Fig. S8C); moreover, linkage would increase this variance to some extent. We also compared the pattern of isolation by distance from the field data to that from the pedigrees generated for both the heterogeneous and uniform density scenarios (SM1.4: Fig. S9); we found that the heterogeneous density is a much better fit than the uniform density (Table S5).

Identity disequilibrium (g_2) estimates from the genotypes from simulated pedigrees with heterogeneous density showed substantial variation between the five simulated pedigrees, and between the ten draws of 91 SNPs from each pedigree (Fig. 3). The average g_2 estimated from the five pedigrees (each with 10 replicates) is 0.00264, which is consistent with the observed mean annual g_2 from the field of 0.00262. On the other hand, when assuming a uniform density, the average g_2 of 0.00171 is significantly lower than the field data. Note that the confidence limits for the field data, generated by InbreedR, only include error due to sampling a limited number of individuals. These errors do not account for sampling a limited number of SNPs, or the random variation between evolutionary realisations (see Discussion).

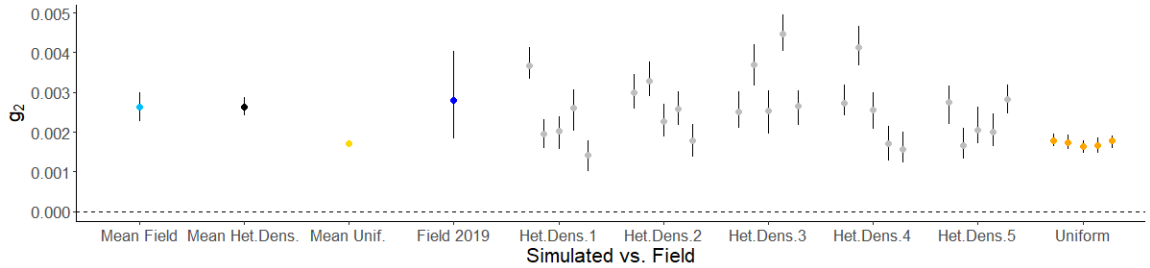

Figure 3. Identity disequilibrium (g_2) calculated from field data versus simulated pedigrees. Five of ten replicates per pedigree are shown (grey: heterogeneous density, with five simulated pedigrees; orange: uniform density, with one simulated pedigree). Mean from the field (light blue) is across 2009-2019, while mean from the heterogeneous (black) and uniform (yellow) simulations is across all replicates. The final year of field data (dark blue) is comparable to g_2 calculated from the final year of pedigree replicates (grey and orange).

<u>Heterozygosity-fitness correlations</u>

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; SM1.5: Fig. S10), justifying use of these traits as fitness proxies. The highest correlation with total fruit was with the number of flowering stems (0.65), which was even higher than with total flowers (0.41).

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, SM1.5: Table S6). This result was still significant after accounting for spatial autocorrelation (data not shown), although it should be noted that the model assumptions for normality of residuals was not met. In

contrast to heterozygosity, the covariates typically were highly significant predictors of fitness proxies (Table S6; further discussed in SM1.5).

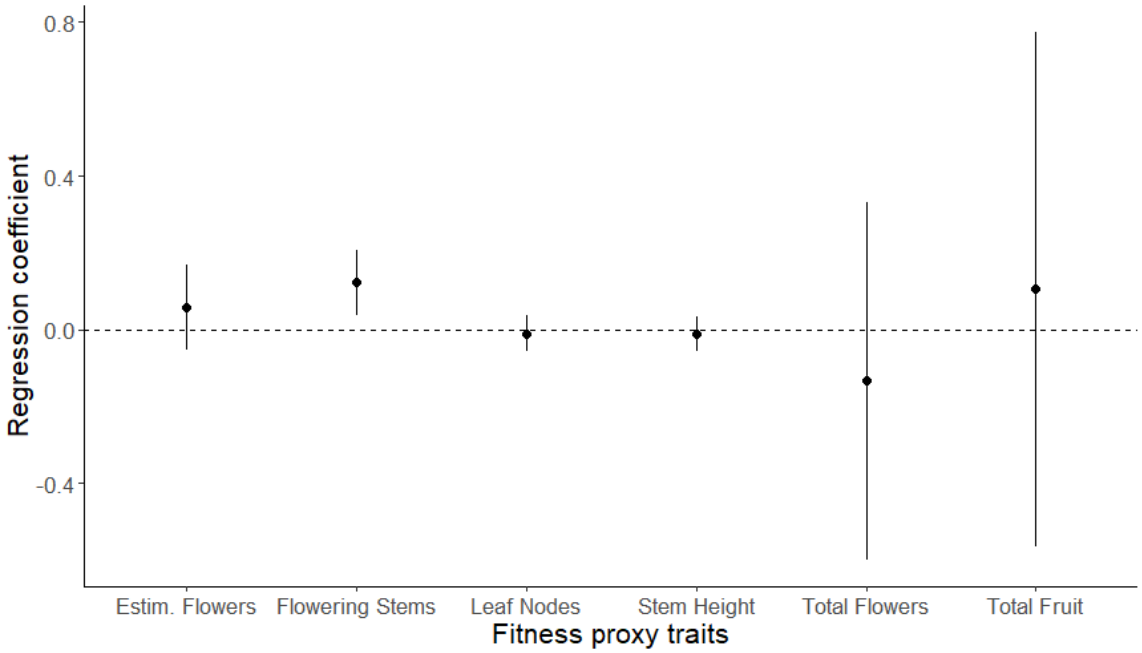

Figure 4. Regression coefficients with 95% confidence intervals for H^* as a predictor of various fitness proxies. Note that coefficients show the percentage change in each trait as H* increases, since the analysis used a log link function.

The regression coefficient of inbreeding on heterozygosity was $\beta_{f,H}$ = -0.173. The maximum magnitude of the inbreeding load, $\beta_{f,W}$, for each trait is shown in SM1.5: Table S7 – this represents 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% fewer flowering stems for each unit increase in inbreeding (mean $\beta_{f,W}$ = -0.70, upper CI = -0.20, lower CI = -1.20; s.d. of flowering stems = 3.47).

Discussion

The negative fitness consequences of inbreeding are well known; what is less well understood is how demographic processes, such as heterogeneous density and the dispersal distribution, influence the distribution of heterozygosity in a population. In this study of over 20,000 plants sampled over 11 years in a natural population, we address two questions central to understanding how demography can influence patterns of heterozygosity and fitness in the field: (i) can the distribution of heterozygosity be predicted for an outcrossing species using key demographic parameters?, and (ii) are there negative effects of inbreeding on fitness proxies? We first confirmed that there is significant correlation in heterozygosity between markers (g_2, a measure of identity disequilibrium). This implies that there is variation in inbreeding, providing the opportunity for inbreeding depression to cause heterozygosity-fitness correlations (HFCs). By simulating offspring from matings between georeferenced, genotyped

individuals, 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, SM1.3: Fig. S2). We found that the observed distribution of heterozygosity is consistent with the known leptokurtic distribution of pollen dispersal. We also simulate the population over 1000 generations using the actual seed and pollen dispersal kernels, and the observed heterogeneous density. We found that this simple model matches the observed identity disequilibrium, whereas a null model with uniform density underestimates it. Despite significant variation in inbreeding, only one of six fitness proxies (number of flowering stems) increased significantly with heterozygosity. This one significant value ($\beta_{f,W}$ = -0.70) indicates quite strong inbreeding depression, but we have low power to detect inbreeding depression in the other traits (SM1.5: Table S7). Taken together, our findings highlight the potential for using demographic parameters to explain variance in inbreeding in spatially continuous populations.

Many studies that test for HFCs find that the excess variation in heterozygosity, g_2, which arises from identity disequilibrium, is low and rarely significant [3]. In our study, we estimate a significant g_2 of 0.0029 (95% CI: 0.0026-0.0033). Although low, this estimate is of the same order as most of the g_2 values found across 105 vertebrate populations in a meta-analysis of 50 HFC studies (average of 0.007) [3], and on the same order as ~60% of the local populations surveyed in a long-lived tree [25]. In our study, beyond simply estimating identity disequilibrium, we use two types of simulation to explore potential demographic factors shaping variation in inbreeding. The first simulation shows that the spatial pattern of pollen dispersal affects the distribution of heterozygosity within a population. Simulated matings with the empirically measured leptokurtic pollen dispersal curve were consistent with the actual g_2, compared to matings with nearest neighbors or a Gaussian pollen dispersal. This result is somewhat surprising because we did not include the complexities of the mating system of *A. majus*. *Antirrhinum majus* has a gametophytic self-incompatibility system (GSI [26]), 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 not only prevents selfing, but also reduces mating among relatives (i.e., biparental inbreeding) because related plants are more likely to share S-alleles [27, 28]. 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. Indeed, we found that a small amount of biparental inbreeding with leptokurtic pollen dispersal may explain an excess of individuals with low heterozygosity, around half the mean (Fig. 2, SM1.3: Fig. S3), although we have low statistical power to distinguish this from rare selfing.

Our second simulation approach asked whether heterogeneous density promotes variation in inbreeding, given strong fine-scale population structure (i.e., pairwise F_{ST} decays rapidly over a few metres, SM1.1: Fig. S1). Indeed, simulated pedigrees with uniformly distributed plants underestimated identity disequilibrium compared to the empirical value. In contrast, simulated pedigrees conditioned on the actual, heterogeneous density of plants were consistent with identity disequilibrium measured in the field, indicating that patchiness combined with leptokurtic dispersal may contribute to observed patterns

of heterozygosity. Simulations with heterogeneous density also better capture empirical isolation by distance patterns than those with a uniform density (SM1.4: Fig. S9). However, the effective population size of 15500 individuals in the heterogeneous-density simulations is an order of magnitude larger than the average number of plants observed in a year (~2500). We estimate that the majority of plants in the population are sampled each year, so the discrepancy is likely not due to significant under-sampling but rather to a seed bank, which is expected to substantially increase the effective population size [30]. Despite simplifications such as non-overlapping generations, no seed bank, and a simple SI system, the heterogeneous-density simulation accurately captures patterns of identity disequilibrium and isolation by distance.

Our estimation of identity disequilibrium illustrates a general problem with statistical comparisons in evolutionary biology. There are three sources of error in estimating g_2: firstly, from sampling a limited number of individuals, secondly, from sampling a limited number of SNP, and thirdly 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 [15]. The second source of error (a limited number of SNPs) is shown by the substantial variation in g_2 of the ten replicates of each of five pedigrees, with variation generated by random meiosis amongst unlinked markers on a fixed pedigree. This variation could be reduced by increasing the number of SNPs, but the effective number of segregating sites that can be included in the analysis is fundamentally limited by the length of the genetic map. Finally, there is additional variation between pedigrees, due to the random assignment of parents in the simulations, which generates a random pedigree. The wide variation in estimates of g_2 due to random meiosis, and to the random generation of the pedigree (Fig. 3) is an important reminder that estimates of parameters are typically limited by the randomness of evolution, and can be far higher than that due to the limited number of individuals or SNPs sampled.

Inbreeding depression is expected to be strong, especially in self-incompatible (SI) species such as *Antirrhinum majus* [31]; this prediction is borne out for many SI species when inbreeding depression is measured through experimental crosses [7]. We found strong inbreeding depression for the number of flowering stems, but HFCs were not significant for the five other fitness proxies. Our g_2 value corresponds to a coefficient of variation of non-identity (1-F) of only $\sqrt{0.0029}$ =0.054. Whilst this identity disequilibrium is significant, it is still too low to allow detection of weak HFCs, despite our exceptionally large sample size. Moreover, although the fitness proxies available in our complete dataset are correlated with total fruit set measured on a subset of plants, these relationships are noisy. Estimates of total lifetime fitness are not possible for such a large sample, but future work will improve over using the largest measurement for perennials by integrating fitness proxies across years and comparing to fitness estimates from a pedigree. Environmental and stochastic effects on traits related to size and fecundity (e.g., the strong covariate effects in SM1.5: Table S6) may swamp fitness effects due to heterozygosity. A further potential explanation for the lack of HFCs is the "invisible fraction," the many individuals that die before their fitness can be evaluated [32]. 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 [33]. In a previous study of paternity in *A. majus*, seeds were randomly collected from our field site, germinated in the greenhouse, and genotyped at 70 SNPs [34]. The heterozygosity of 1127 seedlings from 59 mothers shows the same distribution as the flowering adults sampled in our dataset (T. Ellis, unpublished data; data not shown), suggesting that our sampling method is not missing a significant fraction of seedlings with low heterozygosity that fail to reach flowering. However, our methods still would not detect early-acting inbreeding depression that would cause seed abortion or germination failure.

The distribution of heterozygosity has often been measured to estimate inbreeding depression via HFC. Yet, this type of data has been rarely used to investigate population structure in its own right, as a complement to the more widely used pairwise identity, F_{ST}. By bringing together local inbreeding and isolation-by-distance, our study provides a novel assessment of how dispersal and population density can explain both pairwise identity and the distribution of heterozygosity in spatially continuous populations. However, we have only begun to investigate how the distribution of heterozygosity can be shaped by population structure and demographic parameters. Our future work will focus on understanding how other features such as a seed bank influence genetic diversity, with the ultimate goal of deriving information about demographic history from the distribution of heterozygosity in populations that have fewer measured parameters.

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. We thank Tom Ellis for sharing recently updated data with us from his dissertation research. We also thank three anonymous reviewers for their feedback, which substantially improved the manuscript.

References

1. Barrett SCH. Darwin's legacy: The forms, function and sexual diversity of flowers. Philos T R Soc B. 2010;365:351-368.
2. Szulkin M, Bierne N, David P. Heterozygosity-fitness correlations: A time for reappraisal. Evolution. 2010;64(5):1202-17.
3. Miller JM, Coltman DW. Assessment of identity disequilibrium and its relation to empirical heterozygosity fitness correlations: A meta-analysis. Mol Ecol. 2014;23(8):1899-909.
4. Zhao L, Charlesworth B. Resolving the conflict between associative overdominance and background selection. Genetics. 2016;203(3):1315-1334.
5. Charlesworth D, Willis JH. The genetics of inbreeding depression. Nat Rev Genet. 2009;10(11):783-96.
6. Loveless M, Hamrick J. Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Evol S. 1984;15:65-95.

7. 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.
8. Milligan BG, Archer FI, Ferchaud AL, Hand BK, Kierepka EM, Waples RS. Disentangling genetic structure for genetic monitoring of complex populations. Evol Appl. 2018;11(7):1149-61.
9. Bradburd GS, Ralph PL. Spatial population genetics: It's about time. Annu Rev Ecol Evol S. 2019;50:427-49.
10. 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 Syst Evol. 2010;286(3-4):223-34.
11. 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.
12. 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. P Natl Acad Sci. 2018;115(43):11006-11.
13. 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.
14. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
15. 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 Ecol Evol. 2016;7(11):1331-9.
16. Wolfram Research Inc. Mathematica. Champaign, Illinois: Wolfram Research, Inc.; 2019.
17. Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97-159.
18. 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.
19. Wright S. Isolation by distance. Genetics. 1943;28(2):114-38.
20. Malécot G. The Mathematics of Heredity (English translation, 1969). San Francisco: WF Freeman; 1948.
21. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. New York: Springer; 2002.
22. 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.
23. 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.
24. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. nlme: Linear and nonlinear mixed effects models. R package version 3.1-131. 2017. https://CRAN.R-project.org/package=nlme2017.
25. Rodríguez-Quilón I, Santos-del-Blanco L, Grivet D, Jaramillo-Correa JP, Majada J, Vendramin GG, Alía R, González Martínez SC. Local effects drive heterozygosity-

fitness correlations in an outcrossing long-lived tree. Proc R Soc B. 2015;282:20152230

26. 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.
27. Cartwright RA. Antagonism between local dispersal and self-incompatibility systems in a continuous plant population. Mol Ecol. 2009;18(11):2327-36.
28. Charlesworth D, Charlesworth B. Inbreeding depression and its evolutionary consequences. Annu Rev Ecol S. 1987;18:237-68.
29. 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.
30. Heinrich L, Mueller J, Tellier A, Živković D. Effects of population and seed bank size fluctuations on neutral evolution and efficacy of natural selection. Theor Popul Biol. 2018;123:45-69.
31. Lande R, Schemske DW. The evolution of self-fertilization and inbreeding depression in plants. 1. Genetic models. Evolution. 1985;39(1):24-40.
32. 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.
33. 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.
34. Ellis T. The role of pollinator-mediated selection in the maintenance of a flower color polymorphism in an *Antirrhinum majus* hybrid zone [dissertation]. Klosterneuburg, Austria: IST Austria; 2016.