

1 **Supplemental Material 1 (SM1)**

2 for "Effects of fine-scale population structure on the distribution of heterozygosity in a
3 long-term study of *Antirrhinum majus*" by Parvathy Surendranadh, Louise Arathoon,
4 Carina A. Baskett, David L. Field, Melinda Pickup, Nicholas H. Barton

5
6 **Heading order follows the Methods and Results of the main**
7 **text.**

8 9 **SM1.1 SNP panel**

10 11 ***Detailed Methods: SNP Panel***

12 For each individual, DNA was extracted from leaf material collected from the field site,
13 and was genotyped for the SNP panel by LGC Genomics (Middlesex, UK) using the
14 KASP genotyping platform. Due to repeated sampling of the same individuals across
15 years, the error rate of this method could be calculated, and was found to be low (mean
16 error rate < 0.1% per locus).

17
18 Candidate loci were identified using a draft *A. majus* reference genome (~ 630 Mb across
19 eight linkage groups; courtesy of Yongbiao Xue, Beijing Institute of Genomics); see ref
20 [1]. In this study, SNPs were chosen to have overall mean frequency between 0.1 and 0.9;
21 90% had frequency between 0.25 and 0.75. SNPs that showed excessive geographic
22 differentiation were eliminated by requiring a linear regression gradient of allele
23 frequency on east-west distance to be less than 0.09 km^{-1} ; 90% of chosen SNPs had a
24 gradient < 0.03 km^{-1} . Furthermore, we required that $F_{ST} < 0.1$; F_{ST} was calculated by
25 dividing the region into 200m squares, yielding 164 non-empty demes. Finally, the
26 overall heterozygote deficit, F_{IS} , was required to be between -0.1 and 0.2; 90% of chosen
27 SNPs had $-0.04 < F_{IS} < 0.1$. SNPs with heterozygote deficit $F_{IS} > 0.2$ also showed high
28 F_{ST} and/or clinal gradient, whilst those with $F_{IS} < -0.1$ were likely due to genotyping
29 artefacts (e.g., primers binding to more than one site in the genome). After applying these
30 filters, 170 SNPs remained. Finally, we chose to work with the 91 SNPs that were
31 assayed for at least 60% of the Planoles sample (i.e., at least 13,411 individuals).

32 33 34 **SM1.2 Variation in inbreeding**

35 36 ***Detailed Methods and Results***

37 The identity disequilibrium that we find is due partly to associations between linked SNP,
38 and partly to associations between unlinked SNP (73% vs. 27%, respectively). Table S1
39 shows that correlations in h between SNP within linkage groups are consistently positive,
40 averaging Pearson's $r = 0.01126$ - much higher than the average correlation of 0.00240
41 between all pairs of SNP, which are mostly unlinked. For the 155 individuals with $H < 0.3$,
42 the mean correlation within linkage groups, 0.0354, is much higher, reflecting the shared
43 inheritance of large blocks of genome for close relatives. Correlations are higher between
44 adjacent SNP, and yet higher in highly inbred individuals.

45 **Table S1.** Correlations in heterozygosity within the 8 linkage groups (LG). The third and
 46 fourth columns give the mean correlation in H between loci within each linkage group,
 47 for all 22,353 individuals versus for the 155 individuals with H<0.3. The next two
 48 columns give the mean correlations between adjacent SNPs. The last two columns give g_2
 49 values within each linkage group. Note that 1 of the 91 SNP was not assigned to a linkage
 50 group.
 51

LG	No. SNPs	H within LG		adjacent SNP		g_2 within LG	
		all inds	H<0.3	all inds	H<0.3	all inds	H<0.3
1	13	0.01828	0.04423	0.08998	0.15209	0.02266	0.10093
2	15	0.02929	0.05021	0.12139	0.13723	0.03229	0.13187
3	10	0.00459	0.03298	0.00958	0.06359	0.01012	0.07800
4	12	0.00292	-0.00370	0.00294	-0.00530	0.00327	0.00512
5	12	0.02407	0.07854	0.02654	0.10365	0.02834	0.20210
6	15	0.00402	0.03128	0.00982	0.06743	0.00522	0.08369
7	6	0.00388	0.03862	-0.00159	0.03416	0.00571	0.11591
8	7	0.00304	0.01089	0.00334	0.00805	0.00297	0.06129
Mean	90	0.01126	0.03538	0.03275	0.07011	0.01382	0.09736

52

53 **SM1.3 Effects of pollen dispersal on heterozygosity**

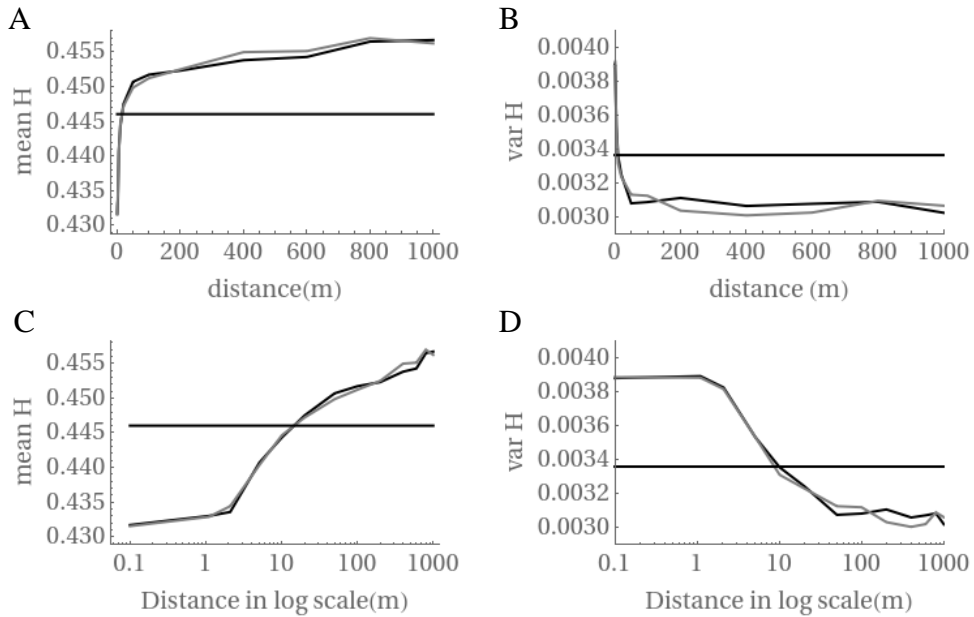
54

55 ***Detailed Methods and Results***

56 The distribution of heterozygosity of offspring depends on distance between parents. We
 57 show this by simulating offspring, using all field-sampled individuals as mothers
 58 (Mathematica notebook in electronic supplementary material [22]). We chose fathers
 59 close to a given distance away, by choosing 12 points evenly spaced on a circle, and
 60 taking the nearest individual to any of those points. The genotype of the offspring was
 61 determined by Mendelian inheritance based on parental genotypes. The mean
 62 heterozygosity of offspring from two parents is linearly related to their pairwise identity;
 63 thus, the increase in mean identity with distance (Fig. S1A,C) is a precise reflection of the
 64 decay in pairwise relatedness. The variance in heterozygosity decreases with distance, as
 65 individuals become less related (Fig. S1B,D). Both mean and variance of H change
 66 sharply over scales of a few metres, and are hardly affected by linkage (compare gray and
 67 black lines in Fig. S1). The observed values from the field data (horizontal lines in Fig.
 68 S1) are consistent with pollination from fathers ~10m away, but are of course the product
 69 of a broad distribution of distances.

70

71



72
73
74
75
76
77

Figure S1. Mean (A, C) and variance (B,D) of H as a function of the distance between parents. Offspring are generated with no linkage (black) or with linkage (gray); observed values from the field data are shown as a horizontal line. Plots C and D show distance in a log scale.

78
79
80
81
82

Table S2. Mean and variance of multilocus heterozygosity (H) and identity disequilibrium (g_2) from field data and offspring simulated from three possible patterns of pollen dispersal (a leptokurtic kernel, a Gaussian kernel, and pollen from the nearest neighbour).

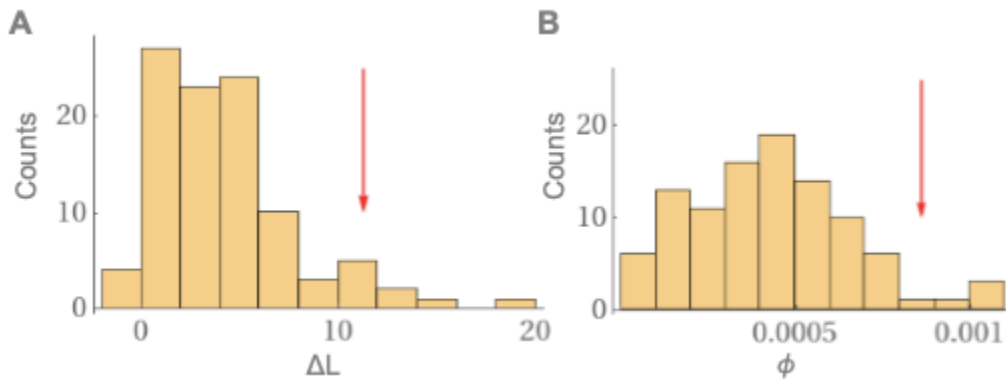
	H mean	H variance	g_2	g_2 CI
Field data	0.4460	0.0034	0.0029	0.0026 - 0.0033
Leptokurtic offspring	0.4458	0.0034	0.0020	0.0016 - 0.0024
Gaussian offspring	0.4323	0.0039	0.0053	0.0049 - 0.0057
Neighbour offspring	0.4314	0.0039	0.0056	0.0051 - 0.0060

83

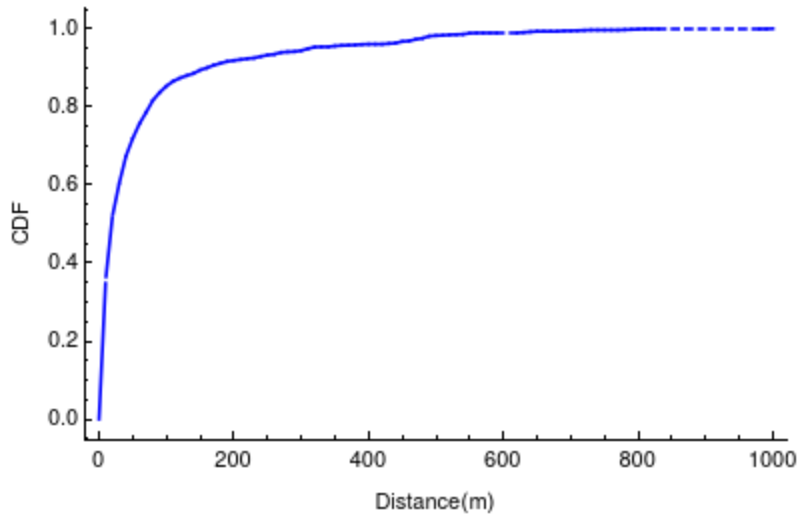
84 **Table S3.** Test statistic and p -value from t -test, F -test and Kolmogorov-Smirnov (KS)
 85 test for each pairwise comparison between heterozygosity calculated from field data and
 86 offspring simulated from leptokurtic, Gaussian, and nearest neighbour matings.
 87

Dataset	Dataset	t -test		F test		KS test	
		t	p	F	p	D	p
Field	Leptokurtic	1.08	0.281	-0.55	0.579	0.015	0.015
Field	Gaussian	24.06	<0.001	-10.02	<0.001	0.094	<0.00001
Field	Neighbour	25.17	<0.001	-9.78	<0.001	0.103	<0.00001
Leptokurtic	Gaussian	23.07	<0.001	-9.80	<0.001	0.086	<0.00001
Leptokurtic	Neighbour	24.19	<0.001	-9.36	<0.001	0.090	<0.00001
Gaussian	Neighbour	1.06	0.29	0.31	0.75	0.009	0.36

88
89



90
 91 **Figure S2:** The distribution of increase in the likelihood (ΔL) (A) and selfing rate, ϕ (B)
 92 between the single and mixed Gaussian distributions from 100 replicates of simulated
 93 matings from leptokurtic dispersal distribution. The red arrow points to the value
 94 observed in the field data.
 95

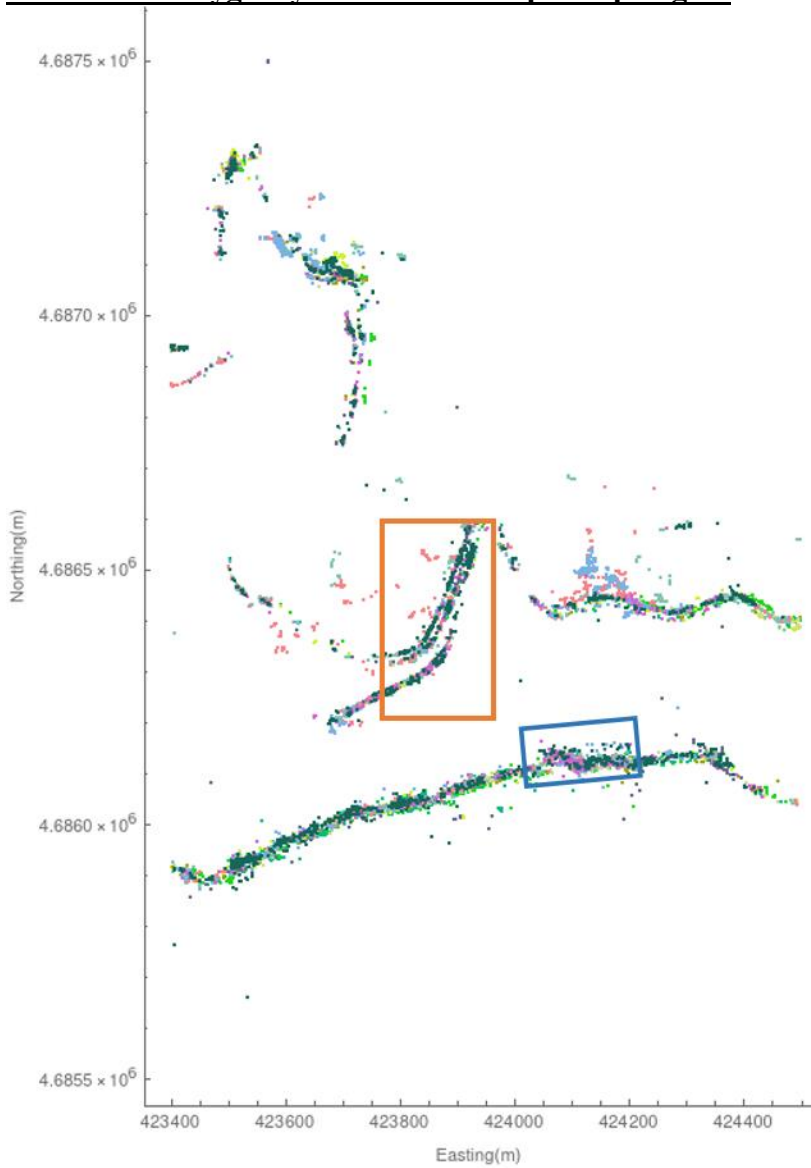


96

97 **Figure S3:** CDF of the empirically measured pollen dispersal distribution. Here, 50% of

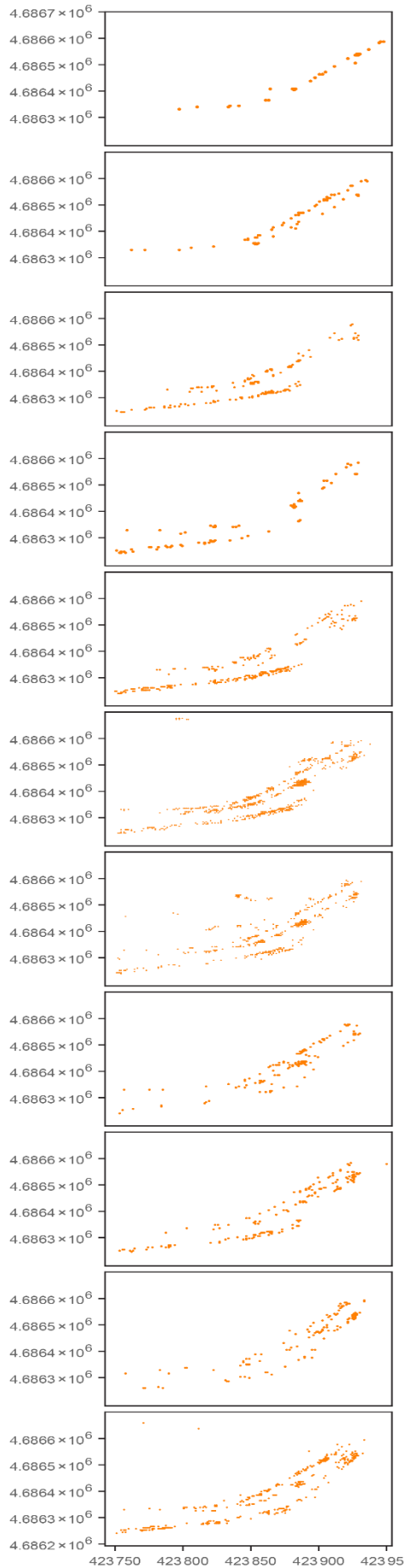
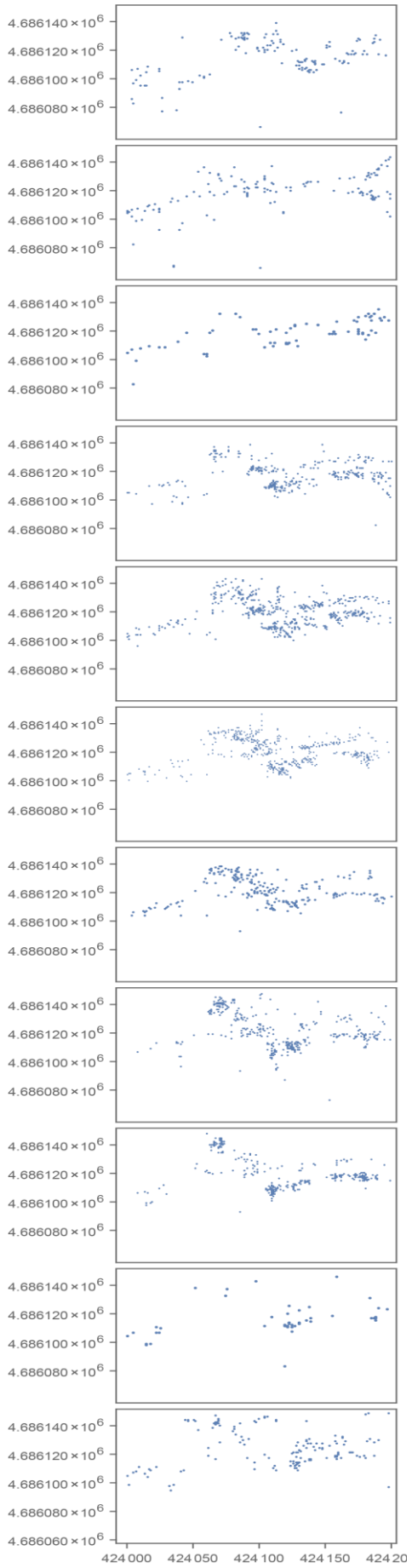
98 matings occur within 20m and 75% of matings occur within 60m.

99 **SM1.4 Heterozygosity in a simulated spatial pedigree**



100
101
102
103

Figure S4. Individual plant locations in the simulated region of the field site. Each colour represents a different year. See Fig. S5 for time series of boxed areas.



105 **Figure S5.** Close-ups of sections of the lower road (left) and upper road (right) from the
 106 field data, showing changes in patchiness over time from 2009 (top) to 2019 (bottom).
 107 Sections are denoted as blue and orange in Fig. S4.

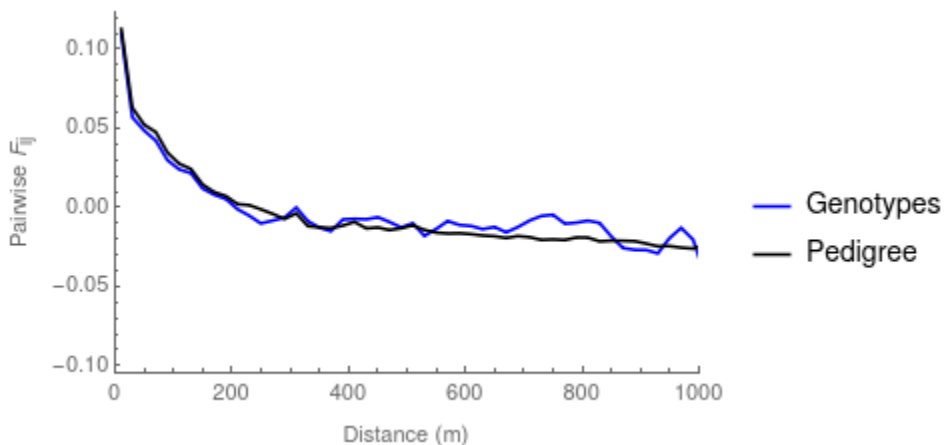
108

109 ***Detailed Methods and Validations for Simulated Spatial Pedigree***

110

111 Since there is no analytical result for probability of identities for a heterogeneously
 112 distributed population with leptokurtic dispersal, we validated the simulation by
 113 comparing pairwise relatedness calculated from the genotypes using 10 replicate
 114 genotypes (using the method described in ‘Heterozygosity in a simulated spatial
 115 pedigree’) against that directly calculated from the simulated pedigree (F). F can be
 116 considered as a $N \times N$ matrix with each element F_{ij} (corresponding to row i and column j)
 117 giving the probability of identity between individuals i and j , where N is the population
 118 size. If we start with a population of unrelated individuals, the probability of identity
 119 matrix at generation 0, F_0 , would contain only 0's. The probability of identity of two
 120 distinct genes from a pair of distinct individuals i and j in generation $g+1$ is $F_{ij,g+1} =$
 121 $\sum_{k,l} M_{ik} F_{kl,g}^* M_{lj}$, where $M_{xy} = 1/2$ if y is a parent of x (with no selfing) and 0 otherwise,
 122 and $F_{kl,g}^* = F_{kl,g}$ if $k \neq l$ and $F_{kl,g}^* = \frac{1}{2} + \frac{1}{2} F_{kk,g}$ if $k = l$. $F_{kl,g}^*$ denotes the probability
 123 of identity of individuals k and l in the previous generation g [2]. F_{ij} as a function of
 124 distance was found from the genotypes and pedigree. Note that we use a smaller $N=1000$
 125 for this calculation due to computational constraints of calculating F from the pedigree.
 126 We see that isolation by distance from the pedigree matches the average from 10 replicate
 127 genotypes (Fig. S6). Furthermore, we verified our algorithm by comparing the proposed
 128 and realized seed and pollen dispersal distributions from the pedigree (Fig. S7). Together,
 129 these two checks validate the estimation of F from the simulated pedigree, and the
 130 algorithm for choosing parents.

131

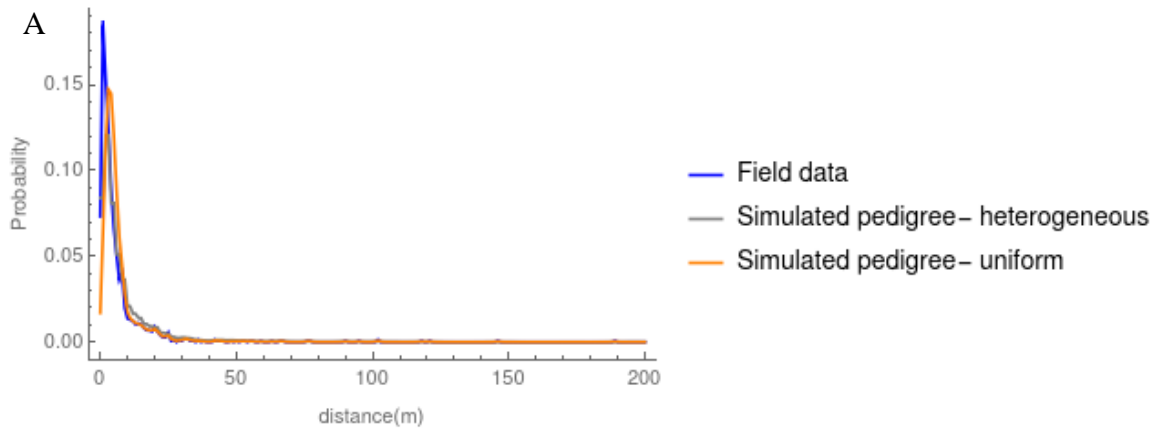


132

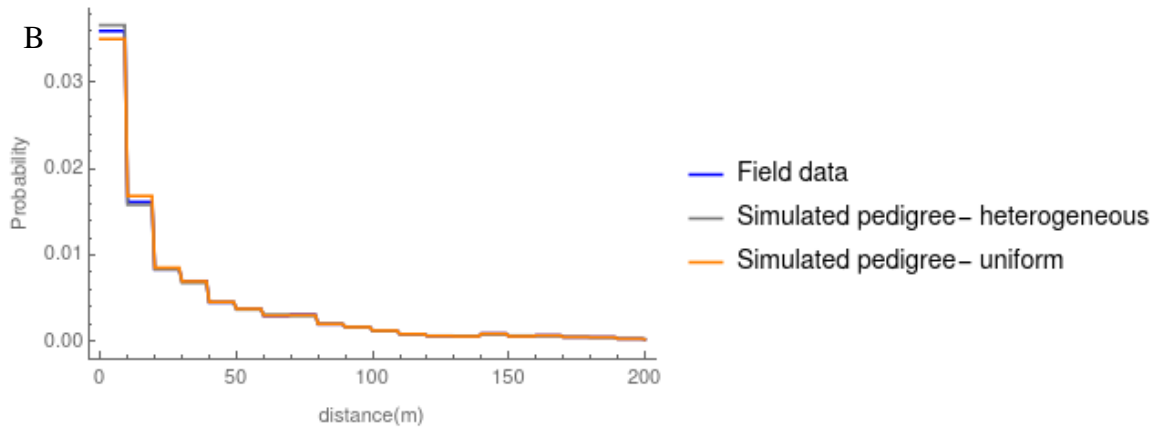
133 **Figure S6:** Isolation by distance calculated for a simulated heterogeneous population of
 134 1000 individuals calculated directly from the pedigree (black) and the average from 10
 135 replicate genotypes (in blue).

136

137



138



139

140 **Figure S7.** Realized (gray and orange) and proposed (blue) seed (A) and pollen (B)
 141 dispersal distribution for the simulated pedigrees with heterogeneous and uniform
 142 population structure. Due to computational constraints, these are calculated from the last
 143 300 generations for the simulated pedigree with uniform density. The pedigree with F_{ST}
 144 closest to that of the field data is shown for the heterogeneous case (also in Fig. S8B, S9).

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

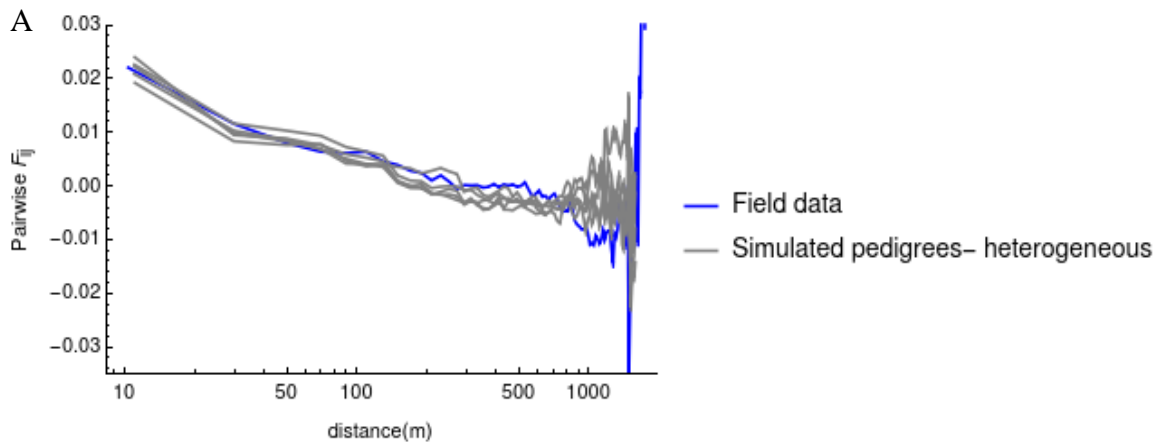
160

161 **Table S4.** Mean and standard deviation (SD) of the proposed and realized seed and
 162 pollen dispersal distributions for the simulated pedigrees with uniform and heterogeneous
 163 spatial structure.
 164

	Seed dispersal		Pollen dispersal	
	Mean	SD	Mean	SD
Proposed	9.52468	38.2242	62.684	119.327
Heterogeneous pedigree 1	12.797	40.1884	62.5955	118.765
Heterogeneous pedigree 2	12.7785	40.0344	62.6493	118.834
Heterogeneous pedigree 3	12.7724	40.0098	62.5716	118.682
Heterogeneous pedigree 4	12.7863	40.0749	62.6127	118.778
Heterogeneous pedigree 5	12.7858	40.0773	62.5902	118.774
Heterogeneous- average	12.784	40.0770	62.6039	118.767
Uniform pedigree	10.4706	38.0741	63.1388	119.122

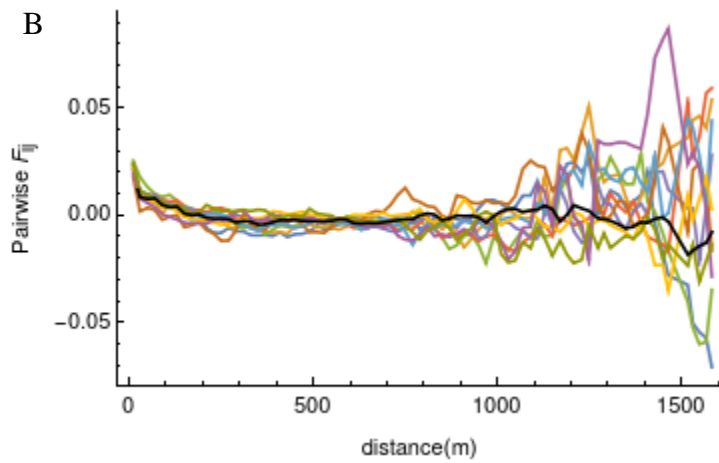
165

166



167

168



169

170

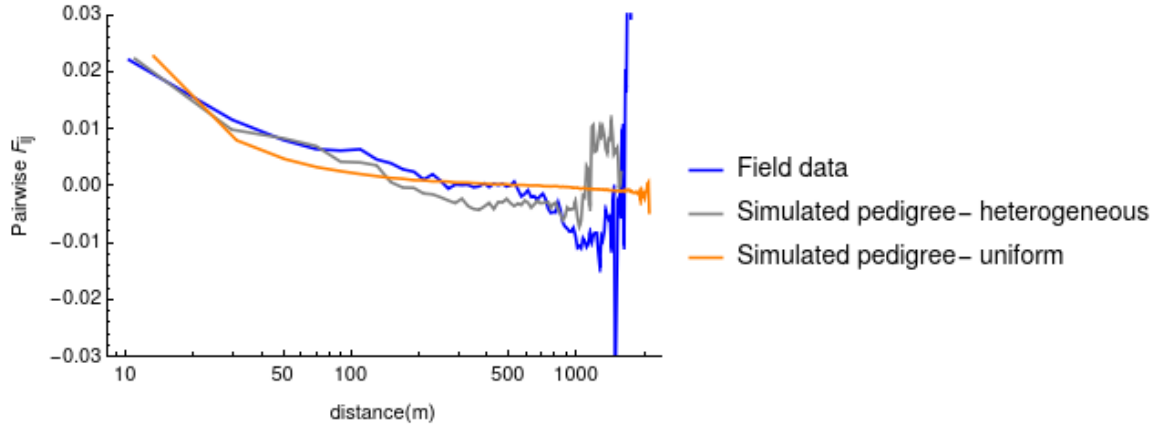
171 **Figure S8.** (A) Isolation by distance for the field data (blue) and five simulated
172 population pedigrees (gray) plotted on a log scale. (B) Isolation by distance from ten
173 replicates of a single pedigree along with their average shown in black.

174

175

176

177



178
179 **Figure S9.** Isolation by distance for the field data (blue), simulated pedigree with realistic
180 spatial structure (gray) and uniform density (orange) plotted on a log scale.
181

182
183 **Table S5.** F_{ST} , F_{IS} and g_2 values from the field data and from simulated (sim.) pedigrees
184 with heterogeneous and uniform density. For the simulations, mean \pm standard deviation
185 of ten replicate sets of genotypes are shown for each pedigree, across the five pedigree
186 means, and across all 50 replicates (ten replicates for five pedigrees).

		F_{ST}	F_{IS}	g_2
Field data		0.022	0.0211	0.00262
Sim. heterogeneous	Pedigree 1	0.0192 \pm 0.00383	0.0216 \pm 0.00236	0.00274 \pm 0.000723
	Pedigree 2	0.0226 \pm 0.00348	0.0254 \pm 0.00179	0.00258 \pm 0.000560
	Pedigree 3	0.0222 \pm 0.00221	0.0247 \pm 0.00149	0.00312 \pm 0.001120
	Pedigree 4	0.0239 \pm 0.00143	0.0244 \pm 0.00132	0.00240 \pm 0.000854
	Pedigree 5	0.0208 \pm 0.00346	0.0259 \pm 0.00160	0.00235 \pm 0.000448
	Across pedigree means	0.0217 \pm 0.0029	0.0244 \pm 0.0017	0.00264 \pm 0.000741
Across all 50 replicates		0.0217 \pm 0.0033	0.0244 \pm 0.0023	0.00264 \pm 0.000797
Sim. uniform	Pedigree 1	0.0226 \pm 0.0009	0.0203 \pm 0.00059	0.00171 \pm 0.000083

187

188 **References**

- 189 [1] M. Li *et al.*, “Genome structure and evolution of *Antirrhinum majus* L,” *Nat. Plants*, vol.
190 5, no. 2, pp. 174–183, 2019, doi: 10.1038/s41477-018-0349-9.
191 [2] B. Charlesworth and D. Charlesworth, *Elements of evolutionary genetics*. W. H. Freeman,
192 2010.

193