

Pedigree-based inbreeding coefficient explains more variation in fitness than heterozygosity at 160 microsatellites in a wild bird population

Pirmin Nietlisbach, Lukas F. Keller, Glauco Camenisch, Frédéric Guillaume, Peter Arcese, Jane M. Reid, Erik Postma

Electronic Supplementary Material:

Electronic Supplementary Material 1: Extended methods and results

Electronic Supplementary Material 2: Determinants of identity disequilibrium

Electronic Supplementary Material 3: Inbreeding with immigration

Electronic Supplementary Material 1: Extended methods and results

Methods

Study system

Mandarte is an island of ~6 ha in the Canadian Gulf Islands archipelago (48° 38' N, 123° 17' W), and the site of a long-term study on song sparrows [1]. All song sparrows resident on Mandarte Island (including immigrants that are caught with mist nets) have been color-banded for individual identification since 1975 [2]. Because effectively all adults alive on Mandarte Island are observed each year [resighting probability = 0.998; 3], the lifespan of resident individuals that died before 2013 (the cutoff for data used in this study) is known exactly. Annual local survival probability of adult song sparrows is ~0.6 [4]. Weather conditions during winter are a major factor affecting mortality [5,6], but predation by visiting birds of prey [7] and locally breeding glaucous-winged gulls (*Larus glaucescens*) [8] also occurs.

All territorial individuals are monitored and nests are typically located during incubation and nestlings are banded ~6 days after hatching [2]. The fate of nestlings is followed until they reach ~24 days of age, at which time they start becoming independent from parental care [9]. This intensive monitoring of reproductive success provides the lifetime number of banded offspring produced, as well as the number of adult offspring. Song sparrows are considered to be adult after surviving their first winter [9]. Sex was determined genetically based on a locus with a known length polymorphism between the sex chromosomes, as described elsewhere [10,11].

Data selection

Because the individuals that hatched in 2006 were the last cohort for which lifetime fitness measures were known for all individuals (they had all died before spring 2013), we limited the dataset to individuals hatched between 1993 and 2006 (but offspring that hatched later are included in the fitness measures of individuals that bred after 2006). As expected, the fitness measures we used were correlated: lifespan and number of banded offspring (Pearson's correlation coefficient $r = 0.88$), lifespan and number of adult offspring ($r = 0.75$), and numbers of banded and adult offspring ($r = 0.84$) (based on all 1919 individuals in this study for which all three fitness measures were known).

Pedigree

A blood sample of ~50 μ L has been taken from the brachial vein of all individuals at banding (at the age of ~6 days after hatching for nestlings, or in adulthood for immigrants) since 1993, and from a subset of individuals hatched during 1987-1992 [12]. Genotyping all individuals for which blood samples are available for 13 microsatellites allowed us to confirm that the social mother is indeed the genetic mother, and to correct the observed pedigree (i.e. the pedigree with the male and female that feed a given young as its parents) for extra-pair paternities, which occur at a rate of ~28% [see below; 13]. We used the package *MasterBayes* [14] in R v3.0.2 [15] to reconstruct the pedigree [13,16], which we confirmed using up to 170 additional microsatellites [see below; 10,17].

Because of complete sampling of all possible parents and the large genetic dataset, this genetically corrected pedigree contains very few errors and captures the great majority of variation in relatedness and inbreeding. The mean and maximum number of known ancestral generations was 8.3 and 16, respectively, with at least 5 ancestral generations known for 82% of the 1966 individuals used for the comparison of F and H . Similarly, for the 1426 individuals used in the analyses involving the number of banded or adult offspring as fitness traits (see main text), mean and maximum number of known ancestral generations were 7.5 and 12, respectively, with at least 5 ancestral generations known for 83% of the 1426 individuals. F was calculated using the R package *pedigreemm* [18].

From 1975 through to 1992, genetic sampling was incomplete and therefore the pedigree for these years is mostly based on observed parental behaviour. However, we grafted the genetically corrected pedigree onto the pedigree of the earlier years because it still contains information about the relatedness of the first genetically sampled individuals, and this is more informative than the alternative assumption that all of these individuals are unrelated [16,19]. Nevertheless, only F of individuals for which all four grandparents were genetically confirmed or which had an immigrant parent (for which it is assumed that $F=0$, see below) were used in the analyses.

Mandarte Island receives a mean of 1.1 immigrating song sparrows per year, which are caught with mist nets, colour-banded, and blood sampled. All immigrants are assumed to have arrived at the age of one year, and to be unrelated to locally hatched individuals. Because parents of immigrants are not known, immigrants have an unknown F , and matings between

immigrants and individuals hatched on Mandarte Island are assumed to produce outbred offspring with $F = 0$ [20–22].

Genotyping errors

The loci used in this study had mean allelic dropout and false allele rates of 0.38% and 0.33%, respectively [10]. To account for these genotyping errors, we masked (i.e. set to unknown) all genotypes at loci where offspring did not match their parents. We also masked the genotypes at loci that gave rise to a range of incompatibilities at the family level [see 10 for details]. Masking affected 0.4% of the 302,684 successfully typed single-locus genotypes (3.9% of the 314,560 single-locus genotypes were not successfully genotyped), resulting in 95.8% (= 301,466) of the single-locus genotypes being known among the 1966 individuals used to investigate the relationship of H and F . We only included individuals in the analyses for which at least 75% of the loci were successfully genotyped.

Effect of sex and phenotype-dependent inbreeding on quantification of inbreeding depression and heterozygosity-fitness correlations

We tested for sex-specific effects of F or H on all four components of fitness by fitting sex and the interaction between sex and F or H , but the effect of the interaction was small and not statistically significant for any of the four measures of fitness. Sex was only a significant predictor of variation in lifespan ($p = 0.03$), but its effect size was small (males lived on average 1.8 months longer than females), and including sex had a negligible effect (less than 0.3%) on the estimated slope of relative lifespan on F . For consistency with the other models, sex effects were therefore not included in the reported models.

Becker *et al.* [23] have shown that inbreeding depression is best quantified in a quantitative genetic framework, as this reduces biases introduced by phenotype-dependent inbreeding. Therefore we also estimated inbreeding depression in an animal model framework [24–26]. However, since the results did not differ appreciably (data not shown), we report only the results obtained with standard regression analyses.

Statistical testing

Although the used fitness measures are heavily zero-inflated, parameter estimates from ordinary least squares regression are unbiased and provide a valid estimate of the strength of selection [27, page 1219], and of the linear relationships derived and discussed in Szulkin *et al.* [28]. Importantly however, p -values accompanying these estimates will be unreliable [29

(page 92),30]. We therefore used randomization tests for hypothesis testing. We randomly ordered the response variable 10,000 times without replacement [31] and estimated slopes using ordinary least squares regression for each randomized dataset. The reported p -value was calculated as the proportion of absolute slopes obtained from randomized data that were more extreme than the observed absolute slope, making this a two-tailed test. In addition, we included a fitness component that was closer to (but still not perfectly) normally distributed: the number of adult offspring produced by adults. Confidence intervals (CI) were obtained by performing 10,000 bootstrap replicates (i.e. sampling datapoints with replacement) and applying the bias-corrected percentile method [29 (page 26),32]. Despite violations of the normality assumption, p -values calculated directly from linear models (data not shown) were very close to those from the randomization tests, testifying to the robustness of linear models to violations of the normality assumption [29 (page 92)]. Similarly, negative binomial generalized linear models also led to equivalent conclusions (data not shown). All statistical analyses were conducted in R v3.0.2 [15].

Testing for local effects

For completeness, we tested for single-locus effects following Szulkin *et al.* [28] by fitting two linear models, one with multilocus H as predictor, and one with single-locus heterozygosities of all 160 loci as separate predictors. If there are single locus effects, an F-ratio test comparing both models should show that the model with locus-specific heterozygosities provides a significantly better fit. The power of this test, especially when the number of loci is high, is however low [28]. Because expected regression coefficients are a function of heterozygosity, we normalized single-locus heterozygosities following Szulkin *et al.* [28 (Appendix 2)] and replaced missing genotypes with mean heterozygosity at that locus.

Results

Testing for local effects

For all four fitness traits, we compared linear models with multilocus H to models with single-locus heterozygosities for all 160 loci as predictors of fitness. There was no evidence that the single-locus heterozygosity models were significantly better than the models with multilocus H : lifespan ($F_{159,1271} = 1.02$, $p = 0.41$), lifetime number of banded offspring ($F_{159,1265} = 1.00$, $p = 0.50$), lifetime number of adult offspring ($F_{159,1265} = 1.16$, $p = 0.09$), and

lifetime number of adult offspring of adults ($F_{159,98} = 1.18, p = 0.19$). However, the statistical power of these comparisons is low.

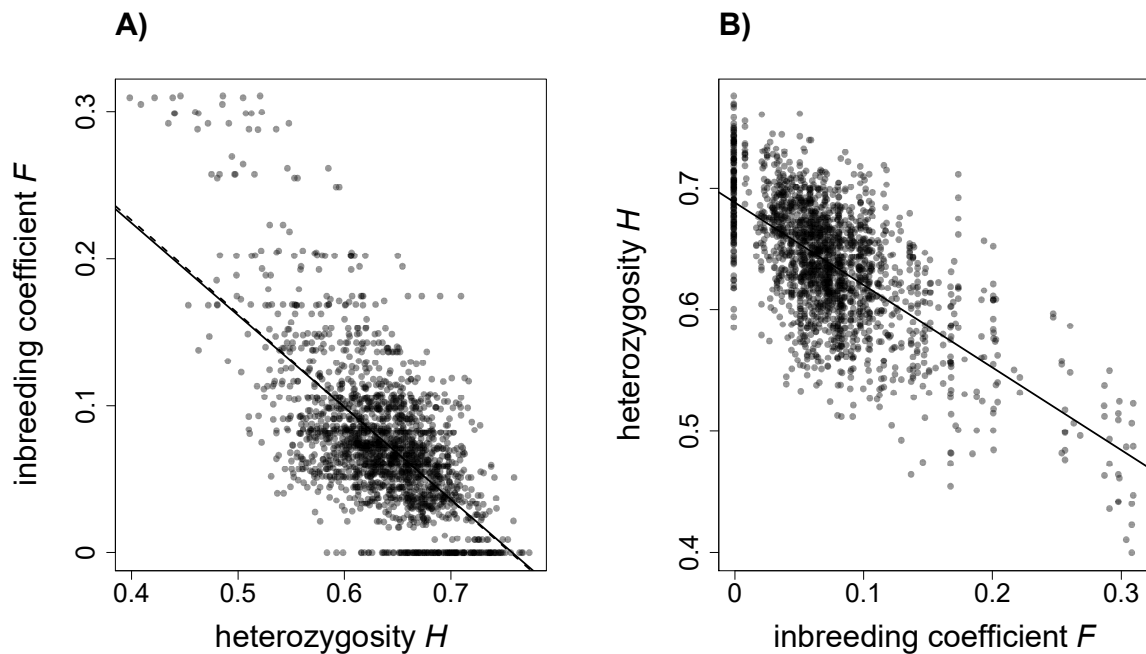


Figure S1. (A) Relationship between multilocus heterozygosity (H) and pedigree-based inbreeding coefficient (F). The observed (solid) regression line matches the predicted (dashed) regression line very closely. Note that we regressed F on H following Szulkin *et al.* [28], even though the causal relationship is in the other direction. **(B)** Regression of H on F . The intercept of a regression of homozygosity on F (or 1 minus the intercept of a regression of H on F) provides an estimate of the IBD-IBS discrepancy [33 (Supplementary Information)]. IBD-IBS discrepancy for our dataset is thus 31.2%.

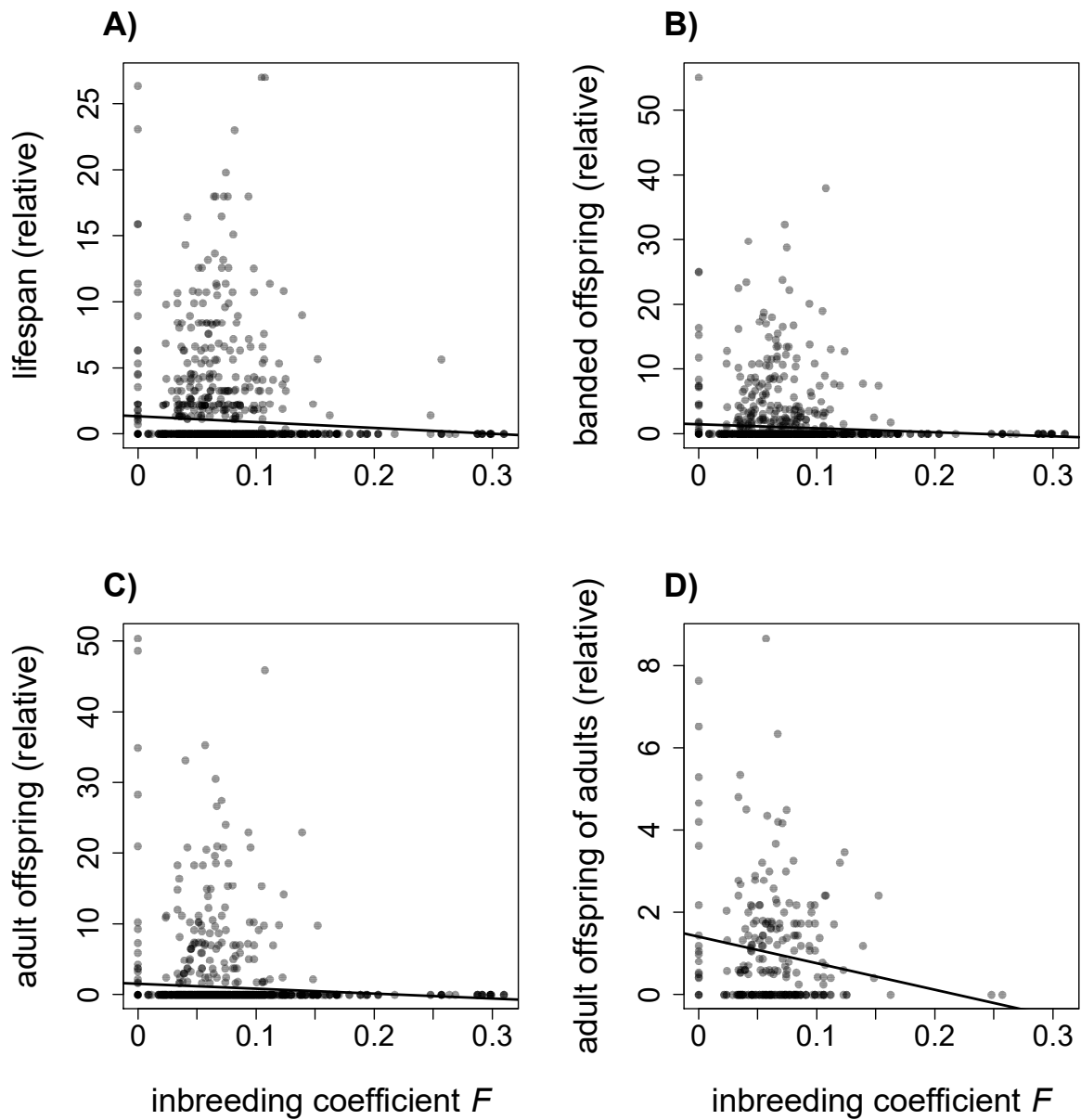


Figure S2. Inbreeding depression in four fitness traits: **(A)** lifespan, **(B)** lifetime number of banded offspring, **(C)** lifetime number of adult offspring, and **(D)** lifetime number of adult offspring produced by adult individuals. All relationships (solid lines) were statistically significant. F explained 0.5% (A), 0.6% (B and C), and 2.6% (D) of the variance in fitness traits.

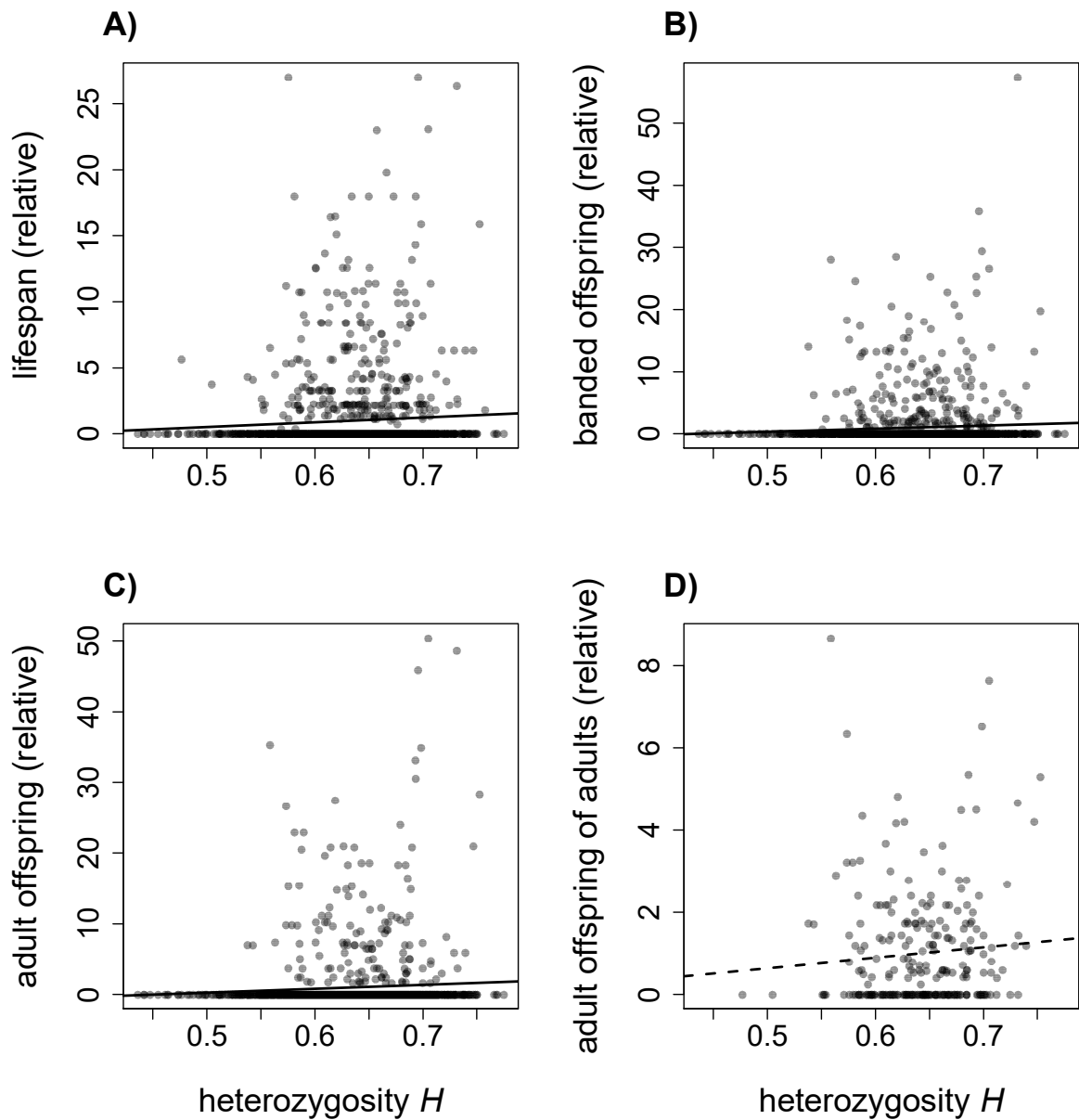


Figure S3. Heterozygosity H at 160 microsatellites predicts variation in four fitness traits: (A) lifespan, (B) lifetime number of banded offspring, (C) lifetime number of adult offspring, and (D) lifetime number of adult offspring produced by adult individuals. Statistically significant relationships (solid lines) were found in (A), (B), and (C), but not in (D) (dashed line). H explained 0.4% (A and B), 0.5% (C), and 0.6% (D) of the variance in fitness (see main text).

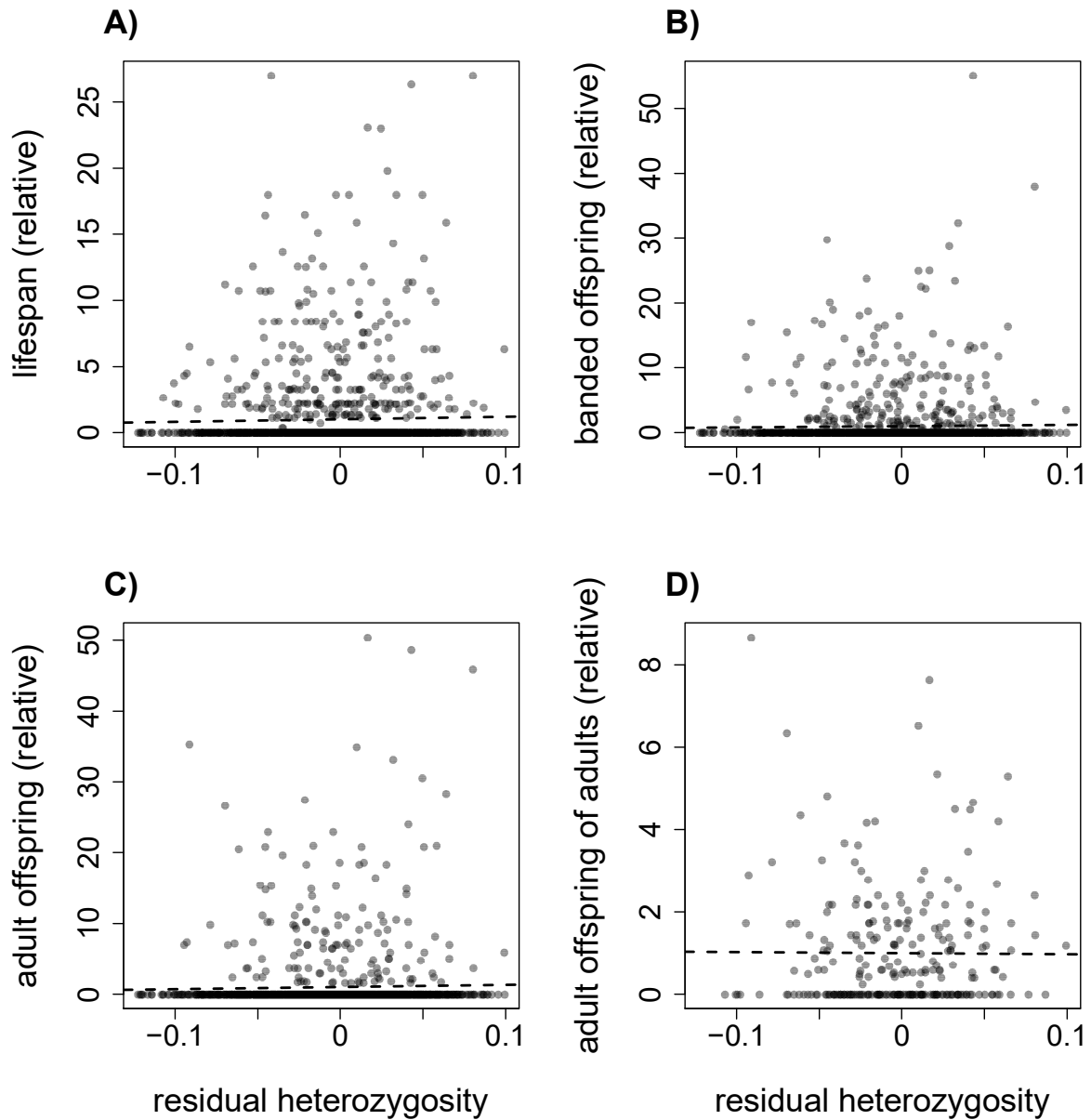


Figure S4. H does not explain variation in four fitness traits above and beyond the variation explained by pedigree-based inbreeding F . Residual heterozygosity (derived from a regression of H on F for illustrative purposes only, not for statistical testing) is plotted against four relative measures of fitness: (A) lifespan, (B) lifetime number of banded offspring, (C) lifetime number of adult offspring, and (D) lifetime number of adult offspring produced by adult individuals. None of the relationships are statistically significant (dashed lines).

Table S1. Expected and observed correlations and slopes for multilocus heterozygosity (H) across 160 (real and simulated unlinked) microsatellites, pedigree-based inbreeding coefficients (F), and four fitness measures: lifespan, lifetime number of banded or adult offspring, and lifetime number of adult offspring produced by adult individuals. Expected heterozygosity-fitness correlations (HFC) are calculated as the product of expected correlation coefficients (or slopes) of F and H , and observed correlation coefficients (or slopes) of fitness and F [28; see equations 1.1 and 1.2 in main text]. Note that these expectations assume the absence of Mendelian noise (the amount of which is unknown for our population) and are therefore downwardly biased (see main text for details).

	correlation				slope			
F and H (observed)	-0.653				-0.627			
F and H (expected)	-0.662				-0.635			
F and H (simulated H)	-0.679				-0.669			
	lifespan	banded	adult	adults of adults	lifespan	banded	adult	adults of adults
fitness and F (observed)	-0.070	-0.078	-0.077	-0.158	-4.460	-6.350	-6.985	-6.436
fitness and H (observed)	0.062	0.064	0.068	0.079	3.598	4.650	5.590	2.530
fitness and H (expected)	0.046	0.052	0.051	0.104	2.832	4.032	4.435	4.087
difference observed to expected	35%	24%	34%	-25%	27%	15%	26%	-38%
fitness and H (simulated H)	0.047	0.053	0.053	0.093	2.912	4.125	4.609	3.051
difference simulated to expected	2%	3%	4%	-11%	3%	2%	4%	-25%

Electronic Supplementary Material 2: Determinants of identity disequilibrium

Introduction

The theory that relates identity disequilibrium g_2 to the mean and variance in pedigree-based F [28] assumes that loci are physically unlinked [34]. However, as discussed in David *et al.* [34] and Szulkin *et al.* [28], an effect of gametic phase disequilibrium (i.e. linkage disequilibrium) among physically linked loci on identity disequilibrium g_2 is theoretically expected. Additionally, the number of genetic markers used to estimate g_2 , as well as relatedness and inbreeding not captured by the pedigree, may influence g_2 . Here we investigate this in our dataset using a series of analyses and simulations.

Methods

Subsets of real microsatellites

To compare g_2 calculated from all microsatellites with subsets of unlinked loci, we capitalised on a genetic recombination map constructed for the Mandarte Island song sparrow population. This allowed us to estimate g_2 from markers with varying degrees of physical linkage and gametic phase disequilibrium [10].

In a first analysis, we selected loci that were at least 50 centiMorgan apart (or part of different linkage groups) and that had a LOD score of less than 2. In a second analysis, we iteratively selected loci that were in significant gametic phase disequilibrium with the fewest other loci in the selected set, allowing loci to be in significant gametic phase disequilibrium with maximally three other loci in the selected set. For both subsets of loci, we calculated g_2 using the approximations in Hoffman *et al.* [35, Supporting Information] and obtained its 95% confidence interval by bootstrapping 10,000 times across individuals.

Number of loci

To investigate how the number of loci influences g_2 , we randomly sampled without replacement the following number of loci from all available 160 loci: 5, 10, 15, 20, 30, 40, 50, 75, 100, 125, 150. For each number of loci, we performed 500 repetitions. We also used the

full dataset of 160 loci. Note that especially for the larger numbers of loci, the same loci will have been included in most of the replicate datasets, and that the full dataset with 160 loci was not resampled. We measured g_2 for each subset of marker data as above. Additionally, we simulated unlinked loci in the same way as described below. For each number of loci (see above), 500 simulations were performed, and all loci were simulated independently.

Pedigree depth and tight linkage

To explore what effect pedigree depth has on g_2 , we conducted simulations using the song sparrow pedigree truncated at three different time points, and four different patterns of linkage among 160 simulated loci.

We used the complete available pedigree dating back to 1975, as well as pedigrees where we removed some or most of the ancestral pedigree information by setting the first year of pedigree data to 1984 or 1993. In each case, g_2 was calculated only across the same 1966 individuals for which H and sufficient pedigree data for calculation of F were available for the main analyses (see main text for details). If unknown founder relatedness and inbreeding has an effect on g_2 in our dataset, we expect g_2 in the simulated datasets to differ if it is calculated with many (pedigree since 1975) or with few ancestral generations (pedigree since 1993) known for the analysed individuals. Because all analysed individuals in the larger two pedigrees had at least two generations of fully known ancestors, but most of them had many more, we expected effects of unknown founder relatedness and inbreeding to be small in this dataset. The simulations were conducted 100 times for all three pedigrees and all four types of linkage patterns, and the mean of g_2 , as well as the range of the central 95% of g_2 values were extracted.

To qualitatively assess if pedigree depth has different effects on g_2 depending on linkage pattern, we first simulated unlinked loci (“unlinked and low LD” in Figure S6). We generated marker data by sampling with replacement alleles of pedigree founders. Subsequently genotypes for all individuals in the pedigree were derived by simulating Mendelian inheritance through the pedigree [i.e. gene dropping; 36]. Note that because for most of the founders no genetic data is available, precise founder allele frequencies are unknown. However, for the purpose of this simulation this is unproblematic as any set of more or less realistic allele frequencies is sufficient, so we used allele frequencies across all genotyped individuals, extracted using the R package *MasterBayes* [14].

Second, we used the same gene dropping procedure as above to simulate 80 pairs of linked loci. Because the founder alleles were sampled independently for each individual, founder alleles were in gametic phase equilibrium. However, gametic phase disequilibrium quickly increased as recombination between the maternally and paternally inherited alleles of the two members of each randomly selected pair of loci at each meiosis was set to 5%. (“linked and low LD” in Figure S6). Although our choice of a recombination frequency of 5% is somewhat arbitrary, it provides a qualitative assessment of the effect of strong linkage on g_2 .

In a third set of simulations, we again set recombination between the maternally and paternally inherited alleles of the two members of each locus pair to a probability of 5% at each meiosis, but to generate the genotypes of the founding individuals we this time assigned the same alleles to both loci of a pair (“linked and high LD” in Figure S6). This resulted in strong gametic phase disequilibrium among the founders, which decreased in subsequent generations due to recombination. Correlation-based estimates of gametic phase disequilibrium among linked loci (only 0.6% of all possible pairs of loci) when using the complete pedigree ranged from 0.24 to 0.85, with a mean of 0.39, and a median of 0.37. Thus, mean and median gametic phase disequilibrium among simulated linked loci was close to the maximum of 0.38 observed in our empirical marker data [10].

Finally for the fourth set of simulations, we simulated a situation where pairs of loci were sampled as before, so starting in extreme gametic phase disequilibrium, but this time we allowed for free recombination among all loci (“unlinked and high LD” in Figure S6).

Simulations of biallelic loci in a bird genome

The simulations described above simulated linkage in a qualitative manner and aimed at examining the effect of pedigree depth on g_2 . More realistic simulations of actual genomes are required to obtain a more quantitative estimate of the influence of linkage on g_2 . We therefore conducted simulations of loci in a realistic bird genome using the genetically explicit simulation framework Nemo v2.3.46r4 [37]. We used data about the size of great tit (*Parus major*) chromosomes (measured in number of nucleotides) from Laine *et al.* [38; accessed on 30/April/2016 on https://genomes.bioinf.nioo.knaw.nl/jbrowse/?data=genomes%2FParus_major_v104&loc=chr%3A1..150265477&tracks=DNA&highlight=].

We randomly distributed 49,998 biallelic loci across the 28 autosomes with available size data. For each of these chromosomes, recombination map length (measured in centiMorgans) was taken from van Oers *et al.* [39] as the mean of the map length of their two study

populations. We used the great tit as an example of a typical bird genome because it is one of the few species with good data on size and recombination map length of most of its chromosomes. Given that number and size of chromosomes is well conserved among birds [40], the great tit genome can be considered as representative of the general structure of most avian genomes.

In order to generate realistic patterns of genetic diversity and gametic phase disequilibrium, we first simulated 10,000 generations in an island model of 50 patches with an adult population size of 100 individuals each. The patches were connected in an island model with a migrant pool consisting of 1% of the individuals. This leads to a mean immigration rate of one immigrant per generation and patch, which is similar to the ~ 1.1 immigrants observed on average per year on Mandarte Island. We simulated a system of randomly paired monogamous reproduction with a proportion of 28% extra-pair paternities, as is the case for the song sparrows on Mandarte Island [13]. The mean number of offspring per female was set to 6. The first generation was initialized by randomly sampling alleles for all biallelic loci (thus mimicking single nucleotide polymorphisms). The mutation rate was set to 0.00001 per locus and generation.

After 10,000 generations, adults from one patch were randomly sampled with replacement to be the founders of the pedigree of the song sparrows of Mandarte Island. As with some simulations described above, the whole pedigree starting in 1975 was used. We then selected the genotypes of the 1966 individuals with sufficient pedigree information (i.e. the same individuals used throughout most of the previous analyses) and calculated identity disequilibrium g_2 as before for various subsets of the simulated loci. First, we calculated g_2 across all loci. Then, we selected polymorphic loci that were at least 10, 20, 30, 40, or 50 centiMorgan apart on the same chromosome (or that were on different chromosomes), and calculated g_2 for those. For comparison, we also randomly selected the same number of polymorphic loci for each subset and calculated g_2 across those loci. These subsets allow us to compare how g_2 changes when it is calculated across increasingly less linked loci as compared to randomly selected loci. Note that removing linked loci reduces the number of loci available, because eventually only the large chromosomes contribute more than one locus. Additionally, we performed the same simulations for 160 unlinked loci and 160 loci with the empirically determined recombination distances of the 160 real short tandem repeat loci of our dataset [10]. All simulations were run 100 times and the mean and central 95% of g_2 values extracted.

Results

Subsets of real microsatellites

Identity disequilibrium g_2 calculated across 50 loci that were at least 50 centiMorgan apart was 0.0032 (95% CI = 0.0022 to 0.0043). g_2 across 32 loci that were in little gametic phase disequilibrium with each other was 0.0028 (95% CI = 0.0007 to 0.0048). Thus, these subsets of markers yielded g_2 estimates that were very close to the pedigree-based expectation of 0.0030, and considerably lower than the value of 0.0043 (95% CI = 0.0037 to 0.0050) for g_2 based on the full set of 160 microsatellites (see main text). However, because these subsets contained relatively few markers, estimates are rather imprecise.

Number of loci

As expected, g_2 averaged across simulations was independent of the number of loci on which it was based, but sampling variance around the mean decreased considerably with an increasing number of loci (Figure S5). Note that because we averaged across replicates, the average amount of linkage is independent of the number of loci that were sampled.

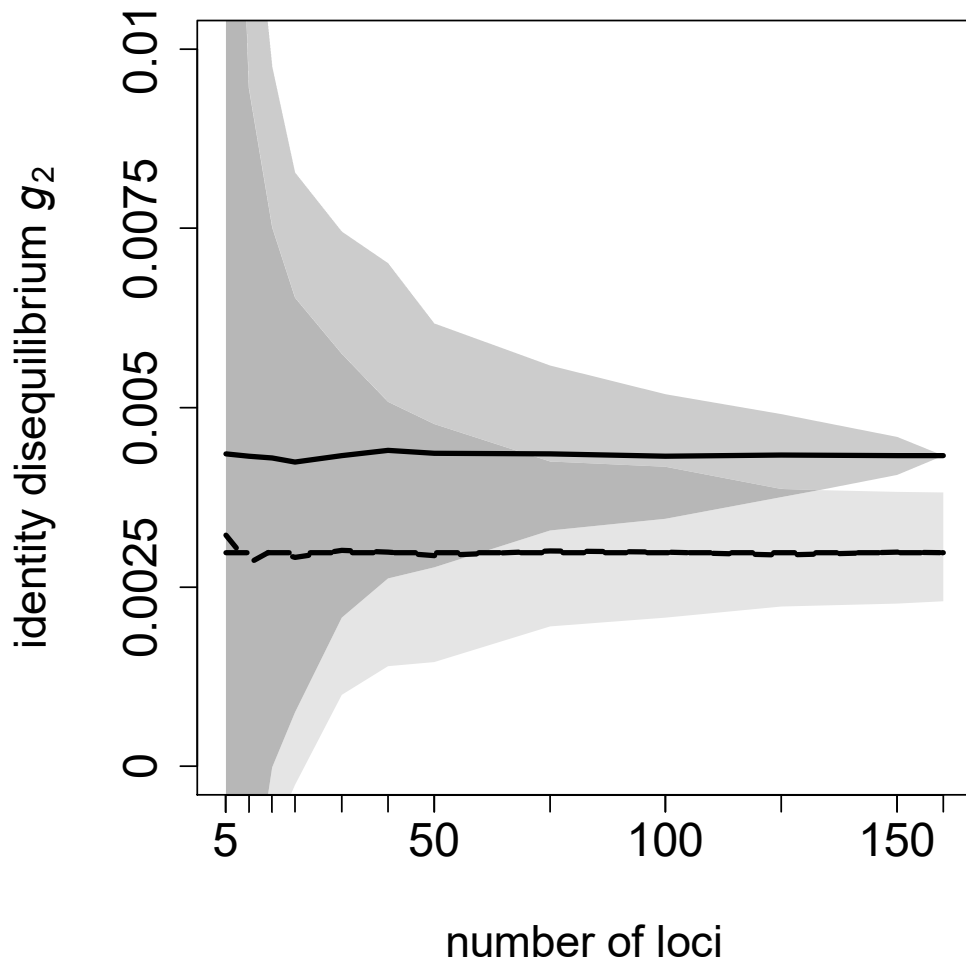


Figure S5. Identity disequilibrium g_2 calculated from different numbers of microsatellites. Estimates from real microsatellites (solid line) very quickly reached a mean value of ca. 0.0043 across 500 randomly selected subsets of increasing numbers of loci (darker grey area represents the spread of the central 95% of values). Values from 500 simulations of Mendelian inheritance at unlinked loci across the song sparrow pedigree (dashed line) quickly approached a mean value of 0.0030 (lighter grey area represents the spread of the central 95% of simulated values), matching exactly with g_2 expected based on mean and variance in pedigree-based F (dotted line; mostly overlaying the dashed line).

Pedigree depth and tight linkage

Figure S6 shows the effects of pedigree depth on g_2 for four different simulated linkage patterns. For unlinked and linked datasets with low gametic phase disequilibrium, g_2 estimated from pedigrees going back to 1975, 1984, or 1993 did not differ. In our dataset, the effects of limited information on founder relatedness and inbreeding had therefore negligible effects on g_2 when using unlinked markers. Only for the simulations of linked loci with strong gametic phase disequilibrium among founders was g_2 estimated from pedigrees going back to 1993 slightly higher than for pedigrees going back to 1975 or 1984. This is because in these simulations gametic phase disequilibrium is highest among founders and will decrease through successive generations of the pedigree due to recombination. Furthermore, in all simulated cases, g_2 measured from pedigrees going back to 1975 and 1984 did not differ. This shows that our pedigree is deep enough so that its estimates of g_2 are not unduly affected by unknown founder relatedness or inbreeding.

Furthermore, Figure S6 shows that strong physical linkage of pairs of loci only slightly increased estimates of g_2 , because linkage without gametic phase disequilibrium appears to have a negligible effect on g_2 . When sampling founder alleles so that they were in high gametic phase disequilibrium, g_2 was considerably higher, but only so if loci were physically linked. If loci were physically unlinked, initial gametic phase disequilibrium had no noteworthy influence on g_2 among the analysed individuals.

In many realistic situations, gametic phase disequilibrium among founders may also occur between unlinked loci and is likely to be the consequence of some sort of population structure. However, these simulations show that our song sparrow pedigree is deep enough so that such unaccounted population structure among founders only affects our estimates of g_2 if loci are physically linked. This indicates that the higher g_2 measured from marker data as compared to its expectation from pedigree-based F is most likely due to the presence of physically linked markers. Although limited pedigree information had only negligible effects on g_2 in our dataset, this may not be generally the case, because also in our most limited case, we still had considerable amounts of pedigree information available for most analysed individuals, as the analysed individuals lived during 1993 until 2012.

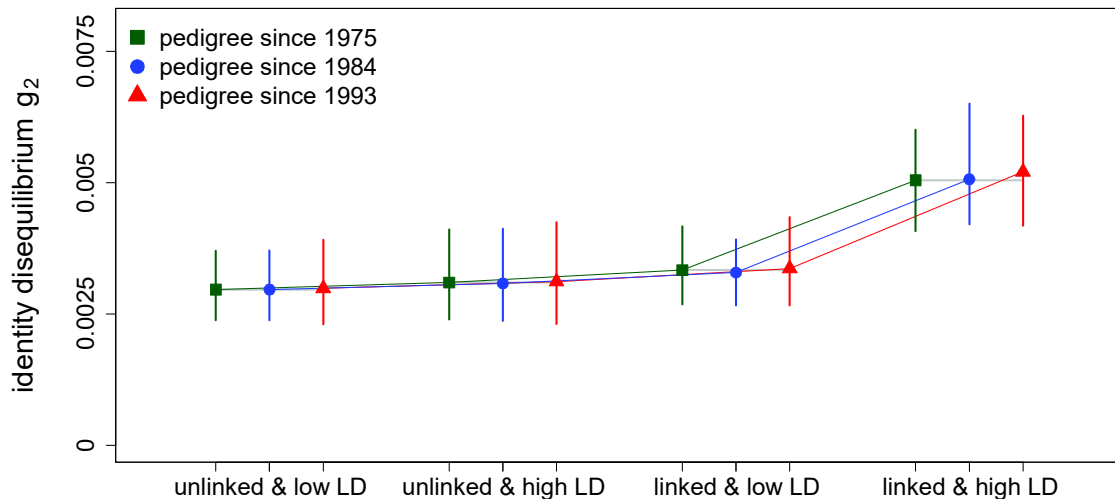


Figure S6. Effects of linkage, gametic phase disequilibrium in the founding population (LD), and pedigree depth on identity disequilibrium. Means (squares, dots, or triangles) and the range of the central 95% of values (vertical lines) are shown for 100 repetitions of each of the twelve different simulations. See main text of the Electronic Supplementary Material for details about the simulations.

Simulations of biallelic loci in a bird genome

Results from simulations of 49,998 biallelic loci distributed across a realistic bird genome are shown in Figure S7. Using unlinked loci yields on average consistently smaller values of g_2 than using all loci or randomly selected loci (i.e. red triangles in Figure S7 are at higher values of g_2 than green squares). As a consequence of increasingly smaller sets of loci with decreasing linkage, precision of g_2 decreases considerably. Note that the ranges of g_2 values calculated from unlinked and randomly selected loci are largely overlapping for our dataset, and unless expected g_2 values are much larger than in our dataset, detecting differences in a single dataset may therefore often be impossible. As expected, mean g_2 calculated from randomly selected loci remains similar to g_2 calculated across all loci. When using only effectively unlinked loci (i.e. loci that are at least 50 centiMorgan apart), mean g_2 across 100 replicates equals $g_2 = 0.0030$, and is therefore equal to g_2 calculated from the pedigree. Similarly, simulation of 160 unlinked loci yields a mean of $g_2 = 0.0031$. However, when simulating 160 loci with the empirically observed linkage patterns, mean $g_2 = 0.0041$, which is very similar to $g_2 = 0.0043$ as measured from our real microsatellite dataset, suggesting

that genotyping errors (which are absent from the simulations) do not contribute to the discrepancy between observed and expected g_2 . Note that the simulations of 160 unlinked loci and 160 loci with empirically observed linkage patterns differ in nothing else except for their patterns of linkage.

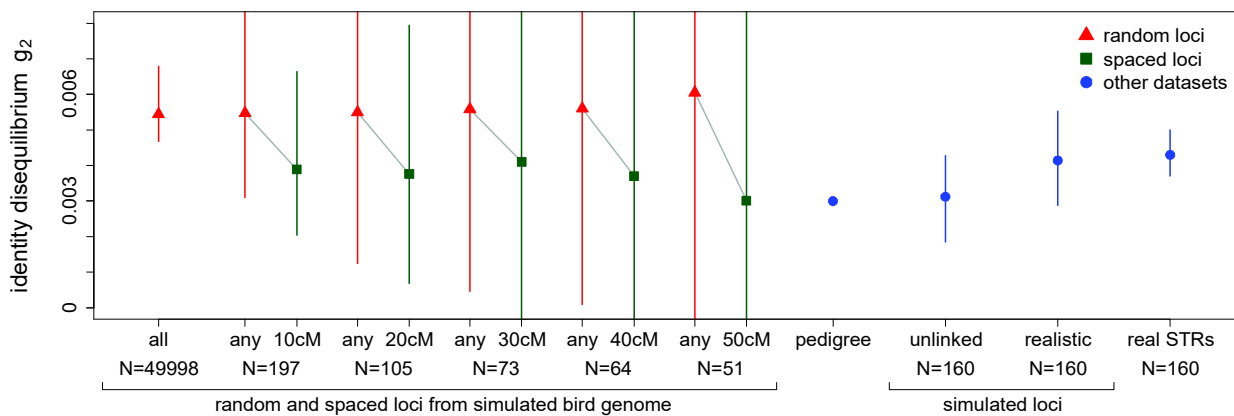


Figure S7. Identity disequilibrium g_2 calculated from a simulated dense dataset of single-nucleotide polymorphisms (“all”), subsets of loci of that dataset that are at least 10, 20, 30, 40, or 50 centiMorgan (cM) apart or from equal numbers of randomly selected (“any”) loci (green squares and red triangles, respectively). Additionally, 160 unlinked loci and 160 loci (“realistic”) with the recombination patterns of the real short tandem repeat dataset were also simulated. Furthermore, g_2 calculated from the pedigree, and from the observed microsatellite dataset (“real STRs”) are also shown. See the main text of the Electronic Supplementary Material for details about the simulations.

Discussion

In summary, we have shown here that unknown relatedness and inbreeding in the founder population has a negligible effect on g_2 in our dataset. This is in line with other simulations that have shown that distant shared ancestors contribute little to variance in inbreeding [41]. Gametic phase disequilibrium among physically linked loci on the other hand increases g_2 , as expected from theory [34]. In line with this, Forstmeier *et al.* [42] found that heterozygosity-heterozygosity correlations between unlinked sets of loci were much lower than between random sets of loci. Thus, we have shown that the fact that in our dataset identity disequilibrium g_2 is higher than expected from the pedigree is attributable to linkage rather than to incomplete pedigree information. Our simulations also show that g_2 calculated from datasets consisting of a large number of genetic loci will lead to values of g_2 that are larger than expected based on pedigree F , and g_2 calculated from closely spaced loci therefore does not relate to the mean and variance of pedigree-based inbreeding in the way described by Szulkin *et al.* [28]. Although this means that g_2 estimated from very dense sets of loci is unsuitable for inferring pedigree-based characteristics such as selfing rate [34], it does provide information about variance in realized identity-by-descent. Thus, our simulations should not be interpreted as advice against using dense marker sets to calculate g_2 , but they rather show that care is required when interpreting values of g_2 .

Electronic Supplementary Material 3: Inbreeding with immigration

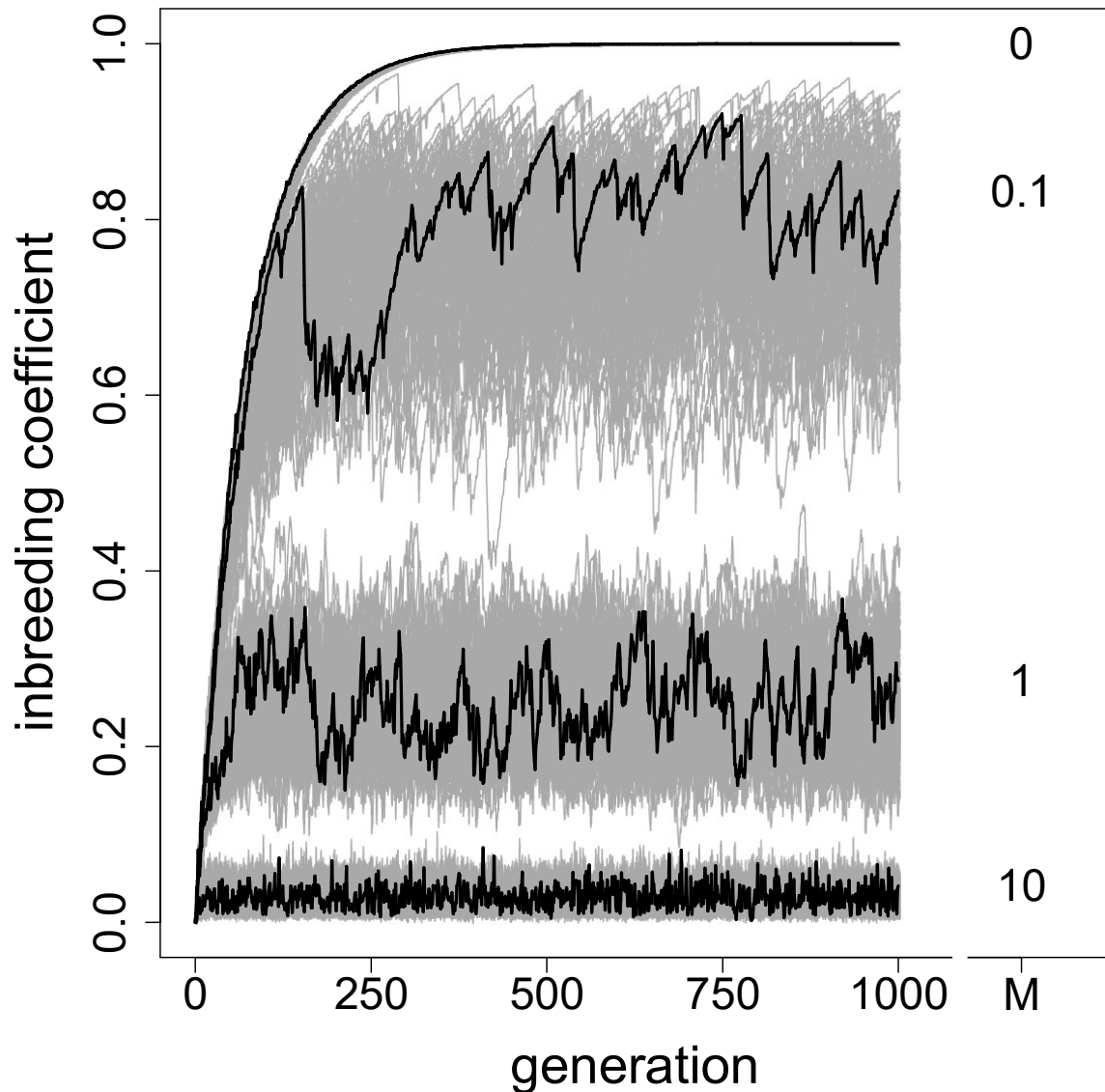


Figure S8. Mean inbreeding coefficients across individuals per generation simulated in a randomly mating population of 50 hermaphroditic individuals capable of selfing. With immigration, the mean inbreeding coefficients fluctuate around an equilibrium value that depends on the expected number of unrelated immigrants M per generation. Simulations were run with M of 0, 0.1, 1, and 10 expected immigrants per generation, and repeated 100 times to illustrate variation. One repetition for each M is plotted as a black line, the remaining repetitions as grey lines. Each individual had the same Poisson distributed probability of producing a mean of 3 offspring. Of all offspring, 50 were randomly selected to form the next generation. Thus, population size was kept constant and generations were non-overlapping. Mutations were ignored in these simulations, and hence average inbreeding goes to unity in the absence of immigration.

References in Electronic Supplementary Material

1. Tompa, F. S. 1964 Factors determining the numbers of song sparrows, *Melospiza melodia* (Wilson), on Mandarte Island, B. C., Canada. *Acta Zool. Fenn.* **109**.
2. Smith, J. N. M. 2006 Song sparrows, Mandarte Island, and methods. In *Conservation and biology of small populations: the song sparrows of Mandarte Island* (eds J. N. M. Smith L. F. Keller A. B. Marr & P. Arcese), New York, United States of America: Oxford University Press.
3. Wilson, S., Norris, D. R., Wilson, A. G. & Arcese, P. 2007 Breeding experience and population density affect the ability of a songbird to respond to future climate variation. *Proc. R. Soc. B* **274**, 2539–2545. (doi:10.1098/rspb.2007.0643)
4. Smith, J. N. M., Marr, A. B. & Hochachka, W. M. 2006 Life history: patterns of reproduction and survival. In *Conservation and biology of small populations: the song sparrows of Mandarte Island* (eds J. N. M. Smith L. F. Keller A. B. Marr & P. Arcese), New York, United States of America: Oxford University Press.
5. Rogers, C. M., Smith, J. N. M., Hochachka, W. M., Cassidy, A. L. E. V, Taitt, M. J., Arcese, P. & Schluter, D. 1991 Spatial variation in winter survival of song sparrows *Melospiza melodia*. *Ornis Scand.* **22**, 387–395.
6. Smith, J. N. M., Marr, A. B., Arcese, P. & Keller, L. F. 2006 Fluctuations in numbers: population regulation and catastrophic mortality. In *Conservation and biology of small populations: the song sparrows of Mandarte Island* (eds J. N. M. Smith L. F. Keller A. B. Marr & P. Arcese), New York, United States of America: Oxford University Press.
7. Arcese, P. 1989 Intrasexual competition, mating system and natal dispersal in song sparrows. *Anim. Behav.* **38**, 958–979.
8. Nietlisbach, P., Germain, R. R. & Bousquet, C. A. H. 2014 Observations of Glaucous-winged gulls preying on passerines at a Pacific Northwest colony. *Wilson J. Ornithol.* **126**, 155–158.
9. Arcese, P., Sogge, M. K., Marr, A. B. & Patten, M. A. 2002 Song sparrow (*Melospiza melodia*). In *The Birds of North America Online* (ed A. Poole), Ithaca, United States of America: Cornell Lab of Ornithology.
10. Nietlisbach, P., Camenisch, G., Bucher, T., Slate, J., Keller, L. F. & Postma, E. 2015 A microsatellite-based linkage map for song sparrows (*Melospiza melodia*). *Mol. Ecol. Resour.* **15**, 1486–1496.
11. Postma, E., Heinrich, F., Koller, U., Sardell, R. J., Reid, J. M., Arcese, P. & Keller, L. F. 2011 Disentangling the effect of genes, the environment and chance on sex ratio variation in a wild bird population. *Proc. R. Soc. B* **278**, 2996–3002. (doi:10.1098/rspb.2010.2763)

12. Jeffery, K. J., Keller, L. F., Arcese, P. & Bruford, M. W. 2001 The development of microsatellite loci in the song sparrow, *Melospiza melodia* (Aves) and genotyping errors associated with good quality DNA. *Mol. Ecol. Notes* **1**, 11–13. (doi:10.1046/j.1471-8278.2000.00005.x)
13. Sardell, R. J., Keller, L. F., Arcese, P., Bucher, T. & Reid, J. M. 2010 Comprehensive paternity assignment: genotype, spatial location and social status in song sparrows, *Melospiza melodia*. *Mol. Ecol.* **19**, 4352–4364. (doi:10.1111/j.1365-294X.2010.04805.x)
14. Hadfield, J. D., Richardson, D. S. & Burke, T. 2006 Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Mol. Ecol.* **15**, 3715–3730. (doi:10.1111/j.1365-294X.2006.03050.x)
15. R Core Team 2013 R: A language and environment for statistical computing.
16. Reid, J. M., Keller, L. F., Marr, A. B., Nietlisbach, P., Sardell, R. J. & Arcese, P. 2014 Pedigree error due to extra-pair reproduction substantially biases estimates of inbreeding depression. *Evolution.* **68**, 802–815. (doi:10.1111/evo.12305)
17. Reid, J. M., Arcese, P., Keller, L. F., Germain, R. R., Duthie, A. B., Losdat, S., Wolak, M. E. & Nietlisbach, P. 2015 Quantifying inbreeding avoidance through extra-pair reproduction. *Evolution.* **69**, 59–74.
18. Vazquez, A. I., Bates, D. M., Rosa, G. J. M., Gianola, D. & Weigel, K. A. 2010 Technical note: an R package for fitting generalized linear mixed models in animal breeding. *J. Anim. Sci.* **88**, 497–504. (doi:10.2527/jas.2009-1952)
19. Reid, J. M., Arcese, P., Sardell, R. J. & Keller, L. F. 2011 Additive genetic variance, heritability, and inbreeding depression in male extra-pair reproductive success. *Am. Nat.* **177**, 177–187. (doi:10.1086/657977)
20. Marr, A. B., Keller, L. F. & Arcese, P. 2002 Heterosis and outbreeding depression in descendants of natural immigrants to an inbred population of song sparrows (*Melospiza melodia*). *Evolution.* **56**, 131–142. (doi:10.1111/j.0014-3820.2002.tb00855.x)
21. Keller, L. F. 1998 Inbreeding and its fitness effects in an insular population of song sparrows (*Melospiza melodia*). *Evolution.* **52**, 240–250.
22. Reid, J. M., Arcese, P. & Keller, L. F. 2006 Intrinsic parent-offspring correlation in inbreeding level in a song sparrow (*Melospiza melodia*) population open to immigration. *Am. Nat.* **168**, 1–13.
23. Becker, P. J. J., Hegelbach, J., Keller, L. F. & Postma, E. 2016 Phenotype-associated inbreeding biases estimates of inbreeding depression in a wild bird population. *J. Evol. Biol.* **29**, 35–46.
24. Henderson, C. R. 1984 *Applications of linear models in animal breeding models*. Guelph, Canada: University of Guelph Press.

25. Kruuk, L. E. B. 2004 Estimating genetic parameters in natural populations using the “animal model.” *Philos. Trans. R. Soc. London Ser. B - Biol. Sci.* **359**, 873–890. (doi:10.1098/rstb.2003.1437)
26. Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B. & Nussey, D. H. 2010 An ecologist’s guide to the animal model. *J. Anim. Ecol.* **79**, 13–26. (doi:10.1111/j.1365-2656.2009.01639.x)
27. Lande, R. & Arnold, S. J. 1983 The measurement of selection on correlated characters. *Evolution.* **37**, 1210–1226.
28. Szulkin, M., Bierne, N. & David, P. 2010 Heterozygosity-fitness correlations: a time for reappraisal. *Evolution.* **64**, 1202–1217. (doi:10.1111/j.1558-5646.2010.00966.x)
29. Quinn, G. P. & Keough, M. J. 2002 *Experimental design and data analysis for biologists*. Cambridge, United Kingdom: University Press.
30. Janzen, F. J. & Stern, H. S. 1998 Logistic regression for empirical studies of multivariate selection. *Evolution.* **52**, 1564–1571.
31. Crowley, P. H. 1992 Resampling methods for data analysis in ecology and evolution. *Annu. Rev. Ecol. Syst.* **23**, 405–447.
32. Efron, B. & Gong, G. 1983 A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am. Stat.* **37**, 36–48.
33. Knief, U., Kempnaers, B. & Forstmeier, W. 2017 Meiotic recombination shapes precision of pedigree- and marker-based estimates of inbreeding. *Heredity.* **118**, 239–248. (doi:10.1038/hdy.2016.95)
34. David, P., Pujol, B., Viard, F., Castella, V. & Goudet, J. 2007 Reliable selfing rate estimates from imperfect population genetic data. *Mol. Ecol.* **16**, 2474–2487. (doi:10.1111/j.1365-294X.2007.03330.x)
35. Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A. S., Lacy, R. C. & Dasmahapatra, K. K. 2014 High-throughput sequencing reveals inbreeding depression in a natural population. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3775–3780. (doi:10.1073/pnas.1318945111)
36. MacCluer, J. W., VandeBerg, J. L., Read, B. & Ryder, O. A. 1986 Pedigree analysis by computer simulation. *Zoo Biol.* **5**, 147–160.
37. Guillaume, F. & Rougemont, J. 2006 Nemo: an evolutionary and population genetics programming framework. *Bioinformatics* **22**, 2556–2557. (doi:10.1093/bioinformatics/btl415)
38. Laine, V. N. et al. 2016 Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* **7**, 10474. (doi:10.1038/ncomms10474)

39. Van Oers, K., Santure, A. W., De Cauwer, I., van Bers, N. E. M., Crooijmans, R. P. M. A., Sheldon, B. C., Visser, M. E., Slate, J. & Groenen, M. A. M. 2014 Replicated high-density genetic maps of two great tit populations reveal fine-scale genomic departures from sex-equal recombination rates. *Heredity*. **112**, 307–316. (doi:10.1038/hdy.2013.107)
40. Ellegren, H. 2010 Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol.* **25**, 283–291. (doi:10.1016/j.tree.2009.12.004)
41. Keller, M. C., Visscher, P. M. & Goddard, M. E. 2011 Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* **189**, 237–249. (doi:10.1534/genetics.111.130922)
42. Forstmeier, W., Schielzeth, H., Mueller, J. C., Ellegren, H. & Kempenaers, B. 2012 Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Mol. Ecol.* **21**, 3237–3249. (doi:10.1111/j.1365-294X.2012.05593.x)