MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/j.1365-294X.2012.05593.x

Heterozygosity–fitness correlations in zebra finches: microsatellite markers can be better than their reputation

WOLFGANG FORSTMEIER,* HOLGER SCHIELZETH,†[1] JAKOB C. MUELLER,* HANS ELLEGREN† and BART KEMPENAERS*

Max Planck Institute for Ornithology, Eberhard-Gwinner-Str. 8, 82319 Seewiesen, Germany, †Department of Evolutionary Biology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden

Abstract

Numerous studies have reported associations between heterozygosity in microsatellite markers and fitness-related traits (heterozygosity–fitness correlations, HFCs). However, it has often been questioned whether HFCs reflect general inbreeding depression, because a small panel of microsatellite markers does not reflect very well an individual's inbreeding coefficient (*F*) as calculated from a pedigree. Here, we challenge this prevailing view. Because of chance events during Mendelian segregation, an individual's realized proportion of the genome that is identical by descent (IBD) may substantially deviate from the pedigree-based expectation (i.e. *F*). This Mendelian noise may result in a weak correlation between *F* and multi-locus heterozygosity, but this does not imply that multi-locus heterozygosity is a bad estimator of realized IBD. We examined correlations between 11 fitness-related traits measured in up to 1192 captive zebra finches and three measures of inbreeding: (i) heterozygosity across 11 microsatellite markers, (ii) heterozygosity across 1359 single-nucleotide polymorphism (SNP) markers and (iii) *F*, based on a 5th-generation pedigree. All 11 phenotypic traits showed positive relationships with measures of heterozygosity, especially traits that are most closely related to fitness. Remarkably, the small panel of microsatellite markers produced equally strong HFCs as the large panel of SNP markers. Both marker-based approaches produced stronger correlations with phenotypes than the pedigree-based *F*, and this did not seem to result from the shortness of our pedigree. We argue that a small panel of microsatellites with high allelic richness may better reflect an individual's realized IBD than previously appreciated, especially in species like the zebra finch, where much of the genome is inherited in large blocks that rarely experience cross-over during meiosis.

Keywords: identity by descent, inbreeding coefficient, inbreeding depression, pedigree, single-nucleotide polymorphism markers

Received 20 December 2011; revision received 13 March 2012; accepted 19 March 2012

Introduction

The rapid development and use of molecular markers has led to a surge of studies reporting correlations between individual-based measurements of heterozygosity across a panel of molecular markers (typically microsatellite markers) and phenotypic traits that are fitness-related or show condition-dependent trait expression (henceforth heterozygosity–fitness correlations, HFCs). Reviews of empirical studies have shown that HFCs on average are positive, but small, so that studies often lack the power to confidently estimate their size (Coltman & Slate 2003; Chapman *et al.* 2009). The weakness of these correlations has been ascribed to the fact that estimates of genome-wide heterozygosity that are based on only a small panel of microsatellite markers (typically 5–20, median 11, see Fig. 3a in Chapman *et al.* 2009) are only weakly correlated with the inbreeding coefficient *F*, calculated from a pedigree (Balloux *et al.*

Correspondence: Wolfgang Forstmeier, Fax: +49 8157 932400; E-mail: forstmeier@orn.mpg.de

[1]Current address: Department of Evolutionary Biology, Bielefeld University, Morgenbreede 45, 33619 Bielefeld, Germany.

2004; Slate *et al.* 2004; Pemberton 2004; Taylor *et al.* 2010; Santure *et al.* 2010).

The observation that correlations between heterozygosity and pedigree *F* are weak has also led to a reinterpretation that HFCs may not arise from genome-wide inbreeding but rather from some of the markers being in direct linkage disequilibrium with loci that show strong heterosis for fitness ('local effect hypothesis', e.g. Balloux *et al.* 2004; Pemberton 2004). This has ignited an ongoing debate about local vs. genome-wide effects (David 1998; Hansson & Westerberg 2002; Olano-Marin *et al.* 2011a,b). The observation of weak correlations between heterozygosity and pedigree *F* has also been used as an argument for obtaining more and better pedigree data (e.g. Pemberton 2004, 2008).

These interpretations seem to be based on the implicit assumption that pedigree-based values of individual *F* reflect the true level of inbreeding, while marker-based estimates may deviate from the true levels because of sampling noise. This is misleading, because an individual's realized inbreeding coefficient (i.e. the proportion of its genome that is identical by descent, IBD) depends on chance events during Mendelian segregation, which can be traced by molecular markers but cannot be captured by general pedigree information (Stam 1980; Leutenegger *et al.* 2003; Hill & Weir 2011). On the basis of pedigree data, for instance, all offspring of first cousins are expected to have 6.25% of their genome IBD (i.e. $F = 0.0625$). Simulations and empirical studies in humans, however, show that in a sample of 1000 such offspring, the true proportion of the genome that becomes IBD varies from about 1% to 16%, with a between-individual standard deviation of 2.4% (Leutenegger *et al.* 2003; Hill & Weir 2011). This individual deviation of realized IBD from pedigree *F* becomes relatively larger for more distant pedigree relationships (increasing coefficient of variance; Hill & Weir 2011), making the more distant pedigree connections particularly imprecise with regard to individual IBD. The segregation-induced amount of variation in realized IBD seems substantial when its magnitude is compared to the low population-wide variation in pedigree-based *F* that is seen in most wild study populations (Slate *et al.* 2004). In such typical wild populations, with small variance in *F*, the uncertainty with which pedigree *F* reflects IBD is therefore relatively large, while this uncertainty is less of an issue in populations where variance in *F* is large, that is, when close inbreeding is frequent.

The human genome, for which the above given numbers are valid, is not an extreme example: it is composed of 22 autosomes and has a total map length of about 37 Morgans (He *et al.* 2011), so it is inherited in many independently segregating units, which reduces the likelihood of extreme outcomes (in terms of the extent of IBD) during segregation. Variation in realized IBD may be even larger in species where a large proportion of the genome is inherited in fewer units. For instance, in the zebra finch (*Taeniopygia guttata*), about half of the autosomal genome is made up of four chromosomes (1, 1A, 2 and 3), and with the exception of regions close to the telomeres, cross-over rates are extremely low for the large interior parts of these chromosomes (16 times lower than in chicken; Backström *et al.* 2010). Thus, nearly a third of the genome typically segregates in only four independent blocks, and while the outcome of this process on an individual basis will often deviate from the pedigree prediction, it is conceivable that a set of only four microsatellite markers that are high in allelic richness and located anywhere in the central parts of each of these chromosomes would allow us to quantify to some extent such individual deviations from the expectation based on pedigree relationships. With its large linkage blocks, the zebra finch might represent a bit of a special case (where microsatellite markers might be especially useful for the study of inbreeding), but future linkage maps of other passerine species will still have to show how special or widespread this situation actually is.

Interestingly, a range of studies with high-quality heterozygosity and pedigree data have found that HFCs remain significant even after statistically controlling for the pedigree-based estimates of *F* (e.g. Bierne *et al.* 1998; Hansson *et al.* 2004; Markert *et al.* 2004; Bensch *et al.* 2006). This finding has often been interpreted as evidence for local heterosis, but it is also expected because of the inevitable noisiness of pedigree-based values of *F* and also supports the hypothesis that HFCs arise owing to genome-wide inbreeding. Although it has previously been recognized that chance events during segregation lead to deviations from the expectation based on *F* (e.g. Bensch *et al.* 2006; Santure *et al.* 2010), its role in explaining the weakness of correlations between heterozygosity and pedigree *F* has not been discussed as prominently as the alternative explanations (i.e. too few markers and local vs. general effects).

Santure *et al.* (2010) have recently addressed the power of microsatellite vs. single-nucleotide polymorphism (SNP) markers for estimating inbreeding in a captive zebra finch population with pedigree data available. They found that mean heterozygosity in a panel of 20 microsatellite markers only poorly reflected the pedigree-based values of *F* and that a panel of only 50 SNPs had an equal power in terms of reflecting pedigree *F*. From this, they concluded that heterozygosity of 20 microsatellites is not a useful proxy for the inbreeding coefficient and therefore of limited value for the study of inbreeding depression. However, their analyses focus on the wrong target, namely on pedigree *F* rather than

on IBD, so the findings may only reflect the inappropriateness of equating pedigree *F* with inbreeding. Hence, it still remains to be seen which marker set has more power in terms of reflecting realized IBD and hence inbreeding depression.

Here, we directly compare the power of all three approaches (microsatellite markers, SNP markers and pedigree-based *F*) in terms of detecting inbreeding depression for phenotypic traits in a large captive population of zebra finches. We examined various measures of heterozygosity using a panel of 11 microsatellite markers (located on nine autosomes) and a panel of 1359 SNP markers (located on 31 autosomes), in combination with data from a pedigree spanning five consecutive generations. As phenotypic traits, we examined a combination of morphological, ornamental, behavioural and fitness traits, all of which showed inbreeding depression (Bolund *et al.* 2010) or have been suggested to be condition dependent or directly related to fitness. Theory predicts that traits closely related to fitness should show the strongest HFCs, although the empirical support for this is still limited (DeRose & Roff 1999; Chapman *et al.* 2009).

Materials and methods

Subjects and inbreeding history

We studied a large population of domesticated zebra finches maintained since 2004 at Seewiesen, Germany (population #18 described in Forstmeier *et al.* 2007b). The founders came from a population maintained since 1985 at the University of Sheffield, which has been studied extensively since then (e.g. Santure *et al.* 2010).

The phenotypic and genetic data covered four generations of birds (comprising 1209 individuals). The ancestors of the first generation were also known, so that the pedigree information covers five generations (1378 individuals). In the last, the fifth, generation, we produced 47 individuals from full-sib matings alongside outbred half-sibs. Using a direct comparison between inbred and outbred individuals in the last generation, we have already published estimates of inbreeding depression for several of the phenotypic traits studied here (Bolund *et al.* 2010). To assess the extent to which the findings presented here depend on the inclusion of these 47 highly inbred individuals, we present most analyses in two versions, once including ('all individuals') and once excluding these 47 inbred birds (referred to as '$F < 0.15$'). This also allows us to evaluate to what extent the findings depend on the magnitude of population-wide variance in *F*.

The numbers of birds in the various categories of pedigree-based *F* are shown in Table 1. These values underestimate the true values of *F* and IBD, because the founders that came from Sheffield are known to have an average *F* of 0.030 ± 0.014 SD (Forstmeier *et al.* 2004), based on an extended 18th-generation pedigree from Sheffield. These additional 3% of the genome that are expected to be IBD (which could roughly be added to each value in Table 1) again may be an underestimate of true IBD, depending on the relatedness between the founders of the Sheffield pedigree. In our 5th-generation pedigree, variance in *F* is 0.00016 (excluding the highly inbred birds), which is close to the value of 0.0002 reported by Santure *et al.* (2010) for the Sheffield zebra finch population (their 20th-generation pedigree), and which is relatively low compared to some wild populations (median = 0.0023; Slate *et al.* 2004). When including the 47 highly inbred individuals, variance in *F* rises to 0.0024 (our 5th-generation pedigree), which is close to the median from the wild populations.

Table 1 Inbreeding coefficients (*F*) of the birds in our 5th-generation pedigree

F	Count of individuals
0	1100
0.01563	24
0.03125	6
0.0625	29
0.125	4
0.25	47

Genotypes

All 1209 individuals were genotyped for 11 microsatellite markers (Table S1, Supporting information). These are located on nine different autosomes (1, 1A, 2, 3, 4A, 5, 6, 9 and 15), with the two largest chromosomes (1 and 2) being covered by two markers each (located 78 and 66 Mb apart from each other, respectively). These markers were selected from a set of 19 loci in total, of which three were excluded because of close linkage (resulting from a special design; see Forstmeier *et al.* 2010) and five were excluded because of evidence for null alleles segregating within our pedigree (null alleles make it impossible to distinguish between homozygotes and heterozygotes carrying a null allele). The here-included ARmicro1 marker also has a null allele, but this is easily recognized because of close linkage with another marker (Forstmeier *et al.* 2010), so it was coded like a regular allele (allele frequency = 1.7%). The 11 selected markers were all tested for heterozygote deficiency using GENEPOP (Raymond & Rousset 1995), which yielded $P > 0.077$ for every marker before Bonferroni

correction. All inheritance errors within our pedigree (about 2–3% of all allele calls) were checked by re-typing and could be identified as calling or genotyping errors (and were corrected) with the exception of five events of mutation (all in ARmicro1; 0.25% mutation rate for this locus). In these five cases, all descendants were assigned to the original parental allele to correctly reflect potential cases of IBD. It is important to note that the 11 used microsatellite markers were not preselected in any way in terms of heterozygosity, and their location in the genome is the outcome of an almost random sampling process (Table S1, Supporting information, Forstmeier *et al.* 2007a, 2010). Hence, any other random set of microsatellite markers in this species should be just as informative. We also tested whether HFCs were driven by one or few of our 11 loci (local effect hypothesis), but this was clearly not the case (details not shown, raw data available).

A subset of 1067 individuals (excluding birds of the second generation that had not been given an opportunity to reproduce) were also genotyped for 1359 polymorphic SNP loci that were included in a linkage map for our population (for further details, see Backström *et al.* 2010, excluding markers on TguZ).

Measures of inbreeding and genetic diversity

For populations with extensive linkage disequilibrium, it is statistically challenging to obtain an accurate estimate of the proportion of an individual's genome that is IBD, and any such estimate would come with an amount of error that is even more difficult to estimate (Leutenegger *et al.* 2003; Hill & Weir 2011). We therefore did not attempt such estimation, but focused on a range of simple and widely used measures of inbreeding or heterozygosity, and examined their ability to predict inbreeding depression for phenotypic traits.

We used one pedigree-based estimate of inbreeding (F), four microsatellite-based estimates of heterozygosity, four SNP-based estimates of heterozygosity and one joint measure of inbreeding (a principal component), as follows:

1 Inbreeding coefficient (F): calculated from our 5th-generation pedigree using Pedigree Viewer 6.4a (Kinghorn & Kinghorn 2010). Data are shown in Table 1.
2 Multi-locus heterozygosity based on microsatellites (MLH_{ms}): number of heterozygous loci divided by the number of genotyped loci (allele frequencies are ignored). The population-wide mean MLH_{ms} was 0.808. We did not calculate standardized MLH_{ms} (Coltman *et al.* 1999), because with complete genotype data (as in our case), the two measures are perfectly correlated.
3 Internal relatedness based on microsatellites (IR_{ms}) gives more weight to rare alleles (Amos *et al.* 2001).
4 Homozygosity by locus based on microsatellites (HL_{ms}): sum of the probability of being heterozygous at homozygous loci (based on population allele frequencies) and the total probability of being heterozygous for all alleles in a given genotype (Aparicio *et al.* 2006).
5 Mean d^2 based on microsatellites (d^2_{ms}): mean squared difference in microsatellite allele length of the two alleles, meant to capture more distant inbreeding (Coulson *et al.* 1998). For the few individuals that were carrying a known null allele at the ARmicro1 locus, we excluded this locus when calculating d^2.
6 Multi-locus heterozygosity based on SNPs (MLH_{snp}): see (2). The population-wide mean MLH_{snp} was 0.302.
7 Standardized multi-locus heterozygosity based on SNPs ($sMLH_{snp}$): mean heterozygosity across all typed loci divided by the mean heterozygosity at typed loci in the population (Coltman *et al.* 1999).
8 Internal relatedness based on SNPs (IR_{snp}): see (3).
9 Homozygosity by locus based on SNPs (HL_{snp}): see (4).
10 Principal component 1 (PC1): a joint measure of inbreeding based on F, IR_{ms} and IR_{snp}.

Because some of these measures quantify heterozygosity and some homozygosity, we adjusted the sign so that higher values indicate increased within-individual diversity (i.e. measures 1, 3, 4, 8 and 9 were multiplied by −1, which we subsequently refer to as 'inverted'). For the calculation of IR and HL measures, allele frequencies were calculated from the $F < 0.15$ subset.

To explore the usefulness of SNP panels of varying sizes (in terms of reflecting inbreeding depression), we randomly sampled 50, 100 and 500 SNPs, respectively. From these, we calculated IR_{snp} and examined the strength of correlation with each phenotypic trait (HFC). The random sampling of SNPs was repeated 1000 times, and the average HFC across the 1000 replicates was used for further analysis.

To estimate the repeatability of marker-based heterozygosities, we randomly assigned markers to two groups of about equal size (as equal as possible; e.g. five vs. six microsatellites), and this was carried out multiple times (1000× for microsatellites and 200× for SNPs). We calculated the respective measures in each iteration for each group and calculated the correlation between the two groups (heterozygosity–heterozygosity correlations, HHC). This Monte Carlo simulation ignored the nonindependence of SNP markers located on the same chromosome. As many SNP markers (and two pairs of microsatellite markers) show strong linkage

(Backström *et al.* 2010), we also applied a simulation where we split SNP markers (and microsatellite markers) into two about equally sized groups, but instead of sampling them randomly, we sampled them in blocks that reflect the autosomes. For this, all the markers from an autosome were allocated to one group as a single unit, with each group consisting of a different set of autosomes. All simulations and calculations were run in R 2.13.1 (R Development Core Team 2011).

Phenotypic traits

We analysed a total of 11 traits (four traits measurable in both of the sexes, four female traits and three male traits). We here briefly describe the data collection and the number of individuals that were phenotyped for a particular trait.

1 Tarsus length: measured on adult birds to the nearest 0.1 mm using a ruler. We adjusted for a small, but statistically significant sexual dimorphism in tarsus length (here and below expressed as the male trait value minus the female trait value: $b = 0.07$ mm ± 0.03 (SE), $t_{1194} = 2.16$, $P = 0.031$) by centring the data within sexes (total $N = 1196$ individuals).
2 Mass adult: measured on a digital scale to the nearest 0.1 g at multiple occasions during nonbreeding conditions. Repeated measures of the same individual (2.9 ± 2.1, mean ± SD) were averaged. We adjusted for a clear sexual dimorphism in adult weight ($b = -0.81$ g ± 0.10 (SE), $t_{1182} = -8.0$, $P = 3*10^{-15}$) by centring the data within sexes ($N = 1184$ individuals). Mass data were cube-root transformed to yield the same dimensionality as length measures.
3 Mass day 8: chicks were measured on a digital scale to the nearest 0.1 g at day 8 posthatch. We adjusted for sexual dimorphism in mass at day 8 ($b = -0.29$ g ± 0.11 (SE), $t_{972} = -2.7$, $P = 0.007$) by centring the data within sexes ($N = 974$ individuals).
4 Beak colour: spectrometric measurements combined into the main axis of sexual dimorphism by discriminant function analysis (see Schielzeth *et al.* 2012 for details). Repeated measures of the same individual (1.8 ± 0.8, mean ± SD) were averaged. We adjusted for strong sexual dimorphism (the discriminant function maximized the difference between the sexes) by centring the data within sexes ($N = 1017$ individuals).
5 Choice activity: female hopping activity in a four-way mate choice chamber (see Schielzeth *et al.* 2010). This trait was included because it was recently suggested to be a condition-dependent quality indicator (Woodgate *et al.* 2010). Data were square-root transformed to approach normality. Repeated measures of the same individual (1.5 ± 0.8, mean ± SD) were averaged ($N = 552$ females).
6 Egg size: egg width and length of the first six eggs (maximally) in each clutch were measured using callipers and converted to a measure of egg volume using the formula: volume = $0.5236 \times \text{length} \times \text{width}^2$. Repeated measures of the same individual were averaged (26.1 ± 19.4 eggs, mean ± SD; $N = 437$ females).
7 Female cage fecundity: number of eggs laid in cages by monogamously paired females divided by the number of weeks paired. Repeated measures of the same individual (2.0 ± 1.0 pairings, mean ± SD) were averaged ($N = 270$ females).
8 Female aviary fecundity: number of eggs laid under aviary conditions (as described in Forstmeier *et al.* 2011), standardized within aviaries to a mean of unity (as commonly done to obtain 'relative fitness'). Repeated measures of the same individual were averaged (117 individuals with one observation and 54 with two observations; $N = 171$ females).
9 Courtship rate: measurement of the duration (in sec) of directed song in staged male–female encounters of bachelor males (as described in Forstmeier *et al.* 2011; $N = 583$ males). Courtship rate was square-root transformed to approach normality.
10 Male attractiveness in a choice chamber: average relative time allocation of four females (tested consecutively) in a choice chamber set-up with four stimulus males (see Schielzeth *et al.* 2010 for details on choice chamber setups; $N = 582$ males).
11 Male siring success in aviaries: number of eggs fertilized under aviary conditions and standardized within aviaries to relative fertilization success (see Forstmeier *et al.* 2011). Repeated measures of the same individual were averaged (102 individuals with one observation and 62 with two observations; $N = 164$ males).

For the purpose of this study, we consider these 11 phenotypic traits as independent measures for revealing inbreeding depression. Table S2 (Supporting information) shows correlation coefficients among these 11 traits: the average correlation was only $r = 0.13$, so the actual nonindependence of some of these traits will introduce only negligible amounts of pseudoreplication in our analyses.

Effects of pedigree shortening

For an average individual in our 5th-generation pedigree, only 2.5 (±0.93 SD) generations of ancestors are known, leading to an estimate of $F = 0$ for 90.9% of all individuals (see Table 1). To examine the extent to which this limits our ability to detect correlations

between pedigree-based F and inbreeding depression, we analysed an additional set of data. In our population, a 6th and 7th generations have been bred, but for these, no genotypic data are available, and phenotypic data are currently limited to tarsus length, adult mass, mass at day 8 and courtship rate (traits 1–3 and 9 listed above). These birds represent the first two offspring generations of lines that were selected for courtship rate. There are six lines in total (made up of 15 breeding pairs each), of which two lines were selected for high courtship rate, two lines were selected for low courtship rate, and two lines were paired randomly as controls. We found no significant line differences in the phenotypic traits 1–3, so we here focus on these traits, but we do not analyse courtship rates, which strongly differ between the lines. Phenotypic traits 1–3 were available for each of 1003 individuals, which we analyse jointly (rather than as six different groups). In this data set, pedigree information spans 6–7 generations, and inbreeding is relatively high (see Fig. S1, Supporting information). Only 39.8% of the individuals are estimated as $F = 0$. Average $F = 0.0357$, and variance in $F = 0.00257$. For this data set, we first examine the correlation between F and phenotypic traits. We then stepwise shorten the pedigree by removing the upmost generation until left with a 3rd-generation pedigree (spanning the 5th to 7th generations). At each step, we calculate F and examine the correlations with phenotypes.

Results

Inheritance of large blocks of IBD

On the basis of the SNP data, Fig. 1 illustrates spatial patterns of homozygosity vs. heterozygosity across the genome for 90 individuals from the fifth generation of our pedigree. Half of the individuals are highly inbred ($F = 0.25$), resulting from brother–sister matings, and the other half is outbred ($F = 0$) according to five generations of pedigree information. Long uninterrupted stretches of homozygosity (in black) are frequently seen in the inbred birds, often covering nearly the full length of a chromosome. Such long stretches are strong indications of identity by descent (IBD) because it is very unlikely that so many SNPs in a row are all homozygous by chance alone. Intriguingly, long stretches of homozygosity can also be found in some of the outbred individuals, which is indicative of inbreeding loops reaching five or more generations back. All of these

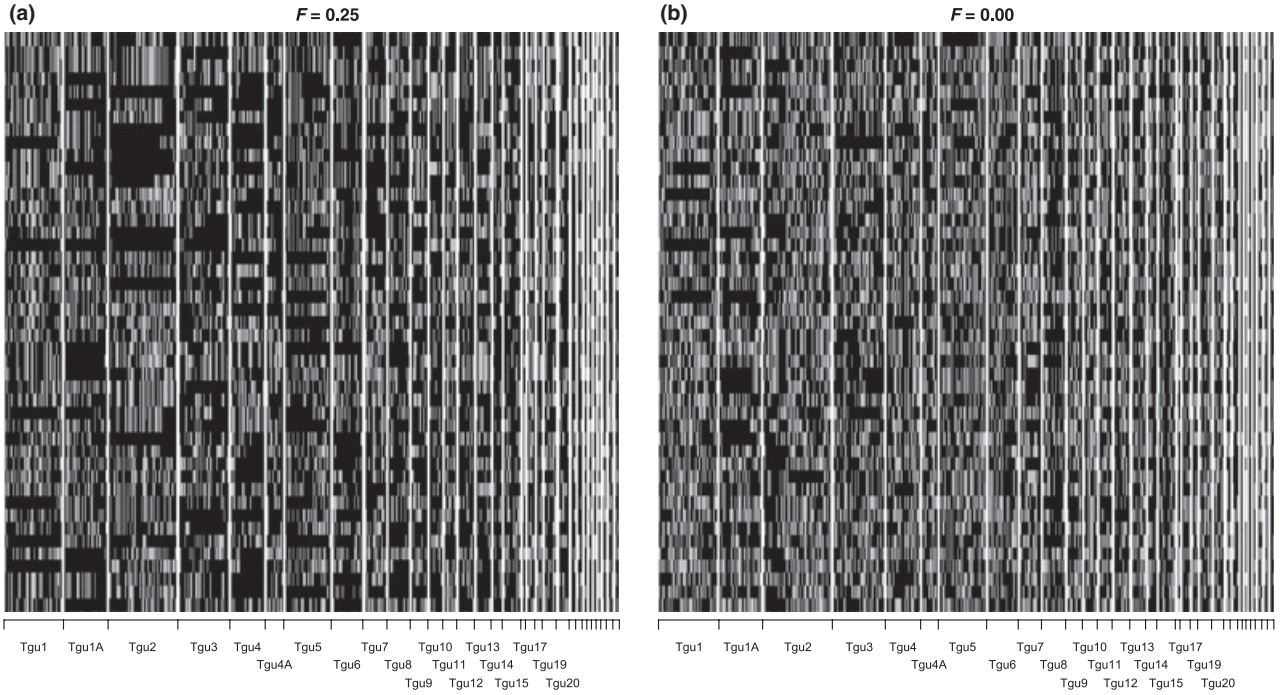

Fig. 1 Homozygosity across the genome for (a) 45 inbred ($F = 0.25$) and (b) a sample of 45 outbred individuals ($F = 0$) from the fifth generation of our pedigree. Each row represents an individual. Markers are ordered along the x-axis according to their position on the genetic map, with macro-chromosomes labelled by their name and micro-chromosomes ordered as in Backström *et al.* (2010). Heterozygous loci are shown in grey, homozygous loci in black and a few missing values in white.

long loops must span ten or more events of meiosis (five or more meioses on both the maternal and the paternal sides of inheritance) in which a block could have been broken up. So, it is remarkable how the scarcity of cross-overs in the interior parts of the large chromosomes (Backström *et al.* 2010) has preserved the integrity of these large blocks.

Heterozygosity–heterozygosity correlations

Table 2 shows the strength of heterozygosity–heterozygosity correlations (HHC) when splitting markers into two subsets of about equal size. When considering all birds (including $F = 0.25$) and the method of random sampling of markers, HHCs for SNPs ($r = 0.68$–0.75) are much higher than those for microsatellite measures ($r = 0.11$–0.13), with d^2 in microsatellites being the least repeatable measure ($r = 0.02$). However, when sampling the markers in a blocked design (markers from half of the chromosomes in one subset, and half of the chromosomes in the other subset), SNP marker HHCs are substantially lower ($r = 0.15$–0.18) and more similar to the HHCs for blocked microsatellite measures ($r = 0.08$). Hence, the high HHCs for SNP markers in the random sampling design follow from spatial heterogeneity in heterozygosity (see Fig. 1), that is, from the nonindependence of SNP markers located on the same chromosome. In contrast, in the blocked design, nonzero HHCs between chromosomes can only arise through individual differences in inbreeding. Accordingly, when restricting the variation in inbreeding (by excluding the $F = 0.25$ individuals), HHCs become lower for both microsatellites ($r = 0.05$) and SNPs ($r = 0.04$–0.05). Among the various measures of heterozygosity, internal relatedness (IR) showed the highest HHCs for both microsatellites and SNPs.

Correlations between pedigree- and marker-based estimates of inbreeding

Table 3 shows pairwise correlations between pedigree-based F and the different measures of heterozygosity in microsatellites and in SNPs. The strongest correlations between these three classes of measures involve measures of IR, further confirming the usefulness of this metric. When focusing on all individuals (high variance in inbreeding), IR in SNP markers was slightly better predicted by IR in microsatellites ($r = 0.59$) than by the pedigree-based F ($r = 0.49$). This difference became more pronounced when restricting the variance in inbreeding by excluding the highly inbred individuals. In this subset ($F < 0.15$), IR in SNPs was still predicted by IR in microsatellites ($r = 0.53$), but not by pedigree-based F ($r = 0.08$). The strength of the SNP–microsatellite correlation shows that there was true variation in the extent of IBD, which was not captured by the pedigree information. Table 3 also shows that d^2_{ms} was the worst microsatellite-based metric in terms of reflecting heterozygosity in SNP markers, which makes sense in the light of its low repeatability (see Table 2).

Table 2 Heterozygosity–heterozygosity correlations (HHC) when splitting markers into two about equal-sized subsets. Sampling was carried out repeatedly ($N = 1000$ for microsatellites, $N = 200$ for SNPs) at random (random sampling) or in spatial blocks keeping markers from the same chromosome in the same subset (blocked sampling). Results are shown once for all individuals and once when excluding 47 highly inbred birds (all $F < 0.15$). Standard deviations (SD) refer to the variation among the 200 or 1000 replicates

	All birds				$F < 0.15$			
	Random sampling		Blocked sampling		Random sampling		Blocked sampling	
	HHC	SD	HHC	SD	HHC	SD	HHC	SD
MLH_{ms}	0.1133	0.0379	0.0773	0.0221	0.0782	0.0368	0.0503	0.0234
IR_{ms}	0.1271	0.0404	0.0850	0.0209	0.0866	0.0394	0.0528	0.0218
HL_{ms}	0.1155	0.0385	0.0784	0.0216	0.0804	0.0375	0.0511	0.0227
d^2_{ms}	0.0189	0.0283	0.0118	0.0263	0.0028	0.0277	−0.0055	0.0258
MLH_{snp}	0.6839	0.0248	0.1555	0.0316	0.6082	0.0262	0.0383	0.0321
$sMLH_{snp}$	0.6856	0.0243	0.1538	0.0317	0.6094	0.0263	0.0355	0.0324
IR_{snp}	0.7478	0.0162	0.1780	0.0302	0.6731	0.0211	0.0435	0.0338
HL_{snp}	0.7017	0.0218	0.1635	0.0290	0.6332	0.0241	0.0538	0.0316

HL_{snp}, homozygosity by locus based on SNPs; HL_{ms}, homozygosity by locus based on microsatellites; IR_{ms}, internal relatedness based on microsatellites; IR_{snp}, internal relatedness based on SNPs; MLH_{ms}, multi-locus heterozygosity based on microsatellites; MLH_{snp}, multi-locus heterozygosity based on SNPs; $sMLH_{snp}$, standardized multi-locus heterozygosity based on SNPs; d^2_{ms}, mean d squared based on microsatellites.

Table 3 Correlations between pedigree-based and marker-based estimates of inbreeding. Correlation coefficients greater than 0.5 are shown in bold. Below the diagonal: all individuals included; above the diagonal: 47 inbred individuals excluded (all $F < 0.15$). The measures *F*, IR and HL were inverted such that all correlations are expected to be positive.

Source	Statistic	*F*	MLH_{ms}	IR_{ms}	HL_{ms}	d^2_{ms}	MLH_{snp}	$sMLH_{snp}$	IR_{snp}	HL_{snp}
Pedigree	*F*	**1**	−0.04	−0.03	−0.04	0.02	0.1	0.1	0.08	0.08
MS	MLH_{ms}	0.21	**1**	**0.98**	**1**	**0.58**	0.45	0.45	0.49	0.45
MS	IR_{ms}	0.24	**0.98**	**1**	**0.98**	**0.57**	0.48	0.48	**0.53**	0.5
MS	HL_{ms}	0.21	**1**	**0.98**	**1**	**0.58**	0.46	0.46	0.5	0.46
MS	d^2_{ms}	0.17	**0.6**	**0.6**	**0.6**	**1**	0.34	0.34	0.35	0.34
SNP	MLH_{snp}	0.46	**0.52**	**0.55**	**0.52**	0.39	**1**	**1**	**0.97**	**0.97**
SNP	$sMLH_{snp}$	0.46	**0.52**	**0.55**	**0.52**	0.39	**1**	**1**	**0.97**	**0.97**
SNP	IR_{snp}	0.49	**0.55**	**0.59**	**0.56**	0.4	**0.98**	**0.98**	**1**	**0.98**
SNP	HL_{snp}	0.45	**0.52**	**0.56**	**0.52**	0.38	**0.98**	**0.98**	**0.98**	**1**

HL_{snp}, homozygosity by locus based on SNPs; HL_{ms}, homozygosity by locus based on microsatellites; IR_{ms}, internal relatedness based on microsatellites; IR_{snp}, internal relatedness based on SNPs; MS, microsatellites; SNP, single-nucleotide polymorphism; MLH_{ms}, multi-locus heterozygosity based on microsatellites; MLH_{snp}, multi-locus heterozygosity based on SNPs; $sMLH_{snp}$, standardized multi-locus heterozygosity based on SNPs; d^2_{ms}, mean d squared based on microsatellites.

Which markers produce the strongest HFCs?

When focusing on all individuals, 108 of 110 correlations between the 11 phenotypic traits and the 10 measures of heterozygosity were positive, with 65 correlations being significant (Table S3, Supporting information). When excluding the highly inbred birds, 103 of 110 correlations were positive, with 31 of them significant. The nine negative correlations (against the prediction) involved either the inverted pedigree-based *F* (one significant and six nonsignificant correlations) or d^2_{ms} (two nonsignificant correlations).

Figure 2 shows the average strength of HFCs by marker type, when averaging across the 11 phenotypic traits. With all individuals included, the strongest HFCs were produced by IR_{snp} ($r = 0.118$) among the SNP-based measures, followed by IR_{ms} ($r = 0.103$) among the microsatellite-based measures, followed by pedigree-based *F* ($r = 0.082$). A principal component composed of IR_{snp}, IR_{ms} and *F* showed the strongest HFCs ($r = 0.131$). Yet, most of these differences in the strength of HFCs were nonsignificant (e.g. paired *t*-test for IR_{snp} vs. *F*: $t_{10} = 1.88$, $P = 0.089$). When excluding the highly inbred birds, HFCs generally became lower, with IR_{ms} ($r = 0.082$) slightly exceeding IR_{snp} ($r = 0.076$), while the HFCs based on *F* (inverted) were practically zero ($r = -0.01$). In this case, IR_{ms} produced significantly stronger HFCs than *F* (paired *t*-test, $t_{10} = 2.53$, $P = 0.03$).

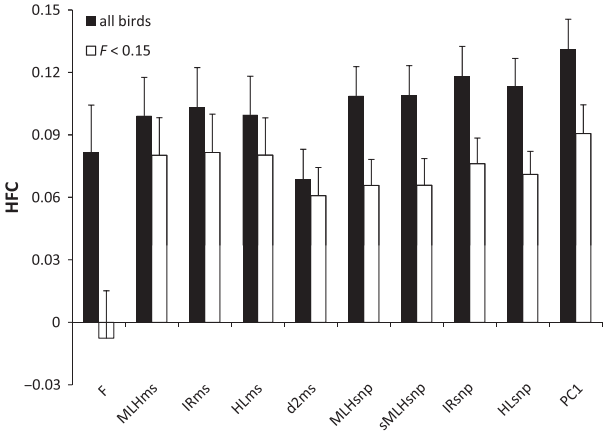

Fig. 2 Average strength of heterozygosity–fitness correlations (HFC) across all 11 phenotypes for the various inbreeding measures: pedigree-based inbreeding coefficient *F* (inverted to produce positive HFCs), microsatellite-based measures of heterozygosity (ms), single-nucleotide polymorphism-based measures (snp) and a principal component (PC1) summarizing all three types of measures. Multi-locus heterozygosity (MLH), standardized multi-locus heterozygosity (sMLH), internal relatedness (inverted; IR) and heterozygosity by locus (HL). Correlations for all birds in the data set are stronger than when excluding highly inbred individuals (all $F < 0.15$). Standard errors were calculated across the 11 phenotypic traits (assuming they are independent; see Table S2, Supporting information).

How many SNPs are needed?

Figure 3 illustrates how the power of SNP marker panels (in terms of producing HFCs) increases with marker numbers (see Table S4, Supporting information). When considering the data set with all birds included, average HFCs increased with marker numbers from $r = 0.052$ (50 SNPs) to $r = 0.069$ (100 SNPs), $r = 0.105$ (500 SNPs) and $r = 0.118$ (1359 SNPs). In comparison, the 11 microsatellites yielded $r = 0.103$, which corresponds to the HFC that would be obtained with

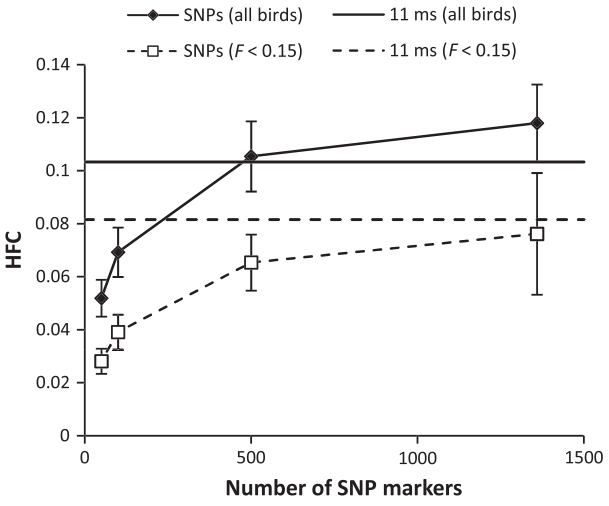

Fig. 3 Average strength of heterozygosity–fitness correlations (HFC) across all 11 phenotypes when reducing the number of single-nucleotide polymorphism (SNPs) used. Internal relatedness (IR) was calculated either from all 1359 SNPs or from 500, 100 or 50 randomly selected SNPs, respectively. Random selections were carried out 1000 times, and average HFCs across the 1000 repetitions were used. Standard errors refer to the variation in HFCs across the 11 phenotypic traits (assuming they are independent; see Table S2, Supporting information). HFCs for IR based on 11 microsatellites (11 ms) are shown for comparison. Correlations for all birds in the data set are stronger than when excluding highly inbred individuals (subset of $F < 0.15$).

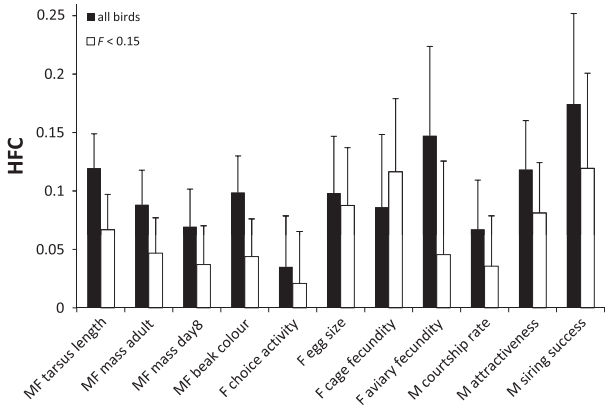

Fig. 4 Average strength of heterozygosity–fitness correlations (HFC) across all nine measures of inbreeding for the various phenotypes measured in both sexes (MF), females only (*F*) or males only (M). Correlations for all birds (including the highly inbred ones) are stronger than those where highly inbred individuals are excluded (subset of $F < 0.15$). Standard errors were calculated for each average correlation coefficient based on the average sample size available according to Sokal & Rohlf (1995, p. 574).

about 450–500 SNPs. For the $F < 0.15$ data set, average HFCs increased with marker numbers from $r = 0.028$ (50 SNPs) to $r = 0.039$ (100 SNPs), $r = 0.065$ (500 SNPs) and $r = 0.076$ (1359 SNPs). In comparison, the 11 microsatellites yielded $r = 0.082$, which corresponds to the HFCs that would be obtained with more than 1500 SNPs (extrapolated).

Effects of pedigree shortening

In the additional set of birds from the 6th and 7th generation of the pedigree, inbreeding coefficients were highly variable (variance (F) = 0.00257) and were fairly evenly distributed across categories (Fig. S1, Supporting information). Stepwise shortening of the pedigree down to three generations reduced variance in F to 0.00085 (Fig. S2, Supporting information) and strongly increased the proportion of birds with $F = 0$ (Fig. S1, Supporting information). On the basis of seven generations of pedigree information, F (inverted) showed an average correlation with morphological phenotypes of $r = 0.120$ (Table S5, Supporting information), which was comparable to the estimates obtained for the main data set from the generations two to five (all birds: $r = 0.139$; subset with $F < 0.15$: $r = 0.069$; Table S5, Fig. S2, Supporting information). Stepwise shortening of the pedigree reduced this correlation initially only slightly (to $r = 0.117$), but eventually down to $r = 0.078$. As expected, the strength of HFCs decreased with decreasing variance in F (Fig. S2, Supporting information).

Which phenotypic traits show the strongest HFCs?

Figure 4 shows the average strength of HFCs by phenotypic trait when averaged across the nine heterozygosity measures (excluding PC1). HFCs are significantly stronger when based on all individuals ($r = 0.100$) than when restricted to the less inbred birds ($r = 0.064$; paired t-test, $t_{10} = 3.66$, $P = 0.004$). When averaging across these two approaches, the strongest HFCs for males were found for siring success in aviaries ($r = 0.147$) followed by attractiveness in a choice chamber ($r = 0.100$). For females, the strongest HFCs involved fecundity in cage breeding ($r = 0.101$) and fecundity in aviary breeding ($r = 0.096$). While this is in line with the expectation that traits closely related to fitness show the strongest HFCs, the large standard errors around these estimates preclude a decisive interpretation (see Fig. 4).

Discussion

In our population of moderately inbred zebra finches, we found that all 11 phenotypic traits showed positive correlations with measures of genome-wide heterozygosity. The average magnitude of these correlations was

fairly small (around $r = 0.1$), which, however, is still larger than the average (around $r = 0.05$) across other animal species (Chapman *et al.* 2009). This somewhat larger effect size may be due to the zebra finch's evolutionary history (large population size, lack of inbreeding; Zann 1996; Forstmeier *et al.* 2007b), which allows recessive deleterious mutations to accumulate, leading to relatively strong inbreeding depression (Bolund *et al.* 2010). By averaging across 11 phenotypic traits, we obtained relatively robust estimates of inbreeding depression. These estimates allowed us to compare various molecular and pedigree-based methods with regard to their ability to reflect inbreeding depression.

We found that a panel of only 11 microsatellite markers produced about equally strong HFCs as a large panel of >1300 SNP markers. When looking at the full data set ('all birds'; where variance in inbreeding was high), HFCs based on our 11 microsatellite markers were about equally strong as those produced by 450–500 SNP markers. When excluding highly inbred birds from this data set (small variance in inbreeding), the microsatellites even outperformed the full set of 1359 SNP markers (Fig. 3). This finding is in strong contrast to earlier predictions that the power of 20 microsatellite markers should be about equal to the power of 50 SNP markers (Santure *et al.* 2010; Smouse 2010). The discrepancy between this prediction and our finding arises from the fact that the prediction was based on the ability of markers to reflect pedigree-based values of *F*. Inbreeding depression, however, results from IBD, and an individual's realized proportion of the genome that is IBD may strongly deviate from the pedigree-based prediction (Leutenegger *et al.* 2003; Hill & Weir 2011). Individual deviations of IBD from *F* will be particularly pronounced in species where much of the genome is inherited in only a few large blocks (see Fig. 1), such that chance events during Mendelian segregation induce major deviations from the average expectation. This effect is especially pronounced in the analyses presented by Santure *et al.* (2010). The 20 microsatellites used in their study were distributed over only four (1, 1A, 2 and 9) of the 32 zebra finch chromosomes. Given the low cross-over rates in the interior of chromosomes 1, 1A and 2, it is conceivable that markers on the same chromosome would often indicate the same state (either all heterozygous or all homozygous), which would lead to large deviations from the pedigree-based probabilities. Thus, the lack of a strong correlation between the microsatellite-based heterozygosity estimate and *F* may reflect the inappropriateness of pedigree-based values of *F*, rather than the inappropriateness of molecular markers for studies of inbreeding depression.

Still, it may seem surprising that a few microsatellite markers have equal power as such a large number of SNP markers (see Fig. 3). This may be explained by the fact that the number of alleles at a microsatellite marker (mean = 11, range = 7–18; see Table S1, Supporting information) is not much lower than the number of unique haplotypes (for a larger genomic region) that is segregating within our population. For instance, using three microsatellite markers located within the oestrogen receptor gene ESR1, we identified 13 different haplotypes for this genomic region in our population (Forstmeier *et al.* 2010), and further sequencing efforts (covering a total of 63 SNPs or indels) revealed an additional four haplotypes (i.e. 17 types in total; unpublished data). Approximately 10% of the individuals carry two copies of the same haplotype, so they seem to be IBD for this region. The three microsatellite markers that are located in this region have an average number of seven alleles, with an average of 22.5% of individuals being homozygote for a given microsatellite marker (i.e. they are identical by state, IBS). Hence, in a bit less than half of the cases, IBS will actually be due to IBD. In contrast, at a typical SNP marker, about 70% of the individuals will be homozygote (IBS), so in the great majority of cases, IBS does not reflect IBD. Individual SNP markers would only become highly informative regarding IBD when the number of haplotypes segregating within the population would approach two.

In our data set, marker-based estimates of heterozygosity tended to be more informative about inbreeding depression than pedigree-based estimates of *F*. This begs the question about the quality of our pedigree information. For our main data set (generations 2–5), the average number of generations for which pedigree information was available was 2.5, which corresponds to about 10 known ancestors. This pedigree information appears to be free of errors, because no pedigree error was detected by the extensive SNP genotyping (Backström *et al.* 2010). The shortness of our pedigree is certainly not ideal for estimating *F*, but it might still be representative for studies of such wild populations where a considerable proportion of animals are immigrants coming from an unknown background, which leads to a limited depth of the informative pedigree. When we included the highly inbred individuals ($F = 0.25$) that were generated in the fifth generation of our pedigree, variance in *F* was rather high (0.0024), and *F* produced correlations with phenotypes that were only slightly (and nonsignificantly) lower than marker-based HFCs (Fig. 2). When excluding these individuals, variance in *F* was reduced to 0.00016. Note that this low variance in *F* is not because of the shortness of the pedigree *per se*, because Santure *et al.* (2010) obtained a quite similar variance in *F* (0.0002) from their 20th-generation pedigree. Hence, it is more the absence of individuals with high inbreeding coefficients than the

large proportion of individuals with $F = 0$ that leads to such low variance in F. With this low variance in F, the average correlation with phenotypes was effectively zero (Fig. 2), but slightly higher values were obtained for the four phenotypes that were measured in both sexes with resulting higher statistical power (average $r = 0.053$; see Table S3, Supporting information).

Analysing a second data set from generations 6–7 revealed that correlations between F and phenotypes remain quite modest (Fig. S2, Supporting information) even when variance in F is high and pedigree information is very detailed (Fig. S1, Supporting information). Despite the fact that in this additional data set 60.2% of the individuals had an $F > 0$, the average correlation between F and phenotypes was not higher ($r = 0.120$) than in the main data set ($r = 0.139$), where only 9.1% of individuals had an $F > 0$. Systematic shortening of the pedigree led to somewhat weaker correlations between F and phenotypes, yet much of this effect could be explained by the reduced variance in F in such shortened pedigrees (Fig. S2, Supporting information). Given that these results are based on only three phenotypic traits, a cautious interpretation may be necessary, but the results may suggest that HFCs were not primarily limited by the length of our pedigree. This is in line with theoretical expectations. Keller *et al.* (2011) showed that in a randomly mating population, very little (<1%) of the variation in pedigree F is missed by ignoring inbreeding resulting from common ancestors more than five generations ago (see also Derrida *et al.* 2000). Moreover, the pedigree estimates of IBD resulting from such long loops come with a large uncertainty (in terms of individual deviations from the expectation; Hill & Weir 2011). As a consequence, longer pedigrees (than about five generations) are not expected to solve the issue that pedigree F gives an imprecise estimate of individual IBD.

Using a high density of SNP markers illustrates (Fig. 1) that Mendelian segregation introduces much between-individual variation in realized IBD, especially in systems like the zebra finch where large blocks are inherited without cross-over. In such systems, it seems plausible that a small number of microsatellites could reflect with some accuracy IBD because of recent inbreeding (left side, $F = 0.25$), but also inbreeding events that reach more than 5 generations back (right side, $F = 0$). The fact that a few highly diverse microsatellites can effectively capture the presence of such large IBD blocks, even when variance in F is low, can also be seen from the very high correlation between IR_{ms} and IR_{snp} ($r = 0.53$ for the $F < 0.15$ subset). In contrast, for the same set of birds, pedigree-based F failed to reflect IR_{snp} ($r = 0.08$). Note that in populations with low variance in F, where most individuals carry only one (or zero) large IBD block in their genome (like in Fig. 1, right side) microsatellites may yield significant HFCs without a need for HHCs to be high (e.g. $r = 0.053$ for IR_{ms}, blocked sampling in the $F < 0.15$ subset, Table 2). Hence, contrary to common perception, significant HHCs are not a prerequisite for significant HFCs.

Estimates of genome-wide heterozygosity from molecular markers can be calculated in a range of different ways. Our data showed that IR was more informative than any other measure. Sub-sampling of markers showed the highest repeatabilities (HHCs) for IR, both for microsatellite and for SNP markers. Correlations between microsatellite and SNP-based measures were also the highest for IR (compared to other measures). Finally, IR also produced the strongest HFCs. Yet for all these measures, the difference between IR and, for instance, HL measures would be too small to reach significance, so this difference should not be emphasized too much. In line with earlier suggestions (Tsitrone *et al.* 2001; Goudet & Keller 2002; Coltman & Slate 2003), mean d^2 proved the least informative measure.

In line with theoretical expectations (Kristensen *et al.* 2010), we found that traits closer to fitness (female fecundity, male siring success) showed greater inbreeding depression than morphological, ornamental or behavioural traits. In an earlier study (Bolund *et al.* 2010), we compared the highly inbred individuals from the 5th generation to their outbred half-siblings and concluded that the two most-studied sexual traits (courtship rate and beak colour) also showed the greatest amount of inbreeding depression compared to other traits. The additional data presented here (the $F < 0.15$ subset) do not confirm the previous conclusion. HFCs for these two traits only ranked 10th and 8th in our list of 11 phenotypic traits (Fig. 4). In both the earlier and the present study, the standard errors for inbreeding effects or HFCs were so large that such between-trait comparisons should be regarded with caution. The overall weakest (and hence nonsignificant) HFCs were found for female hopping activity in a choice chamber, a finding that does not support the idea of condition dependence of this behavioural trait (Woodgate *et al.* 2010). Alternatively, the notoriously high variance associated with the measurement of behavioural traits may have hindered the detection of small magnitude HFCs.

The above conclusions regarding the apparent superiority of microsatellite markers over pedigree information are drawn from an empirical data set rather than from an analytical approach. Moreover, this data set might be regarded as a special case, where the zebra finch's large linkage blocks (Fig. 1) and high allelic richness in microsatellites (Table S1, Supporting information) work in favour of microsatellite markers being informative about genome-wide IBD. An extensive sim-

ulation study would be required to determine the generality of our findings. The relative information content of pedigree vs. microsatellite data in relation to pedigree length and marker numbers would probably depend on the markers' allelic richness, the number and frequency distribution of haplotypes segregating in a population, variance in inbreeding, the number and size distribution of independently segregating units and the distribution of cross-over rates over those chromosomes. The complexity of this multidimensional problem is such that we could not address this within our empirical study. A general conclusion of whether 10 microsatellite markers or five generations of pedigree data will be better at reflecting inbreeding depression will have to await such modelling work. However, from a practical point of view, it may often be easier and cheaper to gather the molecular data, because the collection of pedigree data may be time-consuming and will often already require such molecular data to confirm or correct the putative pedigree links.

Acknowledgements

We thank M. Schneider, N. Backström and H. Mellenius for laboratory work; E. Bolund, C. Burger, S. Janker and K. Martin for help with data collection; S. Bauer, E. Bodendorfer, P. Brekke, A. Grötsch, J. Hacker, M. Halser, K. Martin, J. Minshull, P. Neubauer, J. Pellatt, A. Peter, N. Roddis, M. Ruhdofer, A. Türk, F. Weigel, B. Wörle, R. Yeates and P. Young for providing animal care; T. Birkhead for providing birds; and J. Olano-Marin, S. Nakagawa and three anonymous reviewers for critical comments. This study was funded by financial support from the Max Planck Society (B.K.), the Swedish Research Council (H.E.), the Knut and Alice Wallenberg Foundation (H.E.), the European Research Council (H.E), the German Research Foundation grants FO340/1-1, FO340/1-2 and FO340/1-3 (W.F.) and the European Union grant HPMF-CT-2002-01871 (W.F.).

References

Amos W, Worthington Wilmer J, Fullard K *et al.* (2001) The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society B-Biological Sciences*, **268**, 2021–2027.

Aparicio JM, Ortego J, Cordero PJ (2006) What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology*, **15**, 4659–4665.

Backström N, Forstmeier W, Schielzeth H *et al.* (2010) The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research*, **20**, 485–495.

Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology*, **13**, 3021–3031.

Bensch S, Andren H, Hansson B *et al.* (2006) Selection for Heterozygosity Gives Hope to a Wild Population of Inbred Wolves. *PLoS ONE*, **1**, e72.

Bierne N, Launey S, Naciri-Graven Y, Bonhomme F (1998) Early effect of inbreeding as revealed by microsatellite analyses on Ostrea edulis larvae. *Genetics*, **148**, 1893–1906.

Bolund E, Martin K, Kempenaers B, Forstmeier W (2010) Inbreeding depression of sexually selected traits and attractiveness in the zebra finch. *Animal Behaviour*, **79**, 947–955.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity-fitness correlations in animal populations. *Molecular Ecology*, **18**, 2746–2765.

Coltman DW, Slate J (2003) Microsatellite measures of inbreeding: a meta-analysis. *Evolution*, **57**, 971–983.

Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution*, **53**, 1259–1267.

Coulson TN, Pemberton JM, Albon SD *et al.* (1998) Microsatellites reveal heterosis in red deer. *Proceedings of the Royal Society B-Biological Sciences*, **265**, 489–495.

David P (1998) Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity*, **80**, 531–537.

DeRose MA, Roff DA (1999) A comparison of inbreeding depression in life-history and morphological traits in animals. *Evolution*, **53**, 1288–1292.

Derrida B, Manrubia SC, Zanette DH (2000) On the genealogy of a population of biparental individuals. *Journal of Theoretical Biology*, **203**, 303–315.

Forstmeier W, Coltman DW, Birkhead TR (2004) Maternal effects influence the sexual behaviour of sons and daughters in the zebra finch. *Evolution*, **58**, 2574–2583.

Forstmeier W, Schielzeth H, Schneider M, Kempenaers B (2007a) Development of polymorphic microsatellite markers for the zebra finch (*Taeniopygia guttata*). *Molecular Ecology Notes*, **7**, 1026–1028.

Forstmeier W, Segelbacher G, Mueller JC, Kempenaers B (2007b) Genetic variation and differentiation in captive and wild zebra finches (*Taeniopygia guttata*). *Molecular Ecology*, **16**, 4039–4050.

Forstmeier W, Mueller JC, Kempenaers B (2010) A polymorphism in the oestrogen receptor gene explains covariance between digit ratio and mating behaviour. *Proceedings of the Royal Society B-Biological Sciences*, **277**, 3353–3361.

Forstmeier W, Martin K, Bolund E, Schielzeth H, Kempenaers B (2011) Female extrapair mating behavior can evolve via indirect selection on males. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 10608–10613.

Goudet J, Keller L (2002) The correlation between inbreeding and fitness: does allele size matter? *Trends in Ecology and Evolution*, **17**, 201–202.

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, **11**, 2467–2474.

Hansson B, Westerdahl H, Hasselquist D, Akesson M, Bensch S (2004) Does linkage disequilibrium generate heterozygosity-fitness correlations in great reed warblers? *Evolution*, **58**, 870–879.

He C, Weeks DE, Buyske S *et al.* (2011) Enhanced genetic maps from family-based disease studies: population-specific comparisons. *BMC Medical Genetics*, **12**, e15.

Hill WG, Weir BS (2011) Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research*, **93**, 47–64.

Keller MC, Visscher PM, Goddard ME (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, **189**, 237–249.

Kinghorn BP, Kinghorn AJ (2010) *Pedigree Viewer 6.5*. University of New England, Armidale, Australia.

Kristensen TN, Pedersen KS, Vermeulen CJ, Loeschcke V (2010) Research on inbreeding in the 'omic' era. *Trends in Ecology & Evolution*, **25**, 44–52.

Leutenegger AL, Prum B, Genin E *et al.* (2003) Estimation of the inbreeding coefficient through use of genomic data. *American Journal of Human Genetics*, **73**, 516–523.

Markert JA, Grant PR, Grant BR *et al.* (2004) Neutral locus heterozygosity, inbreeding, and survival in Darwin's ground finches (Geospiza fortis and G-scandens). *Heredity*, **92**, 306–315.

Olano-Marin J, Mueller JC, Kempenaers B (2011a) Correlations between heterozygosity and reproductive success in the blue tit (Cyanistes caeruleus): an analysis of inbreeding and single locus effects. *Evolution*, **65**, 3175–3194.

Olano-Marin J, Mueller JC, Kempenaers B (2011b) Heterozygosity and survival in blue tits (Cyanistes caeruleus): contrasting effects of presumably functional and neutral loci. *Molecular Ecology*, **20**, 4028–4041.

Pemberton J (2004) Measuring inbreeding depression in the wild: the old ways are the best. *Trends in Ecology & Evolution*, **19**, 613–615.

Pemberton JM (2008) Wild pedigrees: the way forward. *Proceedings of the Royal Society B-Biological Sciences*, **275**, 613–621.

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Santure AW, Stapley J, Ball AD *et al.* (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology*, **19**, 1439–1451.

Schielzeth H, Bolund E, Forstmeier W (2010) Heritability of and early-environmental effects on variation in mating preferences. *Evolution*, **64**, 998–1006.

Schielzeth H, Kempenaers B, Ellegren H, Forstmeier W (2012) QTL linkage mapping of zebra finch beak colour shows an oligogenic control of a sexually selected trait. *Evolution*, **68**, 18–30.

Slate J, David P, Dodds KG *et al.* (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity*, **93**, 255–265.

Smouse PE (2010) How many SNPs are enough? *Molecular Ecology*, **19**, 1265–1266.

Sokal RR, Rohlf FJ (1995) *Biometry*, 3rd edn. Freeman, New York.

Stam P (1980) The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genetical Research*, **35**, 131–155.

Taylor SS, Sardell RJ, Reid JM *et al.* (2010) Inbreeding coefficient and heterozygosity-fitness correlations in unhatched and hatched song sparrow nestmates. *Molecular Ecology*, **19**, 4454–4461.

Tsitrone A, Rousset F, David P (2001) Heterosis, marker mutational processes and population inbreeding history. *Genetics*, **159**, 1845–1859.

Woodgate JL, Bennett ATD, Leitner S, Catchpole CK, Buchanan KL (2010) Developmental stress and female mate choice behaviour in the zebra finch. *Animal Behaviour*, **79**, 1381–1390.

Zann R (1996) *The Zebra Finch: A Synthesis of Field and Laboratory Studies*. Oxford University Press, New York.

W.F. is an evolutionary geneticist and is working on the genetics of sexual behaviour in the zebra finch. H.S. is an evolutionary biologist and is interested in the genetic architecture of quantitative traits. J.C.M. is a behavioural geneticist interested in the genetic architecture of fitness-relevant trait variation in natural populations. H.E. interest is the integration of genomics and evolutionary biology. B.K. is a behavioural ecologist interested in sexual selection and the evolution of mating behaviour.

Data accessibility

Pedigree and SNP genotypes: DRYAD entry doi: 10.5061/dryad.r044b (Schielzeth *et al.* 2012). Phenotypes and microsatellite genotypes: DRYAD entry doi: 10.5061/dryad.6v0h65g3.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Counts of inbreeding coefficients (F) in the additional data set (generations 6–7; $N = 1003$ individuals), calculated from pedigrees that were progressively shortened from seven to three generations.

Fig. S2 Average strength of correlation between pedigree-based inbreeding coefficient F (inverted) and three phenotypic traits (Tarsus length length, adult mass, mass day 8) in relation to variance in F in two data sets.

Table S1 Information on the 11 microsatellite markers used in this study.

Table S2 Correlations between pairs of phenotypic traits.

Table S3 Correlations between various measures of heterozygosity and phenotypic traits for the main data set (generations 2–5).

Table S4 Correlations between internal relatedness (IR) (inverted to give positive correlations) based on varying numbers of SNPs and phenotypic traits for the main data set (generations 2–5).

Table S5 Correlations between pedigree-based *F* (inverted to produce positive correlations) and phenotypic traits in the additional data set from generations 6–7 (all $N = 1003$ individuals).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.