ORIGINAL ARTICLE

Estimating genome-wide heterozygosity: effects of demographic history and marker type

JM Miller[1], RM Malenfant[1], P David[2], CS Davis[1], J Poissant[3], JT Hogg[4], M Festa-Bianchet[5] and DW Coltman[1]

Heterozygosity–fitness correlations (HFCs) are often used to link individual genetic variation to differences in fitness. However, most studies examining HFCs find weak or no correlations. Here, we derive broad theoretical predictions about how many loci are needed to adequately measure genomic heterozygosity assuming different levels of identity disequilibrium (ID), a proxy for inbreeding. We then evaluate the expected ability to detect HFCs using an empirical data set of 200 microsatellites and 412 single nucleotide polymorphisms (SNPs) genotyped in two populations of bighorn sheep (*Ovis canadensis*), with different demographic histories. In both populations, heterozygosity was significantly correlated across marker types, although the strength of the correlation was weaker in a native population compared with one founded via translocation and later supplemented with additional individuals. Despite being bi-allelic, SNPs had similar correlations to genome-wide heterozygosity as microsatellites in both populations. For both marker types, this association became stronger and less variable as more markers were considered. Both populations had significant levels of ID; however, estimates were an order of magnitude lower in the native population. As with heterozygosity, SNPs performed similarly to microsatellites, and precision and accuracy of the estimates of ID increased as more loci were considered. Although dependent on the demographic history of the population considered, these results illustrate that genome-wide heterozygosity, and therefore HFCs, are best measured by a large number of markers, a feat now more realistically accomplished with SNPs than microsatellites.
Heredity (2014) 112, 240–247; doi:10.1038/hdy.2013.99; published online 23 October 2013

INTRODUCTION

Individual heterozygosity can be easily measured using genetic markers and is often used as proxy for inbreeding (Hansson and Westerberg, 2002; Balloux *et al.*, 2004). Many studies have examined the relationship between individual genetic diversity and fitness using heterozygosity–fitness correlations (HFCs). HFCs appeal to wildlife and conservation biologists who cannot easily reconstruct pedigrees and directly measure inbreeding in natural populations, especially for endangered species (Balloux *et al.*, 2004; Grueber *et al.*, 2008; Chapman *et al.*, 2009). Meta-analyses of HFCs, however, have revealed that effect sizes are often weak (Coltman and Slate, 2003; Chapman *et al.*, 2009; Szulkin *et al.*, 2010). The modest numbers of genetic markers typically employed may provide inaccurate estimates of genome-wide heterozygosity. It is also unclear whether different types of markers provide similar information content.

Microsatellites have been the most commonly used markers to investigate HFCs. They are relatively abundant in the genome and are highly polymorphic. However, their mutational mechanism is not well understood, and the high mutation rate likely leads to elevated levels of homoplasy, which can underestimate true heterozygosity (Hansson and Westerberg, 2002). In addition, the process of isolating and characterizing novel loci often selects for the most polymorphic markers, resulting in ascertainment bias and an upwardly skewed estimate of genome-wide diversity (Brandstrom and Ellegren, 2008).

Despite their growing use in molecular ecology and evolutionary biology, single nucleotide polymorphisms (SNPs) have been less widely used in HFC studies, perhaps because they are almost exclusively bi-allelic. However, they have some advantages over microsatellites: they are more abundant in the genome, have a well-understood mutational mechanism with low levels of homoplasy and are amenable to high throughput genotyping (Morin *et al.*, 2004). Several authors contend that SNPs may be more suitable than microsatellites for HFCs. Tsitrone *et al.* (2001) used extensive simulation studies to examine the effect of different mutational patterns (corresponding to SNPs and microsatellites) and demographic history on the expected correlation between heterozygosity and fitness. Their results point to a complex interplay between these two factors. The high mutation rate of microsatellites should make them more suitable to detect HFCs that result from recent inbreeding due to crosses between relatives or small population size. The lower mutation rates typical of SNPs may make them better than microsatellites to detect HFCs resulting from ancient inbreeding, such as when two subpopulations accumulate genetic differentiation during a long period of isolation and then come back into contact (Tsitrone *et al.*, 2001). Chakraborty (1981) and DeWoody and DeWoody (2005) argued that correlations between heterozygosity at a set of loci and genomic heterozygosity would be high only when the set of marker loci represents a high fraction of all polymorphisms in

[1]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada; [2]Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, Montpellier, France; [3]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK; [4]Montana Conservation Science Institute, Missoula, MT, USA and [5]Département de biologie et Centre d'Études Nordiques, Université de Sherbrooke, Sherbrooke, Québec, Canada
Correspondence: JM Miller, Department of Biological Sciences, University of Alberta, CW 405 Biological Sciences Building, Edmonton, Alberta T6G 2E9, Canada.
E-mail: jmm1@ualberta.ca

the genome. However, they modeled populations with no inbreeding and no correlations among loci, a condition under which HFC does not occur unless the marker loci themselves are coding for fitness traits, which is not the case for most recently published HFC studies. It remains unclear whether many markers with low genetic diversity (SNPs) or fewer markers with higher diversity (microsatellites) are more suitable to explore general-effect HFCs. This question becomes important as new technologies allow for the development of larger genome-wide marker sets of both SNPs and microsatellites (Baird *et al.*, 2008; Davey *et al.*, 2011).

Studies of HFCs commonly use 10–30 loci (Chapman *et al.*, 2009). However, as the demographic history of a population will heavily influence correlations in marker heterozygosity within individuals (Ljungqvist *et al.*, 2010; Szulkin *et al.*, 2010), such modest numbers of markers may sometimes be insufficient (Balloux *et al.*, 2004; Väli *et al.*, 2008; Ljungqvist *et al.*, 2010; Forstmeier *et al.*, 2012). For example, Väli *et al.* (2008) looked at the correlation between heterozygosity at 10–17 microsatellites and allelic diversity in 10 introns across eight populations of carnivores. They found a positive correlation between average heterozygosity and allelic diversity among populations but not between individual heterozygosity at SNPs and microsatellites. In general, HFCs are not expected within populations without measurable identity disequilibrium (ID), a correlation in identity by descent among markers (Slate *et al.*, 2004; David *et al.*, 2007; Szulkin *et al.*, 2010). ID arises from departures from random mating (for example, inbreeding) or demographic events (for example, a population bottleneck or admixture) that cause the heterozygosity of loci to become associated (Bierne *et al.*, 2000; Szulkin *et al.*, 2010). In the absence of ID, HFCs will be detected only if one or more markers are directly associated with a gene influencing fitness, so-called local or direct effects (Hansson and Westerberg, 2002). These direct effect correlations are difficult to detect because they depend on the specific marker set used in a study.

Here, we examine the contrasting effects of the number of markers considered and marker type on the ability to detect general-effect HFCs. We first use existing models of HFC to derive broad theoretical predictions about how many loci are needed to adequately measure genomic heterozygosity assuming different levels of inbreeding and marker genetic diversity. We then use large sets of both microsatellites and SNPs genotyped in two populations of bighorn sheep (*Ovis canadensis*) to approach this question empirically. Our two study populations have very different demographic histories: one was founded in the 1920's with 12 individuals and experienced a prolonged bottleneck post founding, then recent admixture following a 'genetic rescue' where 15 individuals were intentionally introduced into the population. The other is a native population with no genetic evidence of a comparable bottleneck. These contrasting histories should affect the magnitude of ID and hence our ability to detect HFCs. In the population subject to 'genetic rescue', ID is expected to be higher, arising both from historical inbreeding and admixture following the introductions. In the native population, ID will likely track demography, arising if heterozygosity decreases because of inbreeding. Therefore, power to detect HFCs should be greater in the bottlenecked population than the native one. To test these hypotheses, we first sought to measure the strength of correlations between estimates of heterozygosity from microsatellites and SNPs within individuals. We then examined how many markers are needed to accurately reflect genome-wide heterozygosity and ID in these two populations.

Theory

General-effect HFCs arise as the product of two correlations: the correlation between fitness (*W*) and inbreeding (*f*), and the correlation between *f* and heterozygosity (*h*) (Slate *et al.*, 2004; Szulkin *et al.*, 2010), such that:

$$r(W, h) = r(W, f) \cdot r(f, h). \qquad (1)$$

For the purposes of this paper, we do not consider the correlation between *W* and *f*. Rather, we focus on the power of different marker sets to detect the correlation between *h* and *f*.

Two sources of sampling variance may affect HFCs. One is the sampling of individuals: if a small sample is taken in a population that contains a small proportion of inbred individuals, the proportion of inbred individuals in the sample is subjected to a large variance that directly affects HFC estimates. The estimated HFC will be stronger or weaker than true the HFC in the population simply because the proportion of inbred individuals in the sample happens to be higher or lower than their frequency in the population. This source of error can be large but the only way to reduce it is to sample more individuals. The second source of variance arises from the fact that heterozygosity measured at a set of marker loci is not perfectly correlated with genomic heterozygosity and/or with individual inbreeding level. This error depends on the characteristics of the marker loci (number and genetic diversity). We will mainly concentrate on this second type of error, assuming that all efforts have been made to reduce the first source of error.

The problem is now to estimate how well inbreeding is measured by (or correlated to) heterozygosity in a sample of markers. We consider standardized heterozygosity (*sensu* Coltman *et al.* (1999) denoted H** for consistency with notations in Szulkin *et al.* (2010)) at a set of loci *L*, of size *A*, where h_i is the observed heterozygosity at locus *i* and upper bar denotes expectations

$$H_A^{**} = \frac{\sum_{i \in A} h_i}{\sum_{i \in A} \overline{h}_i} = \frac{\sum_{i \in A} h_i}{L_A \overline{h}_A} \qquad (2)$$

based on Szulkin *et al.* (2010), the expected correlation between H** and inbreeding level (*f*) is

$$r^2(H_A^{**}, f) = \frac{g_2}{\sigma^2(H_A^{**})}, \qquad (3)$$

where g_2 is the covariance of heterozygosity between markers standardized by their average heterozygosity (David *et al.*, 2007) and

$$\sigma^2(H_A^{**}) = \frac{\sum_{i \in A} \overline{h}_i (1 - \overline{h}_i) + 2g_2 \sum \sum_{(j > i) \in A} \overline{h}_i \overline{h}_j}{(L_A \overline{h}_A)^2}. \qquad (4)$$

Assuming that all loci in the set L_A have the same average heterozygosity h_A (for simplicity), this gives

$$r^2(H_A^{**}, f) = \frac{L_A g_2 \overline{h}_A}{1 - \overline{h}_A + (L_A - 1) g_2 \overline{h}_A} \qquad (5)$$

from this it can be seen that the correlation approaches unity as the number of markers increases, and depends mostly on the product of the number of loci and their average heterozygosity: 100 loci with $h = 0.1$ are equivalent to 20 loci with $h = 0.5$. The rate at which the correlation approaches 1 increases with the ID (represented by g_2). When g_2 is null, the correlation is necessarily zero because inbreeding does not vary in the population. In such cases, trying to estimate genome-wide heterozygosity from a small set of markers is pointless because all the variance comes from sampling error.

Often one does not have an independent measure of inbreeding (for example, pedigrees), and therefore the above formula cannot be

checked directly. Instead, what can be performed (and will be performed below using real data) is (i) to check the consistency between two different subsets of marker loci (for example, SNPs and microsatellites) and (ii) to check how fast estimates of heterozygosity based on increasing numbers of loci converge to the most precise estimate available (which uses all loci). Theoretical predictions can be obtained for (i) and (ii). The first is simply the correlation between heterozygosities at two non-overlapping sets of loci; it can be simply computed based on the assumption (underlying the general-effect model) that this correlation emerges only as a result of the common dependency of heterozygosities in both sets of markers on the extent of inbreeding. Therefore,

$$r^2\left(H_A^{**}, H_B^{**}\right) = \frac{(g_2)^2}{\sigma^2(H_A^{**})\sigma^2(H_B^{**})} = r^2\left(H_A^{**}, f\right) r^2\left(H_B^{**}, f\right). \qquad (6)$$

The quantity relevant to point (ii) is the correlation between heterozygosity at L loci and heterozygosity at a subset (S) of these loci, which contain a fraction p_S of the total number of loci. For simplicity, we model a situation in which there are no missing data and all individuals are typed at the same set of loci. If individuals are typed at sets of loci that substantially differ in their average heterozygosity correlations will be reduced; however, this is unlikely to be the case unless substantial amounts of data are missing. As correlations are insensitive to scaling by a constant, we can work here with raw heterozygosities H (not standardized heterozygosity H^{**}). Using raw heterozygosity, total heterozygosity H is the sum of heterozygosity at the S loci (H_S) and at the remaining loci (H_R). Thus,

$$\begin{aligned} r\left(H_S^{**}, H^{**}\right) = r(H_S, H) &= \frac{\mathrm{COV}(H_S, H)}{\sqrt{\sigma^2(H_S)\sigma^2(H)}} \\ &= \frac{\mathrm{COV}(H_S, H_S + H_R)}{\sqrt{\sigma^2(H_S)\sigma^2(H)}} \\ &= \sqrt{\frac{\sigma^2(H_S)}{\sigma^2(H)}} + \sqrt{\frac{\sigma^2(H_R)}{\sigma^2(H)}} r\left(H_S^{**}, H_R^{**}\right) \end{aligned} \qquad (7),$$

where $\sigma^2(H_S)$ is the numerator of Equation 4 ($\sigma^2(H_R)$ and $\sigma^2(H)$ can be computed similarly, making the summations over the appropriate sets of loci). Assuming that all loci have the same heterozygosity h one obtains, after some algebra

$$r^2\left(H_S^{**}, H^{**}\right) = p_S\left(1 + (1-p_S)\frac{L\bar{h}g_2}{1-\bar{h}+\bar{h}g_2(LpS-1)}\right) \qquad (8)$$

in this formula, one can distinguish two terms: the first is simply the proportion of loci included in the subset (p_S) and reflects the fact that subset S will always capture a proportion of the variance in total heterozygosity at all loci because they are part of the total. The ability of the S subset to inform about the other loci (hence about the genome in general) is reflected by the second term, which relies on the existence of ID (g_2): through this disequilibrium, the loci in S inform about the state of other loci and hence capture a more than proportional share of total variance in heterozygosity.

MATERIALS AND METHODS
Study populations
We examined patterns of heterozygosity in bighorn sheep at the National Bison Range (Montana, USA; NBR) and at Ram Mountain (Alberta, Canada; RM). In both populations, long-term studies follow individuals throughout their lives. The National Bison Range population was founded in 1922 via translocation of 12 individuals from Banff National Park (Alberta, Canada). Individual monitoring started in 1979, with genetic sampling beginning in 1988. Beginning in 1985, NBR experienced a 'genetic rescue' via intentional

translocation of 15 individuals from neighboring populations to prevent local extinction after years of isolation and inbreeding (Hogg *et al.*, 2006; Miller *et al.*, 2012). Prior to the introduction, census size and growth rate had been steadily declining (average census size of 48 sheep between 1922 and 1985). Following the supplementation, there has been an increase both in census size (142 sheep in late 2012) and genetic diversity (Hogg *et al.*, 2006; Miller *et al.*, 2012).

In contrast, Ram Mountain is a native population in which individual-based monitoring began in 1972 with genetic sampling starting in 1988 (Jorgenson *et al.*, 1997; Coltman *et al.*, 2002). Between 1988 and 2010, census size fluctuated between 38 and 210 sheep, declining recently because of low recruitment (Jorgenson *et al.*, 1997) and cougar (*Puma concolor*) predation (Festa-Bianchet *et al.*, 2006).

Marker genotyping and selection
SNP genotypes used in this study were generated by typing 27 individuals from NBR and 50 from RM on the OvineSNP50 BeadChip (Miller *et al.*, 2011), yielding 853 variable loci. For this study, we excluded loci that were genotyped in <90% of individuals in both populations ($N = 38$), had <5% minor allele frequency ($N = 392$) and did not conform to Hardy–Weinberg expectations following a Bonferroni correction ($N = 2$). We also excluded any loci mapped to the X chromosome ($N = 9$). This resulted in a final data set of 412 SNPs (Supplementary Table 1). We included loci polymorphic in one population but monomorphic in the other. Note that because of their discovery via cross-species application of the OvineSNP50 BeadChip, the SNPs used in this study are widely distributed in the genome, and are mostly intergenic as few are expected to be in or near genes based on annotation of the domestic sheep genome (Miller *et al.*, 2011).

Microsatellite loci used in analyses were a subset ($N = 200$; Supplementary Table 2) of those used to construct a bighorn sheep linkage map (Poissant *et al.*, 2010). Primer information and PCR conditions for the markers can be found in Poissant *et al.* (2009, 2010) and references therein. All loci conformed to Hardy–Weinberg expectations following a Bonferroni correction. Loci were retained only if they were genotyped in both populations and had <25% missing genotypes in the samples from either population. As with the SNP set, loci on the X chromosome were excluded.

In total, 26 individuals from NBR and 48 from RM were genotyped at both sets of markers and included in subsequent analyses. The individuals from NBR were born between 1981 and 2004 and include descendants of the original founders of the population ($N = 4$), transplanted individuals ($N = 2$) and their progeny ($N = 20$).

Statistical analyses
We calculated individual standardized multilocus heterozygosity (stMLH) following Coltman *et al.* (1999). The relationship between individual stMLH from each marker set was assessed using reduced major axis regression with 1000 jackknife iterations, as implemented in RMA version 1.21 (Bohonak and van der Linde, 2004). Reduced major axis regression was chosen to account for the uncertainty associated with stMLH measures used as both dependent and independent variables. For resampling tests, 100 random subsets of markers were sampled without replacement from the full data sets using the 'sample' function in R version 2.13.0 (R Development Core Team, 2005). For microsatellites, subsets of 5, 10, 20, 30, 50, 75, 100, 125, 150 and 175 loci were extracted, whereas for SNPs the subsets consisted of 20, 50, 75, 100, 150, 200, 250, 300, 350 and 400 loci. We then calculated stMLH for each subset using a custom Perl script. The coefficients of determination (r^2) were compared between the stMLH calculated for each subset and a total stMLH calculated from the concatenation of the SNP and microsatellite data sets (all 612 loci). In addition, we calculated r^2 between each subset and total stMLH for all loci of their respective marker type. We then compared these results to the theoretical predictions described in the previous section.

Estimates of ID and expected power to detect HFCs
To measure ID, we used the program RMES (David *et al.*, 2007) to calculate the g_2 statistic. Significant covariance among marker genotypes can be attributed to inbreeding, admixture or a bottleneck (David *et al.*, 2007;

Szulkin et al., 2010). Assessment of significant levels of ID ($g_2 > 0$) utilized 1000 resampling iterations. Calculations were performed for both populations on the full microsatellite set, the full SNP set, the concatenated marker set and all marker subsets used in the stMLH resampling calculations. We also examined the effect of the number of individuals sampled on the accuracy of g_2 estimates. For this analysis, we bootstrapped both the full microsatellite and full SNP data sets in each population, generating 100 replicates containing different numbers of individuals: 5, 10, 20, 30, 50 or 100 individuals.

To explore the power of different marker sets to detect HFCs, we calculated the expected correlation between f and stMLH using Equation 3 based on empirical estimation of variance in heterozygosity as well as its theoretical value based only on the number and average heterozygosity of the markers (Equation 5). Equation 5 has the advantage that it can be applied to assess the power of a study before actually performing it, as it requires only

Table 1 Estimate of identity disequilibrium (g_2) and expected r^2 between inbreeding (f) and stMLH (H_A) for the different full marker sets in each population of sheep**

	s.d. of stMLH	g_2	s.d. of g_2	Expected r^2 (H_A**, f)
NBR				
Microsatellites	0.22245	0.05554	0.02773	0.95695
SNPs	0.23662	0.06481	0.03081	0.91378
Combined	0.24891	0.06074	0.02920	0.96095
RM				
Microsatellites	0.07086	0.00198	0.00412	0.38321
SNPs	0.09843	0.00456	0.00519	0.43703
Combined	0.07093	0.00309	0.00254	0.54937

Abbreviations: SNP, single nucleotide polymorphism; stMLH, standardized multilocus heterozygosity.

approximated parameter values. As with our estimates of g_2, we calculated $r^2(f, h)$ for the full SNP, microsatellite and concatenated data sets, as well as all subsets. When the estimate of g_2 was negative we set $r^2(f, h)$ to 0.

RESULTS

Summary statistics of markers

In NBR, the average ($\pm$ s.d.) minor allele frequency for SNP loci was 0.197 (± 0.160), and average observed heterozygosity (Ho) was 0.279 (± 0.202). In RM, average minor allele frequency was 0.212 (± 0.151) and Ho 0.292 (± 0.178). For microsatellite loci, Ho was 0.643 (± 0.161), and number of alleles per locus ranged from 2 to 9 (average 4.39 ± 1.43) in NBR, whereas in RM Ho was 0.610 (± 0.157) and the number of alleles per locus ranged from 2 to 10 (average 4.21 ± 1.48).

Estimates of ID and expected power to detect HFCs

All estimates of g_2 based on total marker sets were greater than zero for both populations ($P < 0.001$; Table 1). However, g_2 was much stronger in NBR than RM. Across both populations, the full SNP set produced higher estimates of g_2 than the full microsatellite set, and the combined data sets were intermediate. Our subsampling analyses showed that average values of g_2 on par with the genome-wide estimates of that same marker type were obtained even when few markers were considered (Figure 1a–d). However, there was considerable variation around these estimates. For example, in RM the s.d. estimates were larger than the average values of g_2 when fewer than 75 microsatellites or 150 SNPs were examined (Figure 1b and d).

Bootstrapping the full data sets similarly showed that a stable average value of g_2 can be estimated with small sample sizes; however, larger sample sizes increase the precision (Supplementary Table 3).
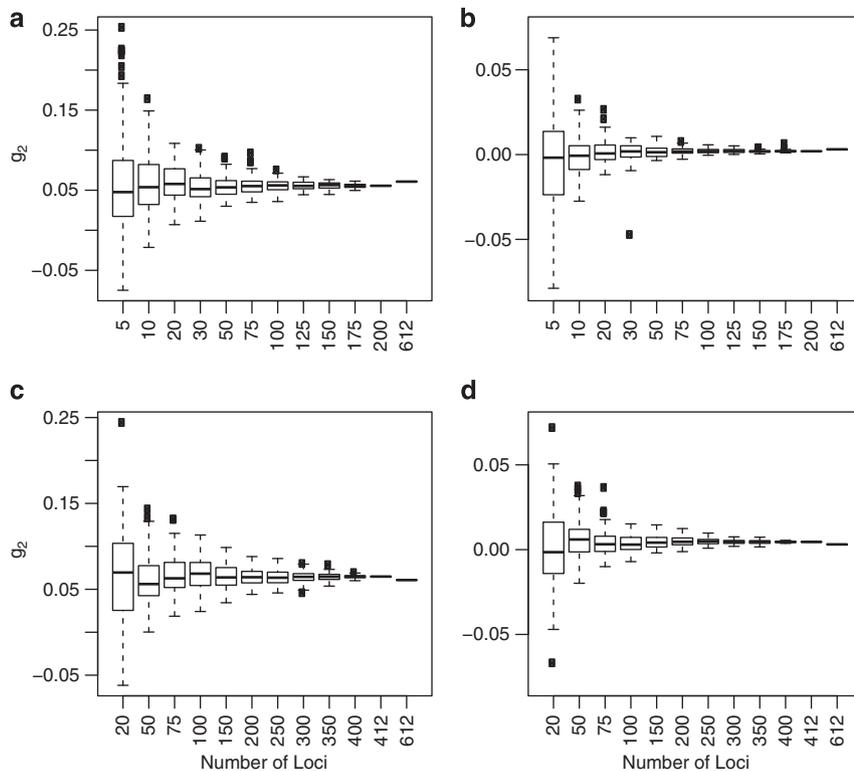
Figure 1 Box plots showing the average level of identity disequilibrium (g_2) for the different marker subsets. Each subset was generated by sampling markers from the total data set 100 times. Plots A and C correspond to NBR at microsatellites (**a**) and SNPs (**c**), whereas plots B and D correspond to RM at microsatellites (**b**) and SNPs (**d**).

One might have expected the effects of increasing sample size to be more apparent in NBR, where the data set contains a few highly inbred individuals and the chances of sampling them would therefore lead to large s.d. around the estimates of g_2. However, when scaled as a percentage of the average estimate, the effect of sampling is greater for RM than NBR. Even when 100 individuals were assumed, s.d. values representing $>14\%$ of the average g_2 estimate in NBR and $>30\%$ in RM were seen.

Expected r^2 between f and stMLH for the various full marker sets (Table 1) were stronger in NBR than RM. In both populations, the strongest r^2 was seen from the combined marker data set. The expected r^2 increased and the variation around estimates decreased as the number of markers increased (Figure 2a–d). There was an apparent upward bias of the r^2 between f and stMLH when measured by the full marker sets compared with the subsets. This bias is likely an artifact given that there is only one estimate based on the full marker sets (rather than 100 permutations) and that the correlation is based on a g_2 statistic that still has error associated with it (Table 1).

Correlations between marker types and among subsets

Individual stMLH was significantly positively correlated between marker types in both populations (Figure 3). However, the correlation was much stronger in NBR (NBR $r = 0.954$, $t_{24} = 15.656$, $P \ll 0.001$; RM $r = 0.360$, $t_{46} = 2.620$, $P = 0.011$). Based on Equation 6, the expected correlations were $r = 1.15$ and $r = 0.44$ for NBR and RM, respectively. Slight differences between the predicted and observed values (as well as the $r > 1$) likely arise because the combined g_2 value is measured with error (Table 1). To ensure that the larger sample size in RM was not the main driver of the difference in correlation

between RM and NBR we jackknifed our data from RM resampling 100 sets of 26 individuals, this yielded an average correlation of 0.355 ± 0.121 (s.d.).

In our resampling analyses, r^2 values were stronger in NBR than RM regardless of the number or type of markers examined
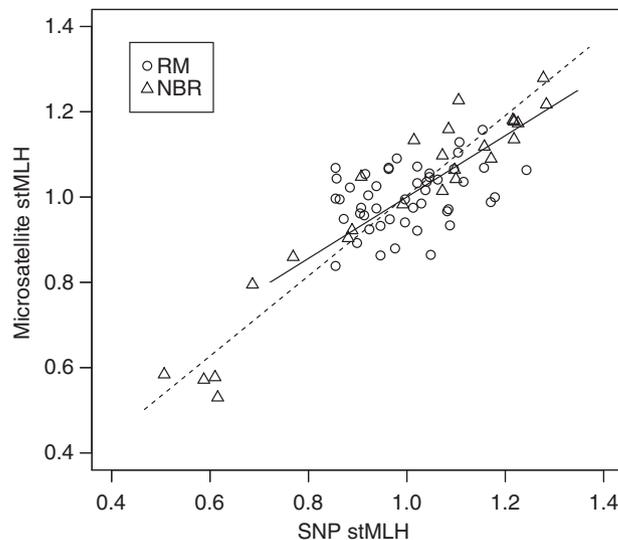
Figure 3 Correlation between individual heterozygosity at SNPs and microsatellites in RM and NBR. Reduced major axis regression lines are shown: $y = 0.7199x + 0.2802$ for RM (solid line); $y = 0.9401x + 0.0631$ for NBR (dashed line).
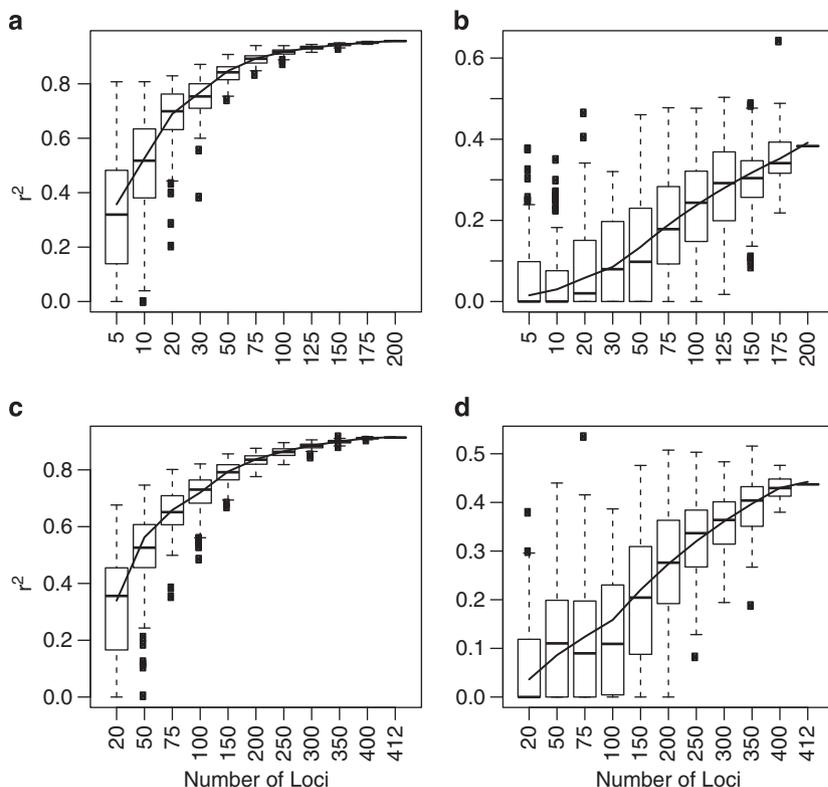
Figure 2 Box plots showing the average expected r^2 between inbreeding and heterozygosity for the different marker subsets. Plots A and C show correlation in NBR at microsatellites (**a**) and SNPs (**c**), whereas plots B and D show correlations in RM at microsatellites (**b**) and SNPs (**d**). Solid lines show predicted correlations based on Equation 5.
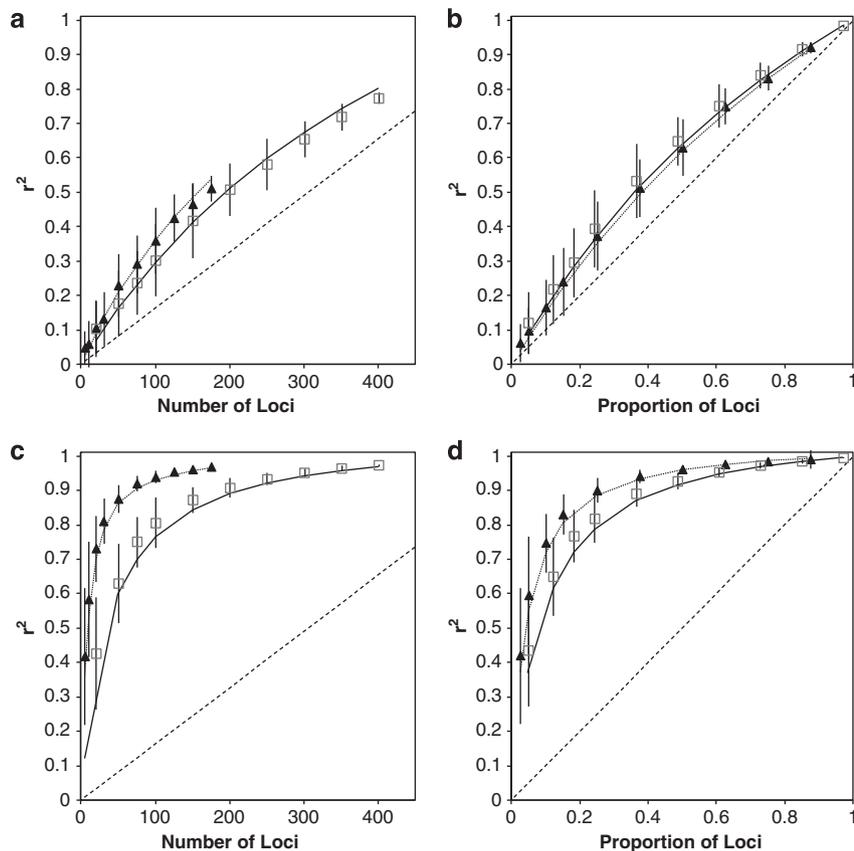
Figure 4 Average r^2 between marker subset stMLH and genome-wide stMLH. Each subset was generated by sampling markers from the total data set 100 times; error bars show s.d. Plots A and C show correlations for SNPs (open squares) and microsatellites (filled triangles) when all 612 loci are considered in RM and NBR, respectively. Plots B and D show correlations when subsets are compared with stMLH exclusively from the same marker type in RM (**b**) and NBR (**d**). Note that the *x* axis is now scaled as a proportion of the total number of either SNPs or microsatellites. Predicted correlations based on Equation 8 are shown for SNPs (solid lines) and microsatellites (dotted lines), dashed lines show predicted correlations among subsets in the absence of identity disequilibrium.

(Figures 4a–d). In both the SNP and microsatellite data sets, the correlation with the total measure of stMLH strengthened with increasing number of markers (Figures 4a and c). In NBR, correlations between microsatellite subsets and total stMLH were higher than those for an equal number of SNPs (Figure 4c). However, an asymptote of strong correlation ($r^2 > 0.9$) was reached with as few as 75 micro-satellites or 200 SNPs. These differences between marker types disappeared when marker subsets were compared with stMLH of only that same marker type and scaled as a proportion of the total number of either SNPs or microsatellites considered (Figure 4d). For RM, when equal numbers of markers were compared, microsatellites produced a marginally higher average correlation to total stMLH than SNPs (Figure 4a), although these differences were not significant. The two marker types gave near-identical correlations when marker subsets were compared with stMLH values from only the same marker type (Figure 4b). Average correlations to marker-specific estimates of stMLH were stronger than when subsets were compared with total stMLH (average increase of 0.12 for SNPs and 0.20 for microsatellites). For both populations, all empirical correlations exceeded the null expectations and closely paralleled predicted correlations (Figure 4).

DISCUSSION
Influence of population history
As expected, across all analyses the strength of association between marker heterozygosities as well as the expected ability to detect HFCs

was highly dependent on the demographic history of the population (Ljungqvist *et al.*, 2010; Szulkin *et al.*, 2010). Bighorn sheep tend to be philopatric and have a highly polygynous mating system, in which a few dominant males sire the majority of offspring (Hogg and Forbes, 1997; Coltman *et al.*, 2002). Thus, even in a native population such as RM a certain level of ID is to be expected. In addition, RM is relatively isolated and rarely receives immigrants (Rioux-Paquette *et al.*, 2010), furthering the likelihood of non-random association of alleles because of inbreeding. Disequilibrium is even more likely in NBR given its population history. Descendants of NBR founders are expected to have low overall genetic diversity after years of inbreeding, translocated individuals from neighboring herds will have relatively higher levels of diversity and their progeny are expected to have the highest heterozygosity as a result of the admixture between the founder and translocated individuals (Hogg *et al.*, 2006; Miller *et al.*, 2012).

Theory predicts that population history as well as mating system (that is, partial inbreeding or selfing), as summarized through the g_2 parameter, determines how well heterozygosity at a set of markers reflects heterozygosity at other loci and, by extension, genomic heterozygosity and inbreeding (Szulkin *et al.*, 2010, our Equations 3, 5 and 6). Theoretical predictions correctly match the observed correlations between heterozygosity at SNPs and heterozygosity at microsatellites in our data. Although significant correlations were seen in both populations, it was much tighter in NBR (Figure 3).

In contrast, Väli et al. (2008) found no significant correlation between individual heterozygosity at SNPs and microsatellites at the level of the individual in four populations of wolves (Canis lupus) and one of coyotes (C. latrans). However, both of these species have high dispersal rates and large effective population sizes (Pilot et al., 2006), which may not allow for such correlations to develop. In addition, Väli et al. (2008) used only 10–17 microsatellites and 25–51 SNPs in 10 introns, which our results suggest may not have had the power to detect an association in a population with low g_2 (for example, 0.001–0.005). The contrasting effect of demographic history is equally apparent when trying to estimate genome-wide heterozygosity from a subset of markers (Figure 4).

Interaction between the number of markers and marker type on correlations in stMLH

In NBR, microsatellites were more highly correlated to total stMLH than SNPs when equal numbers of markers were compared: 20 microsatellites predicted inbreeding as well as 75 SNPs—as expected given their higher average heterozygosity (0.643 compared with 0.279) and our theoretical equations (Equation 8). However, even a small number of either type of marker was highly correlated with inbreeding level and with total heterozygosity at all markers because of the high g_2. The situation was slightly different in RM. Here, SNPs and microsatellites gave essentially the same correlations to our total measure of stMLH when equal numbers of markers were compared. However, given the low g_2, a much larger number of loci (microsatellites or SNPs) is needed to adequately measure inbreeding in RM.

In short, microsatellites are more informative than SNPs because they have higher genetic diversity per locus; however, to find the most efficient strategy, one must consider that it is now becoming technically easier to develop and type a large number of SNPs than an equivalent number of microsatellites (Baird et al., 2008; Davey et al., 2011; Guichoux et al., 2011). These results agree with previous theoretical and empirical studies that suggested that highly heterozygous multi-allelic markers will have higher correlation between MLH and genome-wide heterozygosity than bi-allelic ones (Slate et al., 2004; Ljungqvist et al., 2010; Online Appendix 2 in Szulkin et al., 2010; Forstmeier et al., 2012) except in special cases (Tsitrone et al., 2001).

One factor that could seem to limit the robustness of our conclusions is that correlations to total heterozygosity could be biased because the data set contains a larger number of SNPs than microsatellites (412 versus 200 loci). We do not feel that this is a problem for several reasons. First, individual heterozygosity between the two marker types was correlated (Figure 3) and should therefore show the same patterns no matter the ratio of loci examined. Moreover, if the relative proportion of markers biased our estimates, one would expect correlations to the measure of total stMLH to be constrained by the marker's abundance in the total data set—for example, $r^2 < 0.33$ for microsatellites. Figure 4 shows that this is not the case: both sets of loci quickly rose to essentially perfect correlations in NBR and increased to levels well above their relative proportions in RM. Finally, in general, SNPs are more abundant in a genome than microsatellites, and although the ratio is not 2:1, our data set reflects this difference in abundance. Together, these points suggest that there should not be any substantial bias based on the relative composition of the markers.

ID and expected correlations between f and stMLH

On average, modest numbers of markers seemed to accurately estimate the levels of ID; however, variability was high when only a few markers were considered. Several recent studies have estimated ID for both wild and captive populations (Küpper et al., 2010; Borrell et al., 2011; Grueber et al., 2011; Olano-Marin et al., 2011; Jourdan-Pineau et al., 2012; Wetzel et al., 2012). In all but one case (Olano-Marin et al., 2011), these estimates were nonsignificant, even in the highly endangered takahe (Porphyrio hochstetteri) that had experienced a bottleneck reducing the population to 17 individuals (Grueber et al., 2011). However, all of the studies showing non-significant results used between 7 and 24 microsatellite loci (average 18.2), which we have shown can give an inaccurate picture of diversity, depending on the specific loci examined and the demographic history of the population. In contrast, Olano-Marin et al. (2011) used 80 microsatellites to study a wild population of blue tits (Cyanistes caeruleus).

Time to move towards SNPs for use in HFCs?

Our results suggest that SNPs are more suited for HFCs than previously thought. First, significant correlations between individual stMLH at SNPs and microsatellites indicate that there is no loss of information when using a bi-allelic rather than a multi-allelic marker to estimate heterozygosity. Second, SNPs may be more suited for examining the various hypotheses that underlie HFCs, such as direct effects and local effects (Hansson and Westerberg, 2002), given that they are more abundant in the genome than in microsatellites and can be more readily genotyped at ultra-high density. However, to perform the same job as microsatellites, SNPs need to be more numerous as they are on average less heterozygous. The exact number of markers needed to obtain high correlations depended heavily on the demographic history of the population. For populations such as NBR that have experienced a severe bottleneck or admixture fewer markers will be needed to obtain significant g_2 estimates and detect HFCs. In this situation, it will be more beneficial for researchers to type additional individuals, getting an accurate estimate of the variance in inbreeding and fitness, rather than typing more markers in fewer individuals. For populations with no history of a bottleneck or severe inbreeding, such as RM, significantly more markers will be needed to accurately estimate genome-wide heterozygosity. It is then questionable whether lots of effort should be invested into typing the required number of markers, whatever their type, given that the signal (HFC and inbreeding) is necessarily very weak in such situations. Our equations can be directly used to assess the required number of loci needed to achieve a given accuracy in the measure of inbreeding (or genomic heterozygosity), provided a value of g_2 is available (or can be estimated from preliminary data with fewer loci). For example, in RM using the combined g_2 estimate from all markers (the most precise value available), Equation 5 can be used to predict that no less than 1853 microsatellites or 7053 SNPs would be needed for stMLH to be highly correlated ($r^2 = 0.9$) to inbreeding.

Although SNPs are still moderately expensive to develop for wild species, new methods allow rapid discovery of large marker panels (Baird et al., 2008; Davey et al., 2011) at diminishing costs. Once discovered, new technologies, such as array-based genotyping assays (Shen et al., 2005) and genotype-by-sequencing approaches (Baird et al., 2008), will allow for SNP data sets to be rapidly genotyped in many individuals. Comparable methods for scaling up the genotyping of microsatellites are not currently available.

Although we were unable to directly compare SNPs and microsatellites in terms of their ability to detect HFCs for specific traits because of the small number of individuals genotyped at the sets of SNPs and microsatellites used here, we now have an indication of the number of markers that would be needed for future efforts. More

broadly, our results highlight that accurate calculations of stMLH, assessment of ID, and thereby detection of HFCs will likely require a large number of markers, be they SNPs or microsatellites. However, the exact number is highly dependent on the demographic history (or mating system) of the population being examined, the key parameter being the ID (which can be estimated with g_2). Efforts should be directed towards precisely estimating this parameter in natural populations. To this end, assuming that the number of loci available in population genetic studies will continue to increase, the main limitation will become the sample size in terms of numbers of individuals.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3: e3376.

Balloux F, Amos W, Coulson T (2004). Does heterozygosity estimate inbreeding in real populations? Mol Ecol 13: 3021–3031.

Bierne N, Tsitrone A, David P (2000). An inbreeding model of associative overdominance during a population bottleneck. Genetics 155: 1981–1990.

Bohonak AJ, van der Linde K (2004). RMA: Software for Reduced Major Axis regression, Java version. Website, http://www.kimvdlinde.com/professional/rma.html.

Borrell YJ, Carleos CE, Sánchez JA, Vázquez E, Gallego V, Asturiano JF et al. (2011). Heterozygosity–fitness correlations in the gilthead sea bream Sparus aurata using microsatellite loci from unknown and gene-rich genomic locations. J Fish Biol 79: 1111–1129.

Brandstrom M, Ellegren H (2008). Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. Genome Res 18: 881–887.

Chakraborty R (1981). The distribution of the number of heterozygous loci in an individual in natural-populations. Genetics 98: 461–466.

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

Coltman D, Pilkington J, Smith J, Pemberton J (1999). Parasite-mediated selection against inbred Soay sheep in a free-living, island population. Evolution 53: 1259–1267.

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

Coltman DW, Festa-Bianchet M, Jorgenson JT, Strobeck C (2002). Age-dependent sexual selection in bighorn rams. Proc Biol Sci 269: 165–172.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12: 499–510.

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.

DeWoody Y, DeWoody J (2005). On the estimation of genome-wide heterozygosity using molecular markers. J Heredity 96: 85–88.

Festa-Bianchet M, Coulson T, Gaillard JM, Hogg JT, Pelletier F (2006). Stochastic predation events and population persistence in bighorn sheep. Proc Biol Sci 273: 1537–1543.

Forstmeier W, Schielzeth H, Mueller JC, Ellegren H, Kempenaers B (2012). Heterozygosity–fitness correlations in zebra finches: microsatellite markers can be better than their reputation. Mol Ecol 21: 3237–3249.

Grueber C, Waters J, Jamieson I (2011). The imprecision of heterozygosity-fitness correlations hinders the detection of inbreeding and inbreeding depression in a threatened species. Mol Ecol 20: 67–79.

Grueber CE, Wallis GP, Jamieson IG (2008). Heterozygosity–fitness correlations and their relevance to studies on inbreeding depression in threatened species. Mol Ecol 17: 3978–3984.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O et al. (2011). Current trends in microsatellite genotyping. Mol Ecol Res 11: 591–611.

Hansson B, Westerberg L (2002). On the correlation between heterozygosity and fitness in natural populations. Mol Ecol 11: 2467–2474.

Hogg JT, Forbes SH (1997). Mating in bighorn sheep: frequent male reproduction via a high-risk "unconventional" tactic. Behav Ecol Sociobiol 41: 33–48.

Hogg JT, Forbes SH, Steele BM, Luikart G (2006). Genetic rescue of an insular population of large mammals. Proc Biol Sci 273: 1491–1499.

Jorgenson JT, Festa-Bianchet M, Gaillard J-M, Wishart WD (1997). Effects of age, sex, disease, and density on survival of bighorn sheep. Ecology 78: 1019–1032.

Jourdan-Pineau H, David P, Crochet P-A (2012). Phenotypic plasticity allows the Mediterranean parsley frog Pelodytes punctatus to exploit two temporal niches under continuous gene flow. Mol Ecol 21: 876–886.

Küpper C, Kosztolányi A, Augustin J, Dawson DA, Burke T, Székely T et al. (2010). Heterozygosity-fitness correlations of conserved microsatellite markers in Kentish plovers Charadrius alexandrinus. Mol Ecol 19: 5172–5185.

Ljungqvist M, Åkesson M, Hansson B (2010). Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Väli et al. (2008). Mol Ecol 19: 851–855.

Miller JM, Poissant J, Hogg JT, Coltman DW (2012). Genomic consequences of genetic rescue in an insular population of bighorn sheep (Ovis canadensis). Mol Ecol 21: 1583–1596.

Miller JM, Poissant J, Kijas J, Coltman DW, Consortium TISG (2011). A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Mol Ecol Res 11: 314–322.

Morin PA, Luikart G, Wayne RKGrp SNPW (2004). SNPs in ecology, evolution and conservation. Trends Ecol Evol 19: 208–216.

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

Pilot M, Jedrzejewski W, Branicki W, Sidorovich VE, Jedrzejewska B, Stachura K et al. (2006). Ecological factors influence population genetic structure of European grey wolves. Mol Ecol 15: 4533–4553.

Poissant J, Hogg JT, Davis CS, Miller JM, Maddox JF, Coltman DW et al. (2010). Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorph-ism in bighorn sheep. BMC Genomics 11: 524.

Poissant J, Shafer ABA, Davis CS, Mainguy J, Hogg JT, Côté SD et al. (2009). Genome-wide cross-amplification of domestic sheep microsatellites in bighorn sheep and mountain goats. Mol Ecol Res 9: 1121–1126.

R Development Core Team (2005). R: A Language and Environment for Statistical Computing, Reference Index Version 2.13.0. R Foundation for Statistical Computing: Vienna, Austria.

Rioux-Paquette E, Festa-Bianchet M, Coltman D (2010). No inbreeding avoidance in an isolated population of bighorn sheep. Animal Behav 80: 865–871.

Shen R, Fan J-B, Campbell D, Chang W, Chen J, Doucet D et al. (2005). High-throughput SNP genotyping on universal bead arrays. Mutat Res 573: 70–82.

Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE et al. (2004). Under-standing the relationship between the inbreeding coefficient and multilocus hetero-zygosity: theoretical expectations and empirical data. Heredity 93: 255–265.

Szulkin M, Bierne N, David P (2010). Heterozygosity-fitness correlations: A time for reapprasial. Evolution 64: 1202–1217.

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

Väli Ü, Einarsson A, Waits L, Ellegren H (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Mol Ecol 17: 3808–3817.

Wetzel DP, Stewart IRK, Westneat DF (2012). Heterozygosity predicts clutch and egg size but not plasticity in a house sparrow population with no. Mol Ecol 21: 406–420.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)