

ORIGINAL ARTICLE

Quantifying realized inbreeding in wild and captive animal populations

U Knief¹, G Hemmrich-Stanisak², M Wittig², A Franke², SC Griffith^{3,4}, B Kempenaers¹ and W Forstmeier¹

Most molecular measures of inbreeding do not measure inbreeding at the scale that is most relevant for understanding inbreeding depression—namely the proportion of the genome that is identical-by-descent (IBD). The inbreeding coefficient F_{Ped} obtained from pedigrees is a valuable estimator of IBD, but pedigrees are not always available, and cannot capture inbreeding loops that reach back in time further than the pedigree. We here propose a molecular approach to quantify the realized proportion of the genome that is IBD (propIBD), and we apply this method to a wild and a captive population of zebra finches (*Taeniopygia guttata*). In each of 948 wild and 1057 captive individuals we analyzed available single-nucleotide polymorphism (SNP) data (260 SNPs) spread over four different genomic regions in each population. This allowed us to determine whether any of these four regions was completely homozygous within an individual, which indicates IBD with high confidence. In the highly nomadic wild population, we did not find a single case of IBD, implying that inbreeding must be extremely rare (propIBD = 0–0.00094, 95% CI). In the captive population, a five-generation pedigree strongly underestimated the average amount of realized inbreeding ($F_{\text{Ped}} = 0.013 < \text{propIBD} = 0.064$), as expected given that pedigree founders were already related. We suggest that this SNP-based technique is generally useful for quantifying inbreeding at the individual or population level, and we show analytically that it can capture inbreeding loops that reach back up to a few hundred generations.

Heredity (2015) **114**, 397–403; doi:10.1038/hdy.2014.116; published online 14 January 2015

INTRODUCTION

Inbreeding, the mating of genetically related individuals, leads to genomic regions inherited identical-by-descent (IBD) from both parents to their offspring. Thereby recessive deleterious mutations within those regions become homozygous and fully express their deleterious effects, known as inbreeding depression (Keller and Waller, 2002). Thus, inbred individuals are more often affected by recessive diseases (Campbell *et al.*, 2007), show poorer phenotypic traits and reduced fitness (reviewed by Chapman *et al.*, 2009).

The concept of identity-by-descent is important for understanding inbreeding depression, but it may also be confusing because, in any individual, most base pairs in its genome are homozygous due to common ancestry, often reaching back millions of generations. So, in a sense, most of the genome is 'IBD' (Powell *et al.*, 2010). However, such old coancestry is unproblematic, because it unlikely concerns a recessive deleterious mutation. Most recessive deleterious mutations persist in a population at low allele frequencies (because at higher frequencies they get selected against), and they do not persist for very long times (Li, 1975; Kiezun *et al.*, 2013), because rare alleles are frequently lost from a population even by genetic drift alone. Hence inbreeding depression does result from identity-by-descent, but only from the fraction of coancestry that is more recent than the respective deleterious mutation (Powell *et al.*, 2010). So how can we quantify this recent fraction of inbreeding in the complete absence of knowledge

about deleterious mutations? Traditionally, this is done with pedigree information.

The pedigree-based inbreeding coefficient (F_{Ped}) has often been used to quantify the amount of inbreeding in an individual. F_{Ped} is defined as the probability that the two alleles at any autosomal locus are IBD, or equivalently, as an individual's average proportion of the autosomal genome being IBD (Wright, 1922). Here, IBD is defined with respect to a base population (the pedigree founders) in which all individuals are assumed to be unrelated (Powell *et al.*, 2010). F_{Ped} usefully estimates inbreeding as the proportion of the genome being IBD, but it only captures the most recent inbreeding loops that are included in the pedigree, not the ones from when the common ancestor lived before the pedigree founders. Because of this, F_{Ped} systematically underestimates the amount of IBD that leads to inbreeding depression (Broman and Weber, 1999), since it will miss the causal recessive deleterious mutations inherited twice from a common ancestor that lived before the start of the pedigree (but still after the average deleterious mutation arose).

Several estimates of inbreeding have been proposed that do not rely on pedigree data, but use genetic markers instead (reviewed by Coltman and Slate, 2003). In animal studies, microsatellites have usually been used as molecular markers to quantify individual heterozygosity, for example, as the percentage of loci within an individual that is heterozygous. Likewise, for populations, one can

¹Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Seewiesen, Germany; ²Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; ³Department of Biological Sciences, Macquarie University, Sydney, New South Wales, Australia and ⁴School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, Australia
Correspondence: U Knief, Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Eberhard-Gwinner-Strasse, Starnberg (Seewiesen) 82319, Germany.
E-mail: uknief@orn.mpg.de

Received 13 March 2014; revised 10 November 2014; accepted 14 November 2014; published online 14 January 2015

estimate the percentage of individuals that are heterozygous at any given locus or across a range of loci (multilocus heterozygosity). For instance, a heterozygosity of 80% means that 20% of the individuals are homozygous at a given locus. The disadvantage of this approach is that it remains unclear how many of these 20% are homozygous because of recent inbreeding and how many carry two different copies that are not closely related phylogenetically (that is, allozygous) but only happen to be the same by chance. In other words, a locus can be identical-by-state (IBS), but this does not necessarily mean it is IBD. Hence, IBS gives an approximate upper limit of the proportion of the genome being IBD (here 20%). To distinguish true IBD from IBS occurring due to chance, we need to inspect the information content of the flanking regions that surround a polymorphic site (Broman and Weber, 1999). If a homozygous marker is surrounded by other markers, all of which are homozygous too, then this strongly indicates IBD, because the combined probability of all markers being homozygous by chance becomes very small.

In dense marker panels, genetic regions that are IBD stand out as tracts in which all markers are homozygous, so-called ‘runs of homozygosity’ (ROH; McQuillan *et al.*, 2008) and these can be used to quantify realized inbreeding (Broman and Weber, 1999). Because microsatellites are usually not found at high density in a genome, ROH can be better detected with readily available single-nucleotide polymorphisms (SNPs), even though their allelic richness is lower and they are thus individually less informative (Broman and Weber, 1999). Studies on humans suggest that when more than 50 neighboring SNPs are homozygous, one can safely infer IBD because IBS by chance is too unlikely to ever occur (Powell *et al.*, 2010). Across many generations, recombination between the neighboring markers will lead to the breaking up of haplotypes, making runs of homozygosity shorter, the longer back the common ancestor was (McQuillan *et al.*, 2008). As we will show in this paper analytically, such runs of homozygosity allow us to detect IBD arising from a shared ancestor up to a few hundred generations back, much longer than any pedigree information. We propose that this yields a better measure of inbreeding than F_{Ped} , because it captures more of the relevant inbreeding events, while arguably still being on the safe side in that the majority of recessive deleterious mutations are older than those rather long inbreeding loops that are captured (Li, 1975; Fu *et al.*, 2013).

The aim of the present study is to demonstrate that the ROH-based method can be practically useful for quantifying the realized proportion of the genome that is IBD (propIBD) in both wild and captive animal populations. In the wild, when a study species is highly mobile, it is often impossible to compile a pedigree, so the amount of IBD can only be assessed molecularly. Also, in captivity, it is typically unknown how closely related the pedigree founders were, and by how much F_{Ped} underestimates the levels of IBD that are responsible for inbreeding depression (Ruiz-López *et al.*, 2009).

First, we demonstrate the utility of this molecular method, using two available SNP data sets that had been designed for other purposes.

(1) For a wild population of highly nomadic zebra finches, where no pedigree can be compiled, we use SNP data from an association study (ongoing research by the authors), where 18 candidate genes are being examined. Unfortunately, only four of these genes contained enough SNPs ($n=56-75$) to confidently infer the presence or absence of IBD at a locus within an individual and to exclude that IBS occurred by chance alone. Thus, this data set allows us to take four snapshots for every genotyped individual to assess how frequently IBD occurs in this large and presumably panmictic population (Balakrishnan and Edwards, 2009), whose inbreeding levels are hitherto unknown (Zann, 1996).

(2) For a captive population of zebra finches with a five-generation pedigree, we have data on 1395 SNPs spread widely across the genome that were genotyped for the purpose of quantitative trait locus mapping (Backström *et al.*, 2010). From this data set, we selected four genomic regions with a matching set of 56–75 SNPs, that allow us to infer IBD with confidence. In this case, however, the SNPs are spread over much larger genetic distances than in the candidate gene data set (on average 4 cM vs 0.07 cM in the wild population), such that recombination will break up the haplotypes after a few generations. While this prevents us from detecting IBD *via* long inbreeding loops, the outcome is still sufficiently striking to show the utility of the method.

Because both exemplary data sets comprise only four loci (or genomic regions) per individual, we can only assess the population-wide level of inbreeding, which is relevant, for example, for studies in conservation genetics (Jamieson *et al.*, 2003). However, we note that our method is also suitable to study between-individual variation in inbreeding, provided a sufficient number of genomic regions per individual has been genotyped.

Second, we estimate analytically for how many generations haplotypes persist before being broken up by recombination or altered by mutation. In other words, we estimate the average length of inbreeding loops that can be detected with our method.

MATERIALS AND METHODS

Study populations and sample collection

We collected blood samples from 948 wild adult zebra finches (480 females, 468 males) at Fowler’s Gap, NSW, Australia, in two contiguous places (S 30°57′ E 141° 46′ and S 31°04′ E 141°50′) between October and December 2010 and in April/May 2011. More details on the study sites and catching procedure using a walk-in trap at feeders are given in Griffith *et al.* (2008) and Mariette and Griffith (2012).

For a comparison of inbreeding estimates based on the ROH approach and the pedigree, we used data from 1057 individuals from a captive population held at the Max Planck Institute for Ornithology in Seewiesen, Germany. Pedigree information available for this study covered five generations (Backström *et al.*, 2010; Schielzeth *et al.*, 2011, 2012). We calculated Wright’s inbreeding coefficients using Pedigree Viewer v6.5 (Kingham and Kinghorn, 2010) and averaged them to get an average $F_{\text{Ped}_5\text{gen}}$. Founders (hatched in 2001) were known to be related to each other, showing an average $F_{\text{Ped}_18\text{gen}}$ of 0.030 (calculated using Pedigree Viewer v5.1, for details see Forstmeier *et al.*, 2004), estimated from another 18-generation pedigree for the years 1985–2001 (Forstmeier *et al.*, 2004), which however is not owned by the authors. In a small and closed population, the increase in inbreeding coefficients during the first generations is almost linear (Falconer and Mackay, 1996) so the two average inbreeding coefficients $F_{\text{Ped}_5\text{gen}}$ and $F_{\text{Ped}_18\text{gen}}$ can be summed to obtain a rough estimate of the inbreeding coefficient from a 23-generation pedigree.

SNP genotyping and quality assessment

For the wild population, an initial dense SNP panel (> 20 million SNPs) was discovered by sequencing a pooled non-barcoded sample of equal amounts of whole genomic DNA of 100 from the 948 individuals caught at Fowler’s Gap with the Illumina HiSeq 2000 platform. The whole SNP discovery pipeline is described in the Supplementary Material.

In the course of an association study that will be described elsewhere, we genotyped all 948 wild-caught individuals at 685 SNPs located in 18 genes (10–75 SNPs per gene; Supplementary Table S2 and Supplementary Material) with an Illumina Infinium iSelect HD Custom BeadChip (Illumina Inc., Eindhoven, The Netherlands) on the Illumina iScan platform. Genotype quality was checked for each SNP (clustering of genotype calls, Hardy-Weinberg tests, the occurrence of heterozygous deletions (Ziegler *et al.*, 2010; Gogarten *et al.*, 2012)) and we assessed the possible impact of genotyping errors on our results (for details see the Supplementary Material).

In the captive population, all individuals had previously been genotyped for 1395 SNPs using the Illumina GoldenGate Assay (Fan *et al.*, 2003; Backström *et al.*, 2010). Since this SNP panel was originally designed to cover the whole

genome for quantitative trait locus mapping, the average physical marker spacing was much larger than in the wild birds (mean distance between neighboring SNPs \pm s.d. = 701.5 ± 1117.5 kb vs 1.3 ± 4.6 kb, respectively). The genotype quality of these SNP calls had been checked previously and not a single inheritance error in our five-generation pedigree had been found (Backström *et al.*, 2010). Thus, here we assume no genotyping errors in these samples.

ROH-based estimation of population-level inbreeding

A large enough number of markers will, by chance alone, practically never be homozygous at the same time (Broman and Weber, 1999). Because our SNP data are not phased and haplotype frequencies cannot be established with confidence, we used a simplifying approach to get an estimate of the realized population-level of inbreeding (propIBD).

In the wild population, we first selected genes that were covered with enough SNPs so that IBS would not occur by chance alone. We identified four genes with 56–75 SNPs *via* a selection procedure described in the Supplementary Material. In brief, the procedure estimates the probability that all markers in a gene would be IBS by chance, and we then picked the four genes with the lowest probabilities ($P = 6 \times 10^{-6}$ – 7×10^{-5} , translating into $\mu = 0.01$ – 0.07 individuals expected to be IBS by chance alone). Thus, we expected fewer than one individual to be IBS by chance alone (see Supplementary Material for details). The alternative approach is simply to pick all genes covered with more than 55 SNPs, which yielded the same four genes. Please note that this selection was only needed because our genetic data stem from work that was not specifically designed for estimating population-level inbreeding. If a study is designed specifically for estimating inbreeding, then we recommend genotyping at least 75 SNPs per genomic region (which is the maximum number of SNPs per region used in this study, Supplementary Table S2). Yet the precise number of SNPs needed depends on linkage between SNPs, their allele frequencies and genotyping failure rates.

For each of the four selected regions, we calculated propIBD as the proportion of individuals that were IBS (and hence likely IBD). The population-wide propIBD could in principle be estimated from a single genomic region—provided that a large sample of individuals is used—, because each region should theoretically have the same probability of becoming IBD. Empirically, however, there is variation in IBD among regions in a genome (Weir *et al.*, 2006). Because of that several regions should be used to estimate propIBD to minimize the impact of this variation (which could, for example, stem from a region being located within an inversion polymorphism that is the target of disassortative mating; Thorneycroft, 1975). We calculated confidence limits for our estimated propIBD using Blaker's exact confidence interval (CI) (Blaker, 2000) for a binomial proportion with 0 successes (success = all homozygotes) and 3792 trials (4 regions \times 948 individuals = 3792 trials) as implemented in Scherer (2013).

In the captive population, windows containing an equal number of genotyped SNPs as in the selected four candidate genes and spanning the smallest genetic length were selected on chromosomes *Tgu1* (67 SNPs, spanning 4.6 cM and 655 genes), *Tgu1A* (56 SNPs, spanning 4.0 cM and 476 genes), *Tgu2* (75 SNPs, spanning 3.8 cM and 653 genes) and *Tgu4* (62 SNPs, spanning 4.5 cM and 531 genes). Since these regions span around 4 cM, they will be broken up by cross-over events in roughly 4% of the meioses (1 cM is defined as an expected number of 0.01 cross-over events per meiosis). Hence, the true extent of IBD will be somewhat underestimated because cross-over leads to the loss of IBS for the entire genomic region even when a part of it is still IBD.

To obtain a 95% CI for our estimate of propIBD in captivity, we fitted a generalized linear model with the glm function in R (v2.15.3; R Core Team, 2013). We used counts of completely homozygous individuals vs not completely homozygous individuals for each of the four regions (bound with the cbind function in R) as the response variable, and the intercept as the sole predictor. We specified a quasibinomial error distribution and a logit-link function, because the data were overdispersed. Since we used the logit-link function, we back-transformed (inverse-logit) the intercept and the 95% CI (estimated using the confint function in R), to obtain the estimate of propIBD and its 95% CI (which is equivalent to the average proportion of regions being IBS weighted by their sample sizes; Crawley, 2007).

Persistence of ROH

To get an idea about the persistence of IBD segments over the course of many generations, we estimated the recombination rate per region in both populations from Backström *et al* (2010) and used the average of all four selected regions for our calculations (0.068 cM per 41.6 kb and 4.23 cM per 59.56 Mb for wild and captive populations, respectively). Because the zebra finch exhibits very low recombination rates in the center of its macrochromosomes (0.12 cM/Mb; Backström *et al.*, 2010), which is not representative of most other species, we also estimated the persistence of IBD segments for a more typical example, namely the chicken (*Gallus gallus*). We considered a hypothetical locus that is 65 kb long, such that 65 SNPs can easily be found in that region considering the observed diversity in the chicken genome (Wong *et al.*, 2004). This locus then spans 0.20 cM, given the genome-wide average recombination rate of 3.11 cM/Mb (Groenen *et al.*, 2009). The probability of persistence is then given as $(1 - L/100)^{2 \times G}$ where G is the number of generations and L is the length of a region in cM (see also Hayes *et al.*, 2003). We further assumed a mutation rate of 1.2×10^{-8} per nucleotide per generation (based on studies on humans; Kong *et al.*, 2012). The probability of persistence was then calculated as $(1 - 1.2 \times 10^{-8})^{n_{\text{SNP}} \times 2 \times G}$ where G is the number of generations and n_{SNP} is the average number of SNPs genotyped within one region. The calculations show that the mutation rate had almost no effect on the overall persistence of IBD segments: in the absence of recombination, half of the IBD segments would persist for around 440 000 generations. Hence, uncertainty about the mutation rate in the species of interest is likely to be relatively uncritical.

RESULTS

Estimates of probIBD in our study populations

Figure 1 depicts the average heterozygosity calculated across 56–75 SNPs (depending on the region) for $n = 948$ wild and $n = 1057$ captive zebra finches. These averages are approximately normally distributed, and the left tail of the bell-shaped curve is sufficiently far from zero, indicating that IBS is not expected to occur by chance alone. Hence, individuals that are completely homozygous for a gene region strongly indicate IBD. The proportion of completely homozygous individuals (propIBD) is highlighted for each gene region in Figure 1.

Across the four genes from the wild population, we did not observe a single case that would indicate IBD, suggesting a complete absence of inbreeding (Supplementary Table S2, Figure 1, left panels). It is unlikely that genotyping errors were the cause of the absence of IBD regions, for two reasons. First, all four genes had at least two SNPs that were heterozygous in each individual. Second, those individuals with the least number of heterozygous SNPs per gene had ratios of allelic intensities for the heterozygous SNPs that were in the range of the heterozygous SNPs of the whole population (Supplementary Figure S2). On the basis of these four genes, the estimated population level of propIBD was practically 0 (upper 95% confidence value = 0.00094).

For the captive population, the average pedigree-based inbreeding coefficient was $F_{\text{Ped}, 5\text{gen}} = 0.013$. However, pedigree founders were already related by an average $F_{\text{Ped}, 18\text{gen}}$ of 0.030 (Forstmeier *et al.*, 2004). Thus, the birds from our captive population had an $F_{\text{Ped}, 23\text{gen}}$ of approximately $0.030 + 0.013 = 0.043$. On the basis of the SNPs from four selected genomic regions, the estimated realized propIBD for this population was 0.064 (95% CI = 0.036–0.102) (Supplementary Table S3, Figure 1 right panels).

Persistence of ROH

Our calculations show that recombination events plus *de novo* mutations occur at such a low frequency that it should take 508 generations to break up 50% of the haplotypes that we assessed for IBD in the wild zebra finch population (markers spread over an average genetic length of 0.068 cM). For the captive zebra finch

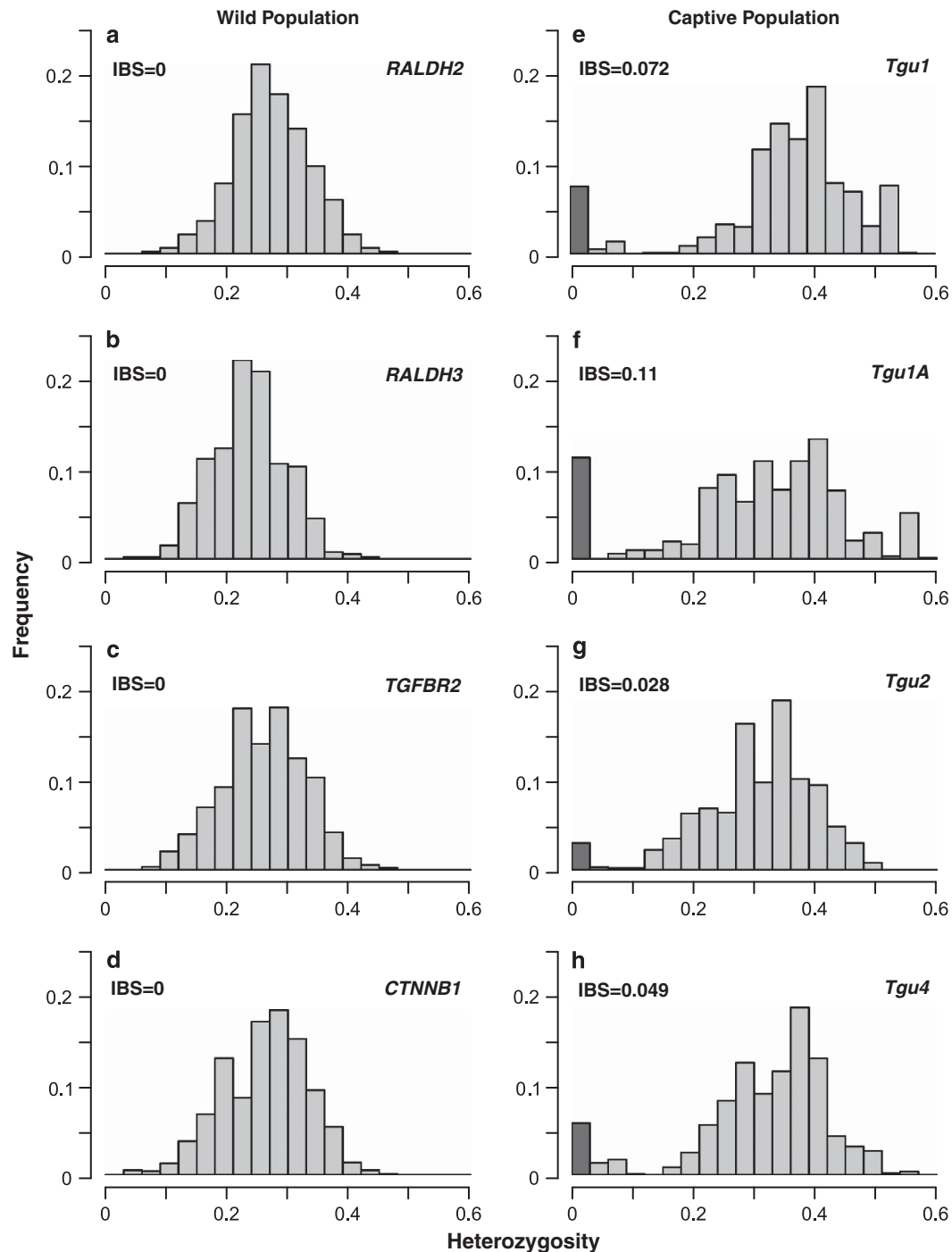


Figure 1 Each histogram shows the mean heterozygosity of every individual in a population for a particular gene region (four regions per population). The wild population ($n=948$ individuals, **a–d**) contains no individual that is fully homozygous in any of the four gene regions (identical-by-state, $IBS=0$). In contrast, between 2.8 and 11.0% of the 1057 individuals of the captive population (**e–h**) are fully homozygous in a given gene region indicating homozygosity by descent and hence inbreeding (dark bars). Gene regions are not the same between the two populations, but were matched for the number of SNPs genotyped per region. $n=67$ SNPs for (a) *RALDH2* and (e) *Tgu1*, $n=56$ SNPs for (b) *RALDH3* and (f) *Tgu1A*, $n=75$ SNPs for (c) *TGFBR2* and (g) *Tgu2*, $n=62$ SNPs for (d) *CTNNB1* and (h) *Tgu4*.

population, however, where our markers were spread over much larger genetic distances (about 4.23 cM), we estimate that 50% of the studied genomic regions would persist for only eight generations. We also estimated the persistence of a hypothetical region in the chicken genome, in which the recombination rate is considerably higher than in the zebra finch (Backström *et al.*, 2010). In such a—potentially

more broadly applicable—hypothetical avian genome, 50% of the 65 kb haplotypes spanning 0.20 cM should persist for 171 generations.

DISCUSSION

We here propose a novel method to estimate inbreeding using ROH of molecular markers without the need for pedigree information, thus

avoiding problems stemming from incomplete pedigree information and relatedness of pedigree founders. The method should not be confused with existing ROH methods that rely on sliding-window approaches to find stretches of markers that are IBS and use miscellaneous methods to discern IBS from IBD (Howrigan *et al.*, 2011). Because these previous methods try to identify every IBD segment within a single genome, they are influenced by variation in linkage disequilibrium and minor allele frequencies of SNPs across windows and tend to overestimate inbreeding when markers are not linkage disequilibrium-pruned before analysis (Polašek *et al.*, 2010). However, this is not the case for our method, because we focus on selected regions that are densely covered with SNPs and hence practically never become homozygous by chance alone (in all four examined regions in the wild zebra finch population together we expect less than 0.2 cases of IBS by chance alone among the $4 \times 948 = 3792$ cases and in the captive population less than 0.7 cases of IBS among the 3958 cases; Supplementary Tables S2 and S3). Thus, theoretically each such region in a genome becomes representative for the population-mean inbreeding and could be used interchangeably to estimate propIBD, if we assume no selection against homozygotes in a region or other special cases like inversion polymorphisms or targets of mate choice. To mitigate errors in population comparisons, normally, the same regions should be used to estimate propIBD in the different populations, which, however, was not possible in the present study because we utilized available SNP data rather than designing the genotyping for our purpose. Yet, reassuringly, all four comparisons (Figure 1) lead to the same conclusion. In the present study, the physical and genetic lengths of the regions were very different between the populations (42 kb and 0.07 cM in the wild vs 60 Mb and 4.2 cM in captivity). As a consequence of using SNPs that were widely spread over long distances, we could only capture rather short inbreeding loops in the captive population, because recombination will have broken up some of the regions studied, leading to an underestimate of propIBD. In that sense, estimates of propIBD are not quite comparable between our two data sets, but the conclusion that inbreeding is much more frequent in the captive than in the wild population only confirms the obvious, and the difference in estimated inbreeding coefficients should be highly conservative.

In the wild population, only four out of 18 genes analyzed could be used to reliably distinguish IBD from cases of IBS occurring by chance alone. These were the four genes with the most SNPs genotyped, emphasizing the need for a dense marker set to reliably infer IBD. Other factors that influence the reliable discrimination between IBD and IBS are population diversity, allele frequencies and linkage between SNPs (Gibson *et al.*, 2006). Australian mainland zebra finches exhibit exceptionally high levels of nucleotide diversity, rapid decay of linkage disequilibrium and high population recombination rates (Balakrishnan and Edwards, 2009), making 56–75 SNPs sufficiently powerful to distinguish IBD from IBS. Although human population demography has been quite different from that of the zebra finch Howrigan *et al.* (2011) suggested that similar marker densities were sufficient for IBD detection in humans.

In our wild study population, not a single individual was completely homozygous in any of the four selected genes, indicating that inbreeding in wild zebra finches is an extremely rare event. With such a low rate of inbreeding, recessive deleterious mutations are not effectively purged and are expected to accumulate in this large panmictic population (Bataillon and Kirkpatrick, 2000). The severe inbreeding depression that has been observed in captive populations

(Bolund *et al.*, 2010; Forstmeier *et al.*, 2012; Hemmings *et al.*, 2012) is in line with such an accumulation of recessive deleterious mutations.

In the captive population, the estimated realized propIBD was 0.064. This value still underestimates the true realized inbreeding because cross-over will have broken up some of the tracts of homozygosity (50% decay after 8 generations). Figure 1 (right panels, especially e and h) shows a few odd cases with heterozygosity < 0.1 but larger than zero (*Tgu1*: $n = 21$, *Tgu1A*: $n = 13$, *Tgu2*: $n = 7$, *Tgu4*: $n = 40$). These might represent IBD segments where just a few of the SNPs had recombined. Consistent with this interpretation, the heterozygous SNPs in those specific cases were not distributed randomly across the examined regions but were concentrated at one of the ends of the regions (data not shown).

In the captive population, there was more variation between regions in the percentage of individuals being IBD than expected by chance (we had to specify a quasi-binomial error distribution in our GLM). Specifically, fewer individuals than expected were IBD for chromosome *Tgu2*. A lack of homozygous individuals for chromosome *Tgu2* had been shown previously in our population (Forstmeier *et al.*, 2007). This could result from non-random mating or be indicative of positive or negative selection. In any case, this emphasizes the need for using multiple regions to estimate population-level inbreeding to ensure against variation in IBD among regions due to evolutionary forces (for example, selection) or structural variants (for example, inversions). In particular, it also illustrates that comparisons of inbreeding levels between populations should normally be based on the same regions in a genome (Weir *et al.*, 2006).

PropIBD was substantially higher than F_{Ped} from a 5-generation pedigree ($F_{Ped, 5gen} = 0.013$). Even when accounting for relatedness from another 18-generation pedigree, $F_{Ped, 23gen} \approx 0.030 + 0.013 = 0.043$ was still lower than the estimated propIBD. This might be surprising because we estimated that 50% of the studied genomic regions would persist for only eight generations. However, both the pedigree and the propIBD estimate may be biased. On the one hand, some individuals have been introduced into the 23-generation pedigree in a later generation, which then are treated as founders, making the pedigree actually shorter and consequently biasing F_{Ped} downwards (that is, 23 generations are the maximum, but not the average length of the pedigree). Furthermore, the founders of the 23-generation pedigree (maintained in a laboratory since 1985 and originating from the population of domesticated birds maintained by aviculturists in the United Kingdom for about one hundred years before that) must have been related to each other to some extent. Indeed, this seems inevitable when founding a captive population from other captive populations. Consequently, F_{Ped} again underestimates the true levels of inbreeding. On the other hand, recombination within a studied region does not always break up the homozygous stretch of SNPs; if the cross-over happens at one end of the studied region the allelic state of the few affected SNPs might not change. Because of such cases, runs of homozygosity may persist for longer than the estimated eight generations. Finally, it should be noted that $F_{Ped, 23gen}$ and propIBD were not significantly different (95% CI overlap).

Our calculations on the persistence of haplotypes assessed for IBD confirmed that our method is able to detect inbreeding loops that reach back in time much further than typical pedigree information obtained from wild animal populations. Even in organisms with high recombination rates like the chicken it should be possible to detect inbreeding loops over more than 100 generations with a sufficiently dense marker panel. From calculations of the mean age of a recessive deleterious allele in a population of constant size (Li, 1975), it is reasonable to assume that in species with a sufficiently large effective

population size the majority of recessive deleterious mutations is much older than 100 generations. We mention this because if most such mutations had arisen only recently, this would undermine the utility of quantifying long inbreeding loops. Instead, this suggests that such long inbreeding loops that reach far back into the past are of importance to study the full extent of inbreeding depression. Thus, our method may be a useful tool in conservation genetics to assess the amount of population-level inbreeding in wild animal populations, even when pedigree information is available. We here show the utility of our method for a large, outbred wild population as well as for a captive population with moderate levels of inbreeding, which could serve as an example for a bottlenecked population under conservation efforts. For future empirical or modelling studies, it would be interesting to assess the utility of our method for populations with high levels of inbreeding, in which the background heterozygosity might not be normally distributed anymore.

Our study suggests that inbreeding can be reliably quantified in a population using ROH based on high-density SNP genotyping without the need for pedigree data. It should be noted that individual variation in inbreeding could also be measured with our method, for example, by genotyping 80 regions each covered with approximately 75 SNPs (that is, a 6k SNP array, yet the exact number of regions and SNPs might depend on the species, research question and marker characteristics). Among others, the false-positive rate of our method decreases with the number of SNPs assessed for a ROH, whereas the false-negative rate increases with the genetic distance covered by the SNPs. Consequently, each region should span only a short genetic distance to detect all relevant stretches of IBD that may cause inbreeding depression.

Only F_{Ped} and ROH-based methods measure inbreeding at the scale that is most relevant for understanding inbreeding depression—namely the proportion of the genome that is IBD. Even if pedigree data are available, the proposed method can identify cases of inbreeding that reach back many more generations than are typically covered by pedigree information. This may be of particular interest because recessive deleterious mutations persist in a population over many more generations than covered by the available pedigrees. High-density SNP genotype data from a large number of individuals are necessary, but these are increasingly becoming available in wild animal populations, for example, through candidate-gene based association studies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Christa Beckmann, Aliza Sager and Mylene Mariette for assistance with sampling the birds. The wild birds were sampled and banded under approval of the Macquarie University Animal Ethics Committee, the Australian Bird and Bat Banding Scheme, and a Scientific License from NSW National Parks and Wildlife Service. We further thank Melanie Schneider for laboratory work in Seewiesen and Markus Schilhabel, the Next-Generation Sequencing team and the Genotyping team at the IKMB in Kiel for laboratory work. UK is part of the International Max Planck Research School for Organismal Biology. This study was funded by the Max Planck Society (BK), with the zebra finch study at Fowler's Gap funded by support to SCG from the Australian Research Council. Genotype data of wild zebra finches are accessible through the Dryad Digital Repository: <http://doi.org/10.5061/dryad.j678b>.

- Backström N, Forstmeier W, Schielzeth H, Møller H, Nam K, Bolund E *et al.* (2010). The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res* **20**: 485–495.
- Balakrishnan CN, Edwards SV (2009). Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the zebra finch (*Taeniopygia guttata*). *Genetics* **181**: 645–660.
- Bataillon T, Kirkpatrick M (2000). Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. *Genet Res* **75**: 75–81.
- Blaker H (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *Can J Stat* **28**: 783–798.
- Bolund E, Martin K, Kempnaers B, Forstmeier W (2010). Inbreeding depression of sexually selected traits and attractiveness in the zebra finch. *Anim Behav* **79**: 947–955.
- Broman KW, Weber JL (1999). Long homozygous chromosomal segments in reference families from the centre d'étude du polymorphisme humain. *Am J Hum Genet* **65**: 1493–1500.
- Campbell H, Carothers AD, Rudan I, Hayward C, Biloglav Z, Barac L *et al.* (2007). Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. *Hum Mol Genet* **16**: 233–241.
- 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 DW, Slate J (2003). Microsatellite measures of inbreeding: a meta-analysis. *Evolution* **57**: 971–983.
- Crawley MJ (2007). *The R Book*. John Wiley & Sons Ltd: Chichester, England.
- Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*. Longman: Harlow, Essex, UK.
- Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL *et al.* (2003). Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* **68**: 69–78.
- Forstmeier W, Coltman DW, Birkhead TR (2004). Maternal effects influence the sexual behavior of sons and daughters in the zebra finch. *Evolution* **58**: 2574–2583.
- Forstmeier W, Schielzeth H, Mueller JC, Ellegren H, Kempnaers B (2012). Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Mol Ecol* **21**: 3237–3249.
- Forstmeier W, Schielzeth H, Schneider M, Kempnaers B (2007). Development of polymorphic microsatellite markers for the zebra finch (*Taeniopygia guttata*). *Mol Ecol Notes* **7**: 1026–1028.
- Fu WQ, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM *et al.* (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**: 216–220.
- Gibson J, Morton NE, Collins A (2006). Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet* **15**: 789–795.
- Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I *et al.* (2012). GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics* **28**: 3329–3331.
- Griffith SC, Pryke SR, Mariette M (2008). Use of nest-boxes by the Zebra Finch (*Taeniopygia guttata*): implications for reproductive success and research. *EMU* **108**: 311–319.
- Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens HJ, Crooijmans RPMA *et al.* (2009). A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Res* **19**: 510–519.
- Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* **13**: 635–643.
- Hemmings NL, Slate J, Birkhead TR (2012). Inbreeding causes early death in a passerine bird. *Nat Commun* **3**: 863.
- Howrigan DP, Simonson MA, Keller MC (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* **12**: 460.
- Jamieson IG, Roy MS, Lettink M (2003). Sex-specific consequences of recent inbreeding in an ancestrally inbred population of New Zealand Takahē. *Conserv Biol* **17**: 708–716.
- Keller LF, Waller DM (2002). Inbreeding effects in wild populations. *Trends Ecol Evol* **17**: 230–241.
- Kiezun A, Pulit SL, Francioli LC, van Dijk F, Swertz M, Boomsma DI *et al.* (2013). Deleterious alleles in the human genome are on average younger than neutral alleles of the same frequency. *PLoS Genet* **9**: e1003301.
- Kinghorn BP, Kinghorn AJ (2010). *Pedigree Viewer 6.5*. University of New England: Armidale, Australia.
- Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G *et al.* (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**: 471–475.
- Li WH (1975). The first arrival time and mean age of a deleterious mutant gene in a finite population. *Am J Hum Genet* **27**: 274–286.
- Mariette MM, Griffith SC (2012). Conspecific attraction and nest site selection in a nomadic species, the zebra finch. *Oikos* **121**: 823–834.
- McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L *et al.* (2008). Runs of homozygosity in European populations. *Am J Hum Genet* **83**: 359–372.
- Polasek O, Hayward C, Bellenguez C, Vitart V, Kolcic I, McQuillan R *et al.* (2010). Comparative assessment of methods for estimating individual genome-wide homozygosity-by-descent from human genomic data. *BMC Genomics* **11**: 139.
- Powell JE, Visscher PM, Goddard ME (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet* **11**: 800–805.

- R Core Team. (2013). *R Foundation for Statistical Computing*. Vienna, Austria.
- Ruiz-López MJ, Roldán ERS, Espeso G, Gomendio M (2009). Pedigrees and microsatellites among endangered ungulates: what do they tell us? *Mol Ecol* **18**: 1352–1364.
- Scherer R (2013). PropCIs: Various confidence interval methods for proportions. *R package version 0.2-4*. <http://CRAN.R-project.org/package=PropCIs>.
- Schielzeth H, Kempenaers B, Ellegren H, Forstmeier W (2012). QTL linkage mapping of zebra finch beak color shows an oligogenic control of a sexually selected trait. *Evolution* **66**: 18–30.
- Schielzeth H, Kempenaers B, Ellegren H, Forstmeier W (2011). Data from: QTL linkage mapping of zebra finch beak color shows an oligogenic control of a sexually selected trait. Dryad Data Repository.
- Thornycroft HB (1975). A cytogenetic study of the white-throated sparrow, *Zonotrichia albicollis* (Gmelin). *Evolution* **29**: 611–621.
- Weir BS, Anderson AD, Hepler AB (2006). Genetic relatedness analysis: modern data and new challenges. *Nat Rev Genet* **7**: 771–780.
- Wong GKS, Liu B, Wang J, Zhang Y, Yang X, Zhang ZJ *et al.* (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432**: 717–722.
- Wright S (1922). Coefficients of inbreeding and relationship. *Am Nat* **56**: 330–338.
- Zann RA (1996). *The Zebra Finch: A Synthesis of Field and Laboratory Studies* Vol 5, Oxford University Press: Oxford, UK.
- Ziegler A, König IR, Pahlke F (2010). *A Statistical Approach to Genetic Epidemiology: Concepts and Applications*. Wiley-VCH: Weinheim, Germany.

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