Performance of marker-based relatedness estimators in natural populations of outbred vertebrates

Katalin Csilléry*, Toby Johnson*, Dario Beraldi*, Tim H. Clutton-Brock†
Dave Coltman‡, Bengt Hansson§, Goran Spong†, Josephine Pemberton*

*Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3JT, United Kingdom
†Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, United Kingdom
‡Department of Biological Sciences, University of Alberta, Edmonton, AB, T6G 2E9, Canada
§Department of Animal Ecology, Lund University, 22362 Lund, Sweden
Abstract

Knowledge of relatedness between pairs of individuals plays an important role in many research areas including evolutionary biology, quantitative genetics and conservation. Pairwise relatedness estimation methods based on genetic data from highly variable molecular markers are now extensively used as a substitute for pedigrees. Although the sampling variance of the estimators have been intensively studied for the most common simple genetic relationships, such as unrelated, half- and full-sib or parent-offspring, little attention has been paid to the average performance of the estimators, by which we mean the performance across all pairs of individuals in a sample. Here we apply two measures to quantify the average performance, first, misclassification rates between pairs of genetic relationships and, second, the proportion of variance explained in the pairwise relatedness estimates by the true population relatedness composition (i.e. the frequencies of different relationships in the population). Using simulated data derived from exceptionally good quality marker and pedigree data from five long-term projects of natural populations, we demonstrate that the average performance mainly depends on the population relatedness composition and may only be improved by the marker data quality only within the limits of the population relatedness composition. Our five examples of vertebrate breeding systems suggest that due to the remarkably low variance in relatedness across the population, marker-based estimates may often have low power to address research questions of interest.
INTRODUCTION

Inferring relatedness among pairs of individuals plays a central role in our understanding of many areas of genetics and population biology. For example, the extent of relatedness between individuals is important in the study of social evolution (e.g. HAMILTON, 1964; CHEVERUD, 1985) and studies incorporating measures of relatedness have influenced our understanding of the mechanism of kin selection in natural populations (e.g. CHOE and CRESPI, 1997). In quantitative genetics the estimation of genetic variance components, allowing estimation of heritability and genetic correlation, requires pairs of individuals with known relatedness (LYNCH and WALSH, 1998). In conservation biology, knowledge of relatedness is essential in captive management, where the goal is to preserve the genetic variation of the wild population from which the founders were drawn (e.g. LACY, 1994). Relatedness estimates are also used when testing hypotheses about inbreeding avoidance (e.g. REUSCH et al., 2001; RICHARDSON et al., 2004) and isolation by distance (e.g. MATOCQ and LACEY, 2004).

Relatedness has traditionally been estimated from pedigrees. Given an outbred source population and good recording, laboratory or managed populations can instantly provide pedigree information. However, many relevant ecological and evolutionary questions can only be addressed in free-living populations with the help of molecular marker data (KRUUK, 2004). When relatedness estimation can be simplified to hypothesis testing over candidate genetic relationships on the basis of some prior life history or partial pedigree information, maximum likelihood methods have been successfully applied (THOMAS, 2005). However, in the absence of prior information, inferences will need to be based solely on marker data. In such cases, with the most commonly available marker numbers (5-20 microsatellite loci), the method of moments estimators are preferred because the ideal properties of the maximum likelihood estimators are only achieved asymptotically, i.e. as the number of loci typed becomes vary large (LYNCH and RITLAND, 1999; WANG, 2002; MILLIGAN, 2003). Thus, the moments estimators developed by QUELLER and GOODNIGHT (1989); LI et al. (1993); RITLAND (1996a); LYNCH and RITLAND (1999); WANG (2002) have became the
most commonly used; hereafter abbreviated QG, L, R, LR, and W, respectively. There is a considerable interest in the performance of these relatedness estimation methods because, in theory, they could make any species accessible for estimating pairwise relatedness (BLOUIN, 2003).

Most previous studies have evaluated the performance of the estimators for the most common first and second order genetic relationships in isolation using Monte Carlo simulations (e.g. LYNCH and RITLAND, 1999; WANG, 2002; VAN DE CASTEELE et al., 2001; MILLIGAN, 2003). These studies, using either theoretical or empirical allele frequency distributions demonstrated that first, the performance of the estimators depends on many factors, including the number of loci and alleles, the shape of the allele frequency distribution, and the relatedness itself (QUELLER and GOODNIGHT, 1989; RITLAND, 1996a; LYNCH and RITLAND, 1999; WANG, 2002; MILLIGAN, 2003), second, estimators generally exhibit a high sampling variance (VAN DE CASTEELE et al., 2001), and third, as a result, the best performing estimators are different depending on the population under investigation (VAN DE CASTEELE et al., 2001).

Although it is useful to know the performance of the estimators for individual relationships, in practice, relatedness is usually estimated among all pairs of individuals in a sample. Thus, we may want to quantify the performance of the estimators across all pairs, i.e. across a range of genetic relationships, what we may call “average performance”. Two measures have been proposed that may be used to measure the average performance.

Although the pairwise relatedness estimators were not developed to classify pairs of individuals to simple genetic relationships BLOUIN et al. (1996) suggested the estimation of the misclassification rate between relationship categories to estimate the error rate if the estimators were used in such way. BLOUIN et al. (1996) defined the misclassification rate between full-sibs and half-sibs as the proportion of pairs that belong to one of the relationships but were classified as the other when using the QG estimator. The midpoint between the two empirically determined means was used as the cutoff point. Error rates were estimated in both directions and used as a performance measure, so that for example, studied as a function of the number of markers. This method has been applied in many recent studies comparing the common first and second order genetic relationships (e.g.
Russello and Amato, 2004; Fraser et al., 2005). Some authors have advocated that the moment estimators can accurately discriminate first order relationships (e.g. Gerlach et al., 2001; Russello and Amato, 2004; Sekino et al., 2004; Fraser et al., 2005).

The other approach was proposed by Van de Casteele et al. (2001), who estimated the proportion of variance explained in the marker-based relatedness estimates by true relatedness, given a population relatedness composition. In the absence of knowledge of the true population relatedness composition they simulated various arbitrary population compositions as a mixture of unrelated, half-sib, full-sib, and parent-offspring pairs. They found that the variance explained by true relatedness was generally high, ranging from 25%-79% (median: 52%) over ten possible relatedness compositions and using the estimator that performed the best for the given set of observed allele frequencies. Of particular note, the $r^2$ was the smallest for the population with the highest proportion of unrelated pairs (60% unrelated). Russello and Amato (2004) applied the same method but tried to estimate the population relatedness composition using a likelihood-based method. The fraction of a particular relationship in the population was estimated as the likelihood of drawing the observed distribution of the relatedness estimates from the sampling distributions of any of the four relationships, unrelated, half-sib, full-sib, parent-offspring. The resulting population compositions were rather similar to those of Van de Casteele et al. (2001), thus, the proportion of variance explained by true relatedness was also high, ranging from 35%-52% (median: 50%) over 11 possible population relatedness compositions.

A major limitation of the previous studies is that the average performance of the relatedness estimators has been investigated without reliable estimates of the true population relatedness composition. This might have led to inaccurate estimates of estimator performance for two reasons. First, the proportions of different relationships may well be different than have been assumed, and second, there may be a non-negligible proportion of higher order relationships in natural populations. In fact, the proportions of highly related pairs were suspiciously high in previous studies; e.g. at least 40% of the pairs had relatedness 0.25 or higher in the simulations of Van de Casteele et al. (2001), which is unlikely to be the case in most natural populations of diploids.
In this paper we assess the average performance of the available method of moments marker-based relatedness estimators using five unique data sets from long term projects of outbreeding vertebrates. Since both deep pedigree and marker data are available we have reliable knowledge of the population relatedness composition as well as the allele frequencies. In particular, we address the point that the population relatedness composition of natural populations may often differ greatly from that assumed by previous investigators (e.g. VAN DE CASTEELE et al., 2001; RUSSELLO and AMATO, 2004). The effect of marker data quality, particularly the number of loci and level of polymorphism, has been emphasised previously (e.g. LYNCH and RITLAND, 1999; WANG, 2002). Here, we also examine the effect of the marker data quality but in the specific context of its importance relative to the population relatedness composition.
METHODS

Observed populations

Long term projects: The starting point for this study is five outbred vertebrate populations, the meerkats, great reed warbler, bighorn sheep, red deer, and Soay sheep, which have each been subjected to intensive individual-based research over several (overlapping) generations. Since the five data sets do differ in some underlying features (e.g. mating system) that may effect performance of estimators, e.g. through influences on relatedness composition of the population, here we provide more details about each species.

Data on meerkats (*Suricata suricatta*) were collected from semi-arid savannah near Vanzylsrus in the Northern Cape of South Africa, where a long term study has been established since 1993 (CLUTTON-BROCK et al., 1998). The study population is spatially continuous and the migration rate is high from the neighbouring areas (i.e. unstudied groups). Meerkats have a nearly monogamous mating system and live in groups of three to 20 adults and sub-adults accompanied by dependent young (CLUTTON-BROCK et al., 1999). Each group is composed of a dominant pair, who produce most, but not all pups, subordinates that were born in the group and a variable number of immigrant males. The life span of meerkats is up to 5-15 years (VAN STAADEN, 1994).

A breeding population of the migratory great reed warbler (*Acrocephalus arundinaceus*) has been studied at Lake Kvismaren in Southern central Sweden since 1983 (HASSELQUIST, 1998; BENSCH, 1996). The population was founded by a few individuals in 1978. Great reed warblers are facultatively socially polygynous with males forming new pair bonds with up to five females each season while about 20% of the males remain unpaired (HASSELQUIST, 1998). The median clutch size is five. Great reed warblers had an average life span of 2.7 years in the study population (HANSSON et al., 2004).

The bighorn sheep (*Ovis canadensis*) population at Ram Mountain, Alberta, Canada has been intensively monitored since 1975 (JORGENSON et al., 1998; COLTMAN et al., 2002). Bighorn sheep are highly polygynous, and rams have up to 22 mating partners through their life time in the
study population. Reproduction is highly seasonal, and after maturation (at age three or four), females producing a single offspring. Bighorn sheep may live for over 10 years (Festa-Bianchet, 1999), but the lifespan of males is significantly shortened by trophy hunting in the study population (Coltman et al., 2003).

Red deer (Cervus elaphus) are the subject of a long-term study in the North Block of the Isle of Rum, Scotland since 1973 (Clutton-Brock et al., 1982). The population was founded by introductions starting in 1845, and was sourced from at least four different British mainland populations. Red deer have a polygynous mating system and males had up to 26 mating partners through their life time in the study population. Reproduction is highly seasonal, females typically become mature when three or four years old and do not necessarily breed every year; if they do they produce a maximum of one calf per year. Average life-span is 10-12 years.

The Soay sheep (Ovis aries) population on the island of Hirta (St. Kilda Archipelago, Scotland) has been intensively studied since 1984 (Clutton-Brock and Pemberton, 2004). The population was founded in 1932 by introduction from a neighbouring island. The Soay sheep is a primitive domestic sheep with a promiscuous, polygynous mating system. Reproduction is seasonal with ewes reproducing once a year. Females start breeding either in the first or second year of their life and produce one or two offspring per year (Clutton-Brock and Pemberton, 2004). The twinning rate is density dependent and fluctuates between 3% and 25%. The life-span is also density dependent and averages to three years.

**Pedigrees:** In all data sets maternity was determined by observation and paternity was assigned using microsatellite genotypes and the likelihood-based paternity inference software CERVUS (Marshall et al., 1998), with the exception of great reed warblers (see Arlt et al. 2004 for details). Pedigrees were further corroborated by identifying parent-offspring mismatches in both maternal and paternal lines and removal of some dubious paternities. In order to justify the comparability of the five pedigrees in terms of their quality, we estimated the depth of the pedigrees by counting the number of ancestors that were present in the pedigree for each individual and averaged over all individuals. Table 1 demonstrates that the pedigrees are similar in quality.
Microsatellite markers: For each population a set of polymorphic unlinked markers was selected (see Table 1). For great reed warbler and Soay sheep a larger set of markers was available, which had been scored for linkage and QTL mapping (Hansson et al. 2005; D. Beraldi, unpublished data). Although many of the markers in these larger data sets are linked, it did not matter for the sake of this study because we only used allele frequency information from these markers for simulating marker genotypes, i.e. as hypothetical sets of unlinked markers.

Simulating populations

In order to simulate realistic populations we estimated the allele frequencies and the relatedness compositions from the five data sets and treated them as known. We estimated the population relatedness composition as the relative frequencies of the relatedness coefficients among all possible pairs in the observed pedigrees. The number of different kinds of relationships was large and many of them were represented by only a few pairs. In order to simplify the analysis, but still keep the complexity of the population relatedness structure, we took a relationship into consideration if it was represented by at least 50 pairs in the sample. In this way less than one percent of the total pairs were lost. Relatedness coefficients were calculated using the kinship package of R (Atkinson, 2005).

Using the observed population allele frequencies 10,000 multilocus genotype pairs were simulated where reference genotypes were drawn randomly according to their Hardy-Weinberg and linkage equilibrium frequencies and genotypes of the pairs were drawn randomly from their conditional genotypic distribution given a particular genetic relationship. All observed genetic relationships were simulated for each of the five populations using the corresponding population allele frequencies. Since any given non-zero relationship can be simulated with more than one pedigree configuration (which greatly increases the number of relationships to be simulated) we assumed, for simplicity, that for a given kind of relationship members of pair were related via a single common ancestor, thus have zero probability of having two genes identical by descent. There was one exception, relatedness 0.5, where we simulated both the parent-offspring and full-sib relationship.
as both of them were present in all five pedigrees.

The marker-based relatedness estimates were calculated for all five published relatedness estimators, but results are presented only for the QG, LR, and W estimators, for simplicity. We choose these estimators because the QG is the most commonly used estimator, the LR estimator is an improved version of the R estimator, and the W can be considered as an improved version of the L estimator. Assumptions of the marker-based relatedness estimators hold in the simulated populations, such that the populations are in Hardy-Weinberg equilibrium, there are no genotyping errors, loci are unlinked and selectively neutral and population allele frequencies are known. Marker-based relatedness estimators were used in their published form (for formulae consult e.g. Wang 2002) and estimates were calculated using functions written in R (R Development Core Team, 2005).

Measuring estimator performance

We quantified the average performance of the estimators using two different measures. First, we used the method suggested by BLOUIN et al. (1996) and calculated the misclassification rates among the unrelated, half-sib, full-sib, and parent-offspring relationships. Error rates were calculated as the proportion of pairs for a given relationship that were misclassified as another relationship or the proportion of pairs that belong to the latter relationship, but classified as the former based on a cutoff point defined as the midpoint between the sampling distributions of the two relationships (subsequently called “naive estimates”). We discourage the use of BLOUIN et al.’s (1996) original terms, Type I and Type II error for the misclassification rates in the two directions because we are not testing a hypothesis of one relationship over another. Then we re-estimated the misclassification rates using knowledge of the population relatedness composition, i.e. knowing that each of the four simple relationships actually incorporate many other higher order relationships. Thus, the sampling distributions of the four relationships were substituted with that of a mixture of higher order relationships observed from the pedigrees. The misclassification rates were determined using the same cutoff points as before since these are what one could determine before the study was
conducted (subsequently called “real estimates”). Comparing the naive and real estimates allowed us to estimate the error of the misclassification rates caused by the assumption that the population is composed of the four simple relationships.

Second, we calculated the proportion of variance explained in the marker-based relatedness estimates by true relatedness (Van de Casteele et al., 2001), for which we used the observed (pedigree-based) population relatedness composition. In order to generate a population of given relatedness composition, 10,000 pairs of different relationships were drawn according to the observed population proportions (see Table 1 for a simplified version of the five population relatedness compositions). The variance explained by true relatedness was estimated as the between group sum of squares divided by the total sum of squares (\(r^2\)) and averaged over 500 realizations of the given relatedness composition.

Any variation in the proportion of variance explained between populations either arises from differences in the population relatedness composition or differences in the marker data (number of loci and/or levels of polymorphism). In order to address how these two factors play a role in driving the proportion of variance explained by true relatedness we analysed in greater detail the great reed warbler and Soay sheep data sets, which turned out to be most and least favourable populations in terms of maximising the \(r^2\), and also the two populations for which the largest sets of marker data were available (Table 1). We randomly selected five different sets of five, ten, 20, 30, 40 marker allele frequencies from the available 101 and 62 in the Soay sheep and great reed warbler populations, respectively and studied \(r^2\) as a function of the number of markers. Since there was variation in the level of polymorphism between markers we also investigated the effect of polymorphism on the average performance. We randomly selected 50 different sets of five markers, calculated the mean number of alleles as a polymorphism measure and compared the effect of polymorphism in the two populations. We choose to investigate the polymorphism effect using sets of five markers instead single locus estimates because, in practice, we are generally interested in multilocus estimates.
RESULTS

A simplified version of the relatedness composition of the five populations illustrates that in all populations the majority of the pairs have relatedness less than 0.25, or in other words less related than e.g. half-sibs, thus would be classified as unrelated using a shallow pedigree (Table 1). The highest proportion of pairs with relatedness equal to or higher than 0.25 were in the meerkat and great reed warbler populations, reflecting the fact that these two populations have the highest number of half-sib and full-sib pairs. Further, the sampling distributions of estimated relatedness for the most common relationships illustrate that the deep pedigrees recovered a large number of relationship categories in all five populations (see Figure 1 for the QG or the LR estimators). The number of relationships is the highest in the meerkat and red deer populations, 347 and 471 respectively, reflecting a higher number of inbreeding events, while it is much less, 72, 103, and 76 in the great reed warbler, Soay sheep, and bighorn sheep populations, respectively. Figure 1 also shows that density curves for the different relationships greatly overlap, especially for the low relationship categories and that only the parent-offspring and full-sib relationships are somewhat distinct from the rest of the relationships, if they are at all. Regarding differences between the two estimators illustrated, the QG has smaller sampling variance for the high relationship categories (density curves are peaked), while when the LR estimator is used the sampling variance is smaller for the low relationship categories. The W estimator is similar to the QG estimator, thus has smaller sampling variance for the high relationship categories (data not shown).

Misclassification rates calculated by assuming that the population comprises only the four relationships, unrelated (UR), half-sib (HS), full-sib (FS), and parent-offspring (PO) (naive estimate) were compared with the misclassification rates calculated using the pedigree, i.e. deconstructing each of the four simple relationship into the observed mixture of higher order relationships (real estimate)(Table 2). Results are illustrated using two estimators, LR and W, and two populations, the bighorn sheep and meerkats, because they represent the best and worse performances, respectively, of BLOUIN et al.’s (1996) misclassification method. Table 2 shows that since relatedness estimates
for the four relationships are lower bounds (relatedness is either that or higher) the misclassification rates to any higher relationships are actually higher (i.e. real estimate is higher than naive; see upper right corners in Table 2) and the misclassification rates to any lower relationships are actually lower (i.e. real estimates are lower than naive; see lower left corners in Table 2). The difference between the naive and real estimates, however, was often very small because the proportion of pairs that are more related than the given relationship is relatively low. This is especially true for the populations of ungulates, e.g. the bighorn sheep (see Table 2; similar figures were obtained for Soay sheep and red deer, results not shown). The difference between the naive and real estimates is the greatest for the meerkats (see Table 2) and great reed warblers (data not shown). Note that these are lower bounds for the biases in the misclassification rate estimates because they are limited by the depth of the observed pedigrees. Regarding the differences between estimators, we found that in all five populations the misclassification rates are the lowest when the LR estimator is used.

We found that the proportion of variance explained in the marker-based relatedness estimates by true relatedness ($r^2$) was generally low, especially in the three populations of ungulates, but two to 14 times higher in the great reed warbler and meerkat populations (Table 3). There is a considerable difference between estimators; noticeably in all five populations the highest proportion of the variance is explained when the LR estimator is used, which reflects the fact that this estimator has the smallest sampling variance for unrelated or low relatedness pairs which are the most common, having over 90% frequency, in all five populations (see Table 1). The W estimator shows the poorest performance in all five populations. Table 3 also highlights two aspects of the populations that are potentially responsible for the between population differences: the number of loci and the variance in relatedness, a summary of the population relatedness composition.

Some of the differences in the $r^2$ among populations may also be contributed to the variation in the number of loci or the quality of the marker data. For example, the great reed warbler and meerkat populations have the most polymorphic markers. We found that in the great reed warbler and Soay sheep populations, using more loci increases the proportion of variance explained, which is expected since with more loci the sampling variance of the marker-based relatedness estimates
decreases (Figure 2). However, there is a striking difference between the two populations; the $r^2$ in the great reed warbler population is nine to 23 times larger than in the Soay sheep population for the studied range of marker number. Figure 2 also shows that the average effect of the number of markers is greater in the great reed warbler population, where adding one locus elevates the $r^2$ by 0.0113 on average, while the equivalent figure is only 0.0014 in the Soay sheep population. Also, in the great reed warbler population there is more variation in the values of $r^2$ for any given number of loci, which probably reflects the fact that there is more variation in the level of polymorphism in the great reed warbler marker set. Results on are shown only for the LR estimator (Figure 2), but the other estimators revealed very similar differences between the two populations (results not shown).

The increased level of polymorphism, expressed as the mean number of alleles, in samples of five loci also had a positive effect on the proportion of variance explained (Figure 3). Note that results are similar using other measures of marker polymorphism, e.g. mean polymorphism information content or heterozygosity (results not shown). Since there is more variation in the great reed warbler markers the two populations can be compared only at the lower end of the polymorphism scale. Again the Soay sheep population has a much lower $r^2$ at all studied polymorphism levels. Increasing the average number of alleles by one increases the $r^2$ in the great reed warbler population by 0.0086 on average, but by only 0.0012 in the Soay sheep population.
DISCUSSION

Relatedness composition of natural populations

Our study demonstrates that in a range of natural populations of vertebrate species the population relatedness composition is different than has been assumed in previous simulation studies and as a result the average performance of marker-based relatedness estimators, defined as the performance across all possible pairs in a sample, is considerably lower than has been previously predicted (e.g. VAN DE CASTEELE et al., 2001; RUSSELLO and AMATO, 2004). We estimated the population relatedness composition from four to six generation deep pedigrees established by long term studies of five species. Analysis of the population relatedness composition reveals that over 90% of the pairs have relatedness less than 0.25, thus would be classified as unrelated using a shallow pedigree, while the proportion of pairs with relatedness of at least 0.5 (e.g. full-sib or parent-offspring pairs) was almost negligible, ranging from 0.1% to 1.7%, as opposed to the 20-50% assumed in previous simulation studies (e.g. VAN DE CASTEELE et al., 2001). Deep pedigrees have also recovered many higher than first and second order relationships, the presence of which may have a non-negligible effect on the average performance of the relatedness estimators, in contrast to previous studies that assumed a simple population composition of unrelated, half-sib, full-sib, and parent-offspring pairs (e.g. BLOUIN et al., 1996).

Our five species represent mating systems ranging from near monogamy to highly skewed polygyny, thus, we argue that our examples of relatedness composition are close to what one would find in many natural populations of vertebrates. Further, if we summarize the population relatedness composition as the variance in relatedness across all pairs, a comparison can be made with a study of monkeyflowers (*Mimulus guttatus*) where the variance in relatedness lies within the range of 0.0025 and 0.01 (RITLAND and RITLAND, 1996), which closely resembles to our estimates (0.0004 - 0.0106). Unfortunately, other examples are scarce in the literature, perhaps because there was no interest in estimating this population parameter (RITLAND, 1996b).
Average performance of relatedness estimators and its consequences

Here, we investigated the consequences of the observed population relatedness composition on the average performance of relatedness estimators by two methods. First, we used the misclassification rate between two relationships described by BLOUIN et al. (1996) and, second, we used the variance explained in the marker-based estimates by true relatedness originally suggested by VAN DE CASTEELE et al. (2001).

The complexity of the relatedness composition of natural populations could lead to misleading estimates of the misclassification rates between the common first and second order relationships. This is because the actual sampling distributions of the first and second order relationships are skewed to the right, towards higher relationships, thus both their sampling variances and means are underestimated. As a result, the optimal cutoff point to classify pairs to different relationships may not be the midpoint between the empirically determined means as has been suggested by BLOUIN et al. (1996), but shifted towards the higher relationships to an extent that is itself dependent on the unknown population relatedness composition. Although this effect will be negligible in some populations (e.g. the bighorn sheep), but in others it will not be (e.g. the meerkats). We would argue that it is generally important to be aware that when one selects highly related pairs from a sample based on BLOUIN et al.’s (1996) method, it is likely that the sample will be diluted with more unrelated pairs than expected. In contrast, we can more confidently select unrelated pairs because the pre-determined error rates will be conservative. These findings are relevant to applications where the aim is to classify pairs to simple first or second order relationships or just to simply select “unrelated” or “related” pairs, e.g. when selecting founders for captive breeding. Unfortunately, endangered species that are selected for captive management are often inbred, and the bias of the pre-determined misclassification rates are expected to be magnified in such scenarios. Finally, we would also generally argue that using the pairwise relatedness estimates as categorical measure should not only be avoided because the misclassification rate depends on the population relatedness composition, but also the estimators themselves are inherently not categorical measures.

The effect of the radical difference between the observed population relatedness composition
and what has previously been assumed is more pronounced on Van de Casteele et al.’s (2001) \( r^2 \) measure, the proportion of variance explained in the marker-based relatedness estimates by true relatedness. This finding is driven by the fact that the low variance in relatedness results in a generally low \( r^2 \). Since increasing the number of markers and/or choosing highly polymorphic markers decreases the sampling variance of relatedness estimates, \( r^2 \) can be improved, but we have shown that even with a hypothetical set of 45 independent, polymorphic microsatellite loci the Soay sheep population, in which the variance in relatedness is an order of magnitude smaller than in great reed warbler population, the \( r^2 \) is on average 10 times smaller. Thus, we suggest that the population relatedness composition sets a limit to the proportion of variance explained in the marker-based relatedness estimates and thus, the average performance may only be improved within the limits of the population relatedness composition.

Knowledge of the proportion of variance explained by true relatedness is essential when pairwise relatedness estimates are used in subsequent analysis as an explanatory variable. When the variance in relatedness is low, as it is expected to be in most natural populations, applications that require the use of relatedness estimates as an explanatory variable will not have sufficient power. This fact closely mirrors recent studies pointing out that there is not sufficient power to detect inbreeding depression in the wild using marker heterozygosity when the variance in inbreeding is low (Slate et al., 2004; Baloux et al., 2004). As a consequence, we argue that some studies might have falsely rejected hypotheses regarding the effect of relatedness. For example, in mate choice experiment with 46 female sticklebacks (Gasterosteus aculeatus) Reusch et al. (2001) claimed to exclude the possibility that preferred males were less related to the females than unpreferred males on the basis of non significant correlation between preference time and pairwise relatedness estimated using seven microsatellite markers (Reusch et al., 2001). As another example, a study found that eider ducks (Somateria mollissima) form non-kin brood-rearing coalitions, thus rejected the kin selection hypothesis on the basis of comparing the relatedness estimates of 24 pairs of brood-rearing females with 24 randomly drawn pairs of females using 6-8 microsatellite markers (Öst et al., 2005).
Pairwise relatedness estimates across all pairs in a sample are again used as the independent variable when applying Ritland’s (1996b) method, which has been developed to estimate quantitative genetic parameters in natural populations. The method was published over 10 years ago, but remarkably few applications have appeared in the literature since, and apparently, most of them are using one of the data sets presented in this paper. As an example, studies comparing heritability estimates based on Ritland’s (1996b) method and traditional pedigree-based methods found that marker-based estimates erratically either under or overestimate the pedigree-based estimates of heritability (e.g. Thomas et al., 2002; Wilson et al., 2003; Colman, 2005), for which the low variance in relatedness and the inaccurate estimate of it should at least partly be responsible. We suspect that there are many more out of range, biologically not sensible, unpublished results.

**Improving the average performance**

If the average performance of the relatedness estimators in natural populations is generally expected to be low, the question arises how to improve it or what alternative methods are available. Here we discuss the choice of relatedness estimator, the importance of marker data quality, the potential choice of study population and/or organism, and the combined use of marker data and pedigrees.

The most popular question is undoubtedly the choice of estimator (Van de Castele et al., 2001). Given the observed population relatedness compositions our study uniformly supports the use of the LR estimator. In all five populations the sampling variance is the lowest when using the LR (and R, data not shown) estimator for the low relationship categories, and since in all studied populations most pairs have relatedness less than 0.25, on average, across all pairs the LR estimator minimises the sampling variance. Thus, in terms of average performance the Lynch and Ritland estimator is recommended on the bases of these five populations and some other published studies as well (Rusello and Amato, 2004; Colman, 2005). In contrast, the most recent simulation study of the moment estimators demonstrated that the single locus sampling variances for the QG, L, and W estimators asymptotically approach the minimum sampling variances (variance in identity-by-descent) with increasing number of alleles, but not when the R and LR estimators are
used (WANG, 2002), because the latter two estimators assume null relationship when calculating weights. WANG (2002) also pointed out that LR and R estimators are sensitive to sample sizes both in terms of variance and bias. In summary, however, when we look at the differences between estimators in terms of e.g. the $r^2$, they are almost negligible in relation to the differences between populations, suggesting that the estimator choice may not be the most crucial question.

The utmost importance of good quality marker data have been emphasised by many, for example, RITLAND (1996b) suggested that for populations with weaker structure simply more polymorphic markers are needed and WANG (2002) concluded that estimators asymptotically approach the minimum sampling variances with increasing number of alleles. In contrast, we argue that only if the population has a high variance in relatedness acquiring more markers and more polymorphic markers deliver substantial improvements in the average performance of the relatedness estimators. In such cases, it is worthwhile to type as many markers as possible in order to achieve the best possible performance. The number of markers available, however, may well be limited by the number of chromosomes, because some of the markers will unavoidably be linked. We argue that the use of linked markers could be useful because estimators should only lose efficiency relative to the same number of unlinked markers. This is because linked markers simply carry less information about identity by descent (THOMPSON, 1986). Regarding the level of polymorphism, which we have even less control over, is generally specific to the species; e.g. mammals were found to have less polymorphic markers than birds in a comparative study using AC repeats (NEFF and GROSS, 2001).

Recognising that the population relatedness composition plays the major role in the average performance of the estimators one may choose to address questions that require the knowledge of relatedness in study organisms where the expected estimator performance is high. RITLAND (1996b) has also suggested selecting study population or taxa where the variance in relatedness is expected to be high and described two potentially favourable situations (RITLAND, 2000). One of them is where one polygynous male is breeding within a lineage of philopatric females, a common breeding system in many mammalian social systems. Unfortunately, our three populations of un-
gulates are like this, but our results show that the variance in relatedness is rather low. RITLAND’s (2000) other example is newly founded populations, with a small number of related founders. The fact that the great reed warbler population is recently founded by a few individuals might have played a role in the fact that this population has a relatively high variance in relatedness, but perhaps it is more likely to be the result of the mating system of the species. On the basis of our five populations we rather recommend using information on the mating system of the species to predict the population relatedness composition and prefer species with large full-sib families. More specifically, one may want to consider using monogamous birds with a large clutch size.

Traditional, pedigree-based methods supported by marker aided parentage inference, where required, offer a good alternative in many applications, for example, when the aim is to classify pairs to different relationships or to estimate quantitative genetic parameters in natural populations (KRUUK, 2004; THOMAS, 2005). However, when relatedness estimates are used as an explanatory variable and the variance in relatedness is low in the study population, even knowledge of the pedigree cannot directly help. In such cases pedigrees may be employed to aid the marker-based relatedness estimation. Even partial or shallow pedigrees could be used to selectively sample highly related pairs or families and, thus artificially generate a population with more favourable population relatedness composition. When only shallow pedigrees are available this is perhaps the preferred method, because marker-based estimates could potentially be more accurate than categorical measures of relatedness, assuming good marker data. This is because marker-based methods estimate the actual relatedness between two individuals, which is the realized relationship and not the mere expectation that the categorical measures estimates of relatedness, like pedigrees, provide (THOMAS, 2005).

Acknowledgements: We are grateful to all who organised and raised funds over many years for the five long term projects, especially to Staffan Bensch, Marco Festa-Bianchet, Dennis Hasselquist. We also thank all field workers and genotypers. The ideas presented in here have been improved by discussions with Bill Hill, Penny Kukuk, Allen Moore, Jon Slate, and Alastair Wilson. K. C. is
funded by a Principal’s Studentship from the University of Edinburgh.
Literature Cited


RICHARDSON, D. S., J. KOMDEUR and T. BURKE, 2004 Inbreeding in the Seychelles warbler: environment-dependent maternal effects. Evolution 58: 2037–2048.


Table 1: Summary of the microsatellite markers and pedigrees of the five data sets from natural populations. Marker data quality parameters for the great reed warbler and Soay sheep populations in parenthesis correspond to the mapping data sets (see text for details). Pedigree depth was estimated by counting the number of ancestors that were present in the pedigree for each individual, averaged over all individuals. The population relatedness composition based on the pedigrees is, for simplicity, summarized as proportion of pairs that would be classified as unrelated, half-sib, or full-sib (parent-offspring) if only a two generation deep pedigree was available. Note that this is great simplification of the true relatedness composition (see text and Figure 1 for details).

<table>
<thead>
<tr>
<th></th>
<th>Meerkat</th>
<th>Great reed warbler</th>
<th>Bighorn sheep</th>
<th>Red deer</th>
<th>Soay sheep</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Marker data quality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of loci</td>
<td>8</td>
<td>15 (62)</td>
<td>30</td>
<td>8</td>
<td>19 (101)</td>
</tr>
<tr>
<td>Median of number of alleles</td>
<td>12.5</td>
<td>10 (5)</td>
<td>5</td>
<td>9</td>
<td>6 (5)</td>
</tr>
<tr>
<td>Range of number of alleles</td>
<td>5-17</td>
<td>4-26 (2-78)</td>
<td>3-10</td>
<td>6-10</td>
<td>4-8 (3-8)</td>
</tr>
<tr>
<td><strong>Pedigree quality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of individuals</td>
<td>792</td>
<td>802</td>
<td>910</td>
<td>3544</td>
<td>3550</td>
</tr>
<tr>
<td>Number of individuals with both parents known</td>
<td>586</td>
<td>706</td>
<td>382</td>
<td>1003</td>
<td>1720</td>
</tr>
<tr>
<td>Pedigree depth</td>
<td>6.3</td>
<td>5.4</td>
<td>4.6</td>
<td>4.6</td>
<td>5.9</td>
</tr>
<tr>
<td><strong>Simplified relatedness composition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentages of pairs with relatedness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- less than 0.25</td>
<td>90.8</td>
<td>96.6</td>
<td>98.9</td>
<td>99.6</td>
<td>99.7</td>
</tr>
<tr>
<td>- less than 0.5, but at least 0.25</td>
<td>7.5</td>
<td>2.5</td>
<td>0.9</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>- at least 0.5</td>
<td>1.7</td>
<td>0.9</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Table 2: Misclassification rates between pairs of genetic relationships illustrated by the W and LR estimators and for the bighorn sheep and meerkat populations. Four relationships are shown on each axis, unrelated (UR), half-sib (HS), full-sib (FS), and parent-offspring (PO). Row headers indicate true relationship and column headers indicate the relationship into which pairs were misclassified. Each cell of the table compares misclassification rates calculated in two different ways, the first value is the “naive estimate” and the second value is the “real estimate” (see Methods for details). Both values are means of 500 independent draws of the population relatedness compositions from 10000 simulated pairs for each genetic relationships.

<table>
<thead>
<tr>
<th>True relationship</th>
<th>Misclassified as</th>
<th>UR</th>
<th>HS</th>
<th>FS</th>
<th>PO</th>
</tr>
</thead>
<tbody>
<tr>
<td>UR</td>
<td>-</td>
<td>0.17 / 0.18</td>
<td>0.027 / 0.031</td>
<td>0.027 / 0.031</td>
<td></td>
</tr>
<tr>
<td>HS</td>
<td>0.145 / 0.142</td>
<td>-</td>
<td>0.147 / 0.151</td>
<td>0.144 / 0.149</td>
<td></td>
</tr>
<tr>
<td>FS</td>
<td>0.015 / 0.015</td>
<td>0.132 / 0.132</td>
<td>-</td>
<td>0.489 / 0.489</td>
<td></td>
</tr>
<tr>
<td>PO</td>
<td>0 / 0</td>
<td>0.045 / 0.044</td>
<td>0.496 / 0.491</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

W estimator, bighorn sheep population
### W estimator, meerkat population

<table>
<thead>
<tr>
<th>Misclassified as</th>
<th>UR</th>
<th>HS</th>
<th>FS</th>
<th>PO</th>
</tr>
</thead>
<tbody>
<tr>
<td>UR -</td>
<td>0.311 / 0.366</td>
<td>0.143 / 0.186</td>
<td>0.143 / 0.187</td>
<td></td>
</tr>
<tr>
<td>HS 0.261 / 0.217</td>
<td>-</td>
<td>0.285 / 0.339</td>
<td>0.287 / 0.341</td>
<td></td>
</tr>
<tr>
<td>FS 0.103 / 0.103</td>
<td>0.246 / 0.246</td>
<td>-</td>
<td>0.479 / 0.479</td>
<td></td>
</tr>
<tr>
<td>PO 0.017 / 0.017</td>
<td>0.146 / 0.146</td>
<td>0.513 / 0.511</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

### LR estimator, bighorn sheep population

<table>
<thead>
<tr>
<th>Misclassified as</th>
<th>UR</th>
<th>HS</th>
<th>FS</th>
<th>PO</th>
</tr>
</thead>
<tbody>
<tr>
<td>UR -</td>
<td>0.087 / 0.098</td>
<td>0.009 / 0.013</td>
<td>0.009 / 0.013</td>
<td></td>
</tr>
<tr>
<td>HS 0.183 / 0.18</td>
<td>-</td>
<td>0.176 / 0.179</td>
<td>0.177 / 0.181</td>
<td></td>
</tr>
<tr>
<td>FS 0.044 / 0.044</td>
<td>0.208 / 0.208</td>
<td>-</td>
<td>0.503 / 0.503</td>
<td></td>
</tr>
<tr>
<td>PO 0.021 / 0.022</td>
<td>0.177 / 0.178</td>
<td>0.52 / 0.522</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

### LR estimator, meerkat population

<table>
<thead>
<tr>
<th>Misclassified as</th>
<th>UR</th>
<th>HS</th>
<th>FS</th>
<th>PO</th>
</tr>
</thead>
<tbody>
<tr>
<td>UR -</td>
<td>0.142 / 0.208</td>
<td>0.054 / 0.097</td>
<td>0.054 / 0.098</td>
<td></td>
</tr>
<tr>
<td>HS 0.37 / 0.321</td>
<td>-</td>
<td>0.269 / 0.313</td>
<td>0.271 / 0.315</td>
<td></td>
</tr>
<tr>
<td>FS 0.199 / 0.199</td>
<td>0.342 / 0.341</td>
<td>-</td>
<td>0.496 / 0.496</td>
<td></td>
</tr>
<tr>
<td>PO 0.151 / 0.152</td>
<td>0.321 / 0.322</td>
<td>0.516 / 0.516</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Proportion of variance explained in marker-based relatedness estimates by true relatedness in simulated populations based on the observed relatedness composition and allele frequencies of five natural population samples. The population relatedness composition is summarized as the variance in relatedness of all pairs. Values are means of 500 independent draws of the population relatedness compositions from 10000 simulated pairs for each genetic relationship.

<table>
<thead>
<tr>
<th></th>
<th>Meerkat</th>
<th>Great reed warbler</th>
<th>Bighorn sheep</th>
<th>Red deer</th>
<th>Soay sheep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance in relatedness</td>
<td>0.0106</td>
<td>0.0044</td>
<td>0.0015</td>
<td>0.0005</td>
<td>0.0004</td>
</tr>
<tr>
<td>Number of loci</td>
<td>8</td>
<td>15</td>
<td>30</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>Variance explained</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QG</td>
<td>0.176</td>
<td>0.172</td>
<td>0.086</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>LR</td>
<td>0.269</td>
<td>0.328</td>
<td>0.14</td>
<td>0.024</td>
<td>0.028</td>
</tr>
<tr>
<td>W</td>
<td>0.161</td>
<td>0.167</td>
<td>0.074</td>
<td>0.013</td>
<td>0.013</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1. Sampling distributions of the most common pedigree-based relationships in five natural populations when pairwise relatedness is calculated using the QG or the LR estimators. Each density curve is based on 10,000 simulated pairs of a given genetic relationship. Observed population allele frequencies were used. The plotted relationships were the most common in the observed pedigrees.

Figure 2. Proportion of variance explained in the marker-based relatedness estimates by true relatedness as a function of the number of loci. Populations were simulated based on the observed relatedness composition and allele frequencies of five natural population samples. For each number of loci five different loci were drawn from the available set of markers, which consisted of 62 loci for the great reed warbler and 101 for the Soay sheep. Relatedness was estimated using the LR estimator.

Figure 3. Proportion of variance explained in the marker-based relatedness estimates by true relatedness as a function of the level of polymorphism expressed as the mean number of alleles at five randomly selected loci. Populations were simulated based on the observed relatedness composition and allele frequencies of five natural population samples. Loci were drawn from the available set of markers, which consisted of 62 loci for the great reed warbler and 101 for the Soay sheep. Relatedness was estimated using the LR estimator.
Figure 1:
Figure 2:
Figure 3: