Molecular Ecology Resources (2008) **8**, 256–263 doi: 10.1111/j.1471-8286.2007.01940.x

TECHNICAL ADVANCES

A maximum-likelihood relatedness estimator allowing for negative relatedness values

DMITRY A. KONOVALOV* and DIK HEG†
*School of Mathematics, Physics & Information Technology, James Cook University, Townsville, Queensland 4811, Australia,
†Department of Behavioural Ecology, Zoological Institute, University of Bern, Hinterkappelen, Switzerland

Abstract

Previously reported maximum-likelihood pairwise relatedness (*r*) estimator of Thompson and Milligan (M) was extended to allow for negative *r* estimates under the regression interpretation of *r*. This was achieved by establishing the equivalency of the likelihoods used in the KINSHIP program and the likelihoods of Thompson. The new maximum-likelihood (ML) estimator was evaluated by Monte Carlo simulations. It was found that the new ML estimator became unbiased significantly faster compared to the original M estimator when the amount of genotype information was increased. The effects of allele frequency estimation errors on the new and existing relatedness estimators were also considered.

Keywords: allele frequency errors, codominant molecular markers, maximum likelihood, relatedness

Received 12 April 2007; revision accepted 4 July 2007

Introduction

The widespread use of highly polymorphic codominant genetic markers such as microsatellite loci attracted considerable interest to marker-based kinship analysis in populations when prior pedigree information is not available. The concept of pairwise relatedness is commonly used in such analysis (e.g. Blouin 2003). Although there exist a number of definitions and interpretations of relatedness (Rousset 2002), in this study we focus on one aspect of the available interpretations: whether negative relatedness coefficient (*r*) as a measure of degree of relatedness is permitted or not; and how negative estimates of *r* are possible within the maximum-likelihood framework. The subject of possible interpretation of *r* in terms of what relatedness coefficient *r* actually measures was extensively dealt with by Rousset (2002). Therefore, we introduce the subject only briefly and only in relation to why the negative estimates of *r* do not contradict the concept of relatedness (see also Wright 1965; Hamilton 1972).

Historically, *r* was conceived as a measure of genetic similarity due to recent shared ancestry given by known pedigree relationships (Rousset 2002). In particular, the coefficient of coancestry $\theta = r/2$ (also known as coefficient of consanguinity) between two individuals (dyad) could be defined as the probability that two alleles, one chosen randomly from each individual, are identical by descent (Lynch & Walsh 1998; Blouin 2003). When neither diploid individual is inbred, the coefficient of relatedness ($r = 2\theta$) could be interpreted as the expected fraction of alleles that are shared identical by descent (Pamilo 1990; Blouin 2003), constraining *r* to the range of [0, 1] and θ to the range of [0, 0.5]. This interpretation will be referred to as the identity by descent (IBD) throughout this study. The IBD interpretation leads to a number of inconsistencies (Rousset 2002), e.g. given a sample of unrelated individuals, any *unbiased* estimator of *r* must produce negative estimations of IBD probability for half (on average) of all sample dyads since the mean value of pairwise *r* must be zero by the definition of the unrelated dyads.

Fortunately, the IBD interpretation is not the only possible interpretation of *r*. Another interpretation (will be referred to as the *regression* interpretation) allows negative *r* estimates and could be understood as *r* measuring how much higher (or lower for negative *r*) the probability of recent coalescence is for a dyad relative to the average probability for all considered dyads or dyads from the reference population (Rousset 2002). The

Correspondence: Dr Dmitry Konovalov, Fax: +617 4781 4029; E-mail: dmitry.konovalov@jcu.edu.au

regression interpretation could be used to interpret negative *r* estimates obtained by the QG (Queller & Goodnight 1989) estimator and such method-of-moments (MoM) relatedness estimators as the LR (Lynch & Ritland 1999) and W (Wang 2002) estimators. Note that the IBD interpretation was used in the maximum-likelihood relatedness estimator (Thompson 1975; Milligan 2003). The estimator will be referred to as the M estimator, after Milligan (2003).

Given that the commonly used QG, LR and W relatedness estimators (e.g. Csillery *et al*. 2006) are unbiased by allowing for the negative *r* estimates and given the validity of such negative values under the regression interpretation (Rousset 2002), the existing maximum-likelihood estimator (Thompson 1975; Milligan 2003) stands out by not permitting the negative *r* estimates. This situation is unsatisfactory at least from the academic point of view, when one would expect for an estimator to be unbiased first before looking into any other statistical properties of the estimator, e.g. see arguments for (and derivation of) the unbiased estimator of variance in most statistics textbooks. By forcing *r* to be non-negative, the M estimator exhibits the larger bias the less related the two individuals are (Milligan 2003). Note that Thomas (2005) had already noticed that constraining *r* to [0, 1] disadvantages the M estimator since the MoM estimators allow negative values of *r* and truncating the M estimates simply introduces positive bias and offers few benefits. Moreover, negative *r* does have a biological meaning by saying that two individuals share *fewer* alleles than expected on the basis of the corresponding Hardy–Weinberg allele frequencies (Queller & Goodnight 1989), i.e. fewer identical-in-state alleles than expected. For example, allowing for negative *r* values, a pairwise *r* frequency plot of unrelated individuals has a normal distribution with zero mean (e.g. Heg & van Treuren 1998). Furthermore, negative value may carry a very practical biological meaning: (i) the more negative the *r* estimate, the more confident we can be that two individuals are unrelated, and (ii) it may allow detection of recent immigrants into the population carrying novel alleles. Hence, the regression interpretation of *r* must be considered, although it has not been known until this study how negative *r* estimates were possible within the maximum-likelihood framework.

The first objective of this study is to define a new maximum-likelihood (ML) estimator of relatedness that is not constrained to the non-negative *r* estimates, thus reducing the bias reported by Milligan (2003). This objective is achieved by establishing (for the first time to our knowledge) the equivalency of the likelihoods used in the KINSHIP (Goodnight & Queller 1999) program and the likelihoods of Thompson (1975) (under certain conditions), and then extending the KINSHIP likelihoods under the regression interpretation of relatedness.

The second objective is to study the effects of uncertainties in estimated allele frequencies on the relatedness estimators. This is important because the considered estimators rely on a population sample of alleles to compare the focal genotypes against, and thus becoming biased at small population sizes (Wang 2002). Preferably, the population sample should be taken independently from the focal genotypes containing only single-generation individuals and should be sufficiently large to capture all alleles present in the population with sufficient accuracy, which is rarely possible in practice.

Materials and methods

Relatedness estimators

Let outbred diploid individuals *X* and *Y* be genotyped at *L* unlinked loci, where (*a*, *b*) and (*c*, *d*) denote their respective codominant alleles at a single locus. Queller & Goodnight (1989) presented several versions of their QG estimator. One such version, which will be referred to by the same 'QG' abbreviation, could be defined by

$$r_{QG}(X, Y) = \frac{\sum_{l=1}^{L} (\delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd} - p_a - p_b - p_c - p_d)}{\sum_{l=1}^{L} (2 + \delta_{ab} + \delta_{cd} - p_a - p_b - p_c - p_d)}, \quad \text{(eqn 1)}$$

where δ_{xy} is the Kronecker delta defined as $\delta_{x,x} = 1$ and $\delta_{x,y\neq x} = 0$, and where the (*a*, *b*, *c*, *d*) alleles together with the corresponding reference population allelic frequencies (p_a, p_b, p_c, p_d) are locus-specific. This version of the QG estimator is the one implemented in the KINSHIP (Goodnight & Queller 1999) program. Lynch & Ritland (1999), Milligan (2003) and Wang (2002) tested their proposed estimators against the following version (denoted by QG2) which is different from the QG version: $r_{QG2}(X, Y) = 1/2(r'_{QG}(X, Y) + r'_{QG}(Y, X))$, where the asymmetric contribution $r'_{QG}(X, Y)$ is given by $r'_{QG}(X, Y) = (1/L)\sum_{l=1}^{L}(0.5(\delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd}) - p_a - p_b)/(1 + \delta_{ab} - p_a - p_b)$. Note that the QG version is symmetric to the interchange of $X \leftrightarrow Y$ and does not need the additional symmetrization.

The symmetric LR (Lynch & Ritland 1999) estimator is defined by $r_{LR}(X, Y) = 1/2(r'_{LR}(X, Y) + r'_{LR}(Y, X))$ from the asymmetric estimator, $r'_{LR}(X, Y) = 1/W \sum_{l=1}^{L}(p_a(\delta_{bc} + \delta_{bd}) + p_b(\delta_{ac} + \delta_{ad}) - 4p_a p_b)/(2p_a p_b)$, where $W = \sum_{l=1}^{L} w_l$ and $w_l = ((1 + \delta_{ab})(p_a + p_b) - 4p_a p_b)/2p_a p_b$. This expression is mathematically identical to the original from Lynch & Ritland (1999) but was reported as numerically more stable by Oliehoek *et al*. (2006). Another benefit of the expression is that it clearly reveals that the estimator is highly sensitive to the frequencies of *rare* alleles which, if present, may yield almost arbitrary values of *r*, depending on the procedure adopted for estimating the population frequencies of rare alleles (Wang 2002).

New maximum-likelihood estimator

Let us examine two parameterizations of the likelihoods for pairwise kinship relationships. The first is via the $k = (k_0, k_1, k_2)$ IBD coefficients (e.g. Blouin 2003), where k_m is the probability that a genotyped dyad shares m alleles that are identical by descent, and where $k_0 + k_1 + k_2 = 1$. The explicit expressions for the corresponding pairwise likelihoods were given by Thompson (1975) and Milligan (2003) and will be referred to as the Thompson likelihoods. The relatedness is then expressed via $r = k_1/2 + k_2$, where the k coefficients represent a special case of the $\Delta = \{\Delta_1, \Delta_2, \ldots, \Delta_9\}$ (Jacquard 1974) coefficients for outbred populations: $\Delta_1 = \ldots = \Delta_6 = 0, \Delta_7 = k_2, \Delta_8 = k_1$ and $\Delta_9 = k_0$ (e.g. Blouin 2003). The second parameterization is via the (R_m, R_p) coefficients, where the R_m parameter was defined as the probability that the maternal alleles in a diploid dyad are identical by descent, and R_p representing the same for the paternal alleles (Goodnight & Queller 1999). The following are the (R_m, R_p) values for some common kinship relationships: full-siblings ($R_m = 0.5$, $R_p = 0.5$), half-siblings ($R_m = 0.5$, $R_p = 0$) or ($R_m = 0$, $R_p = 0.5$), unrelated ($R_m = 0$, $R_p = 0$) and parent–offspring ($R_m = 1$, $R_p = 0$) or ($R_m = 0$, $R_p = 1$) The second parameterization was utilized in the KINSHIP (Goodnight & Queller 1999), KINGROUP (Konovalov *et al*. 2004), FAMILYFINDER (Beyer & May 2003) and ML-RELATE (Kalinowski *et al*. 2006) programs and will be referred to as the KINSHIP parameterization or the KINSHIP likelihoods. For diploids, the pairwise KINSHIP (Goodnight & Queller 1999) likelihoods are given by

$$L(a, b; c, d \mid R_m, R_p) = L(a, c \mid R_m) \times L(b, d \mid R_p), \qquad \text{(eqn 2)}$$

where (a, c) and (b, d) are maternal and paternal alleles of $X = (a, c)$ and $Y = (c, d)$ genotypes, respectively, and where $L(a, c \mid R_m) = p_a(R_m\delta_{a,c} + (1 - R_m)p_c)$ and $L(b, d \mid R_p) = p_b(R_p\delta_{b,d} + (1 - R_p)p_d)$. If maternal and paternal genotypes of the X and Y individuals are not known, the likelihoods are further summed over all possible permutations of the alleles within the same locus,

$$\begin{aligned} L(X, Y \mid R_m, R_p) = {} & L(a, b; c, d \mid R_m, R_p) \\ & + (1 - \delta_{c,d})L(a, b; d, c \mid R_m, R_p) \\ & + (1 - \delta_{a,b})L(b, a; c, d \mid R_m, R_p) \\ & + (1 - \delta_{a,b})(1 - \delta_{c,d})L(b, a; d, c \mid R_m, R_p). \end{aligned} \qquad \text{(eqn 3)}$$

The Thompson and KINSHIP likelihood expressions yield identical values for the same kinship relationships when $\Delta_1 = \ldots = \Delta_6 = 0$, where the conversion between the (R_m, R_p) and (k_0, k_1, k_2) parameters is given by $k_1 = R_m(1 - R_p) + R_p(1 - R_m)$ and $k_2 = R_mR_p$.

Since the two likelihood versions are identical, the use of one or the other is merely a subject of convenience for a given problem. For the regression interpretation of relatedness, the KINSHIP parameterization is mathematically more transparent by mapping (R_m, R_p) to negative value of r via $r = 1/2(R_m + R_p)$, where $-1 \le r \le 1$ if R_m and R_p are varied in the range of [–1, 1]. The upper level of the range was set to '1', which is the highest possible value under any interpretation of relatedness. The lower value of the range was set to '–1' as it is the lowest value possible for r as a correlation coefficient (Wright 1965) and in line with r being interpreted as the regression coefficient between the genotypic values of the two individuals concerned (Hamilton 1972; Pamilo 1990). The maternal (paternal being treated identically) contribution to the KINSHIP likelihood, $L(a, c \mid 0 \le R_m \le 1) = p_a(R_m\delta_{a,c} + (1 - R_m)p_c)$, yields the probability of observing allele a in Y being higher (for $R_m > 0$) than dictated by Hardy–Weinberg equilibrium (allele frequency p_a). However, the likelihood equally well describes the case when the probability of observing a in Y is lower than p_a for negative R_m obtaining exactly the same likelihood expression, $L(a, c \mid R_{\min} \le R_m < 0) = p_a(R_m\delta_{a,c} + (1 - R_m)p_c)$, where the likelihood remains to be non-negative with $R_{\min} = -p_a/(1 - p_a)$ for $a = c$, and never exceeds one with $R_0 = -1$ for $a \ne c$. The construction of the diploid likelihood remains the same as per equations 2 and 3.

The proposed extension of the KINSHIP likelihoods provides analytic continuation of the likelihood expression into the region corresponding to negative r. Figure 1 shows some typical examples of the likelihood surfaces plotted over the full range of (R_m, R_p) values. The analytic continuation is biologically justified under the regression interpretation (Rousset 2002) and is a standard mathematical technique (see any textbook on theory of functions of a complex variable). The (R_m, R_p) coefficients were then found (within 0.001 accuracy) by searching for the global maximum of the likelihood surface using a hill-climbing algorithm. We used the standard conjugate gradient method of Fletcher & Reeves (1964), which was modified to handle the cases when a maximum was located on the border, e.g. Fig. 1(C). The original M estimator of Milligan (2003) was reproduced by limiting both R_m and R_p to the range of [0, 1]. In summary, the new ML estimator is based on the non-negativity of the likelihoods, while the M (Thompson 1975; Milligan 2003) estimator is based on the non-negativity of r.

Simulations

The uniform Dirichlet (Milligan 2003; Wang 2002) ($p_{ml} = C_l\varepsilon_{ml}$) and triangular ($p_{ml} = C_lm$) allele frequency distributions are considered, where: l denotes locus index $1 \le l \le L$; L is the total number of loci; m denotes allele index $1 \le m \le k$; k is the total number of alleles per locus; ε_{ml} is a uniformly distributed random variable $0 \le \varepsilon_{ml} < 1$; C_l is a normalization

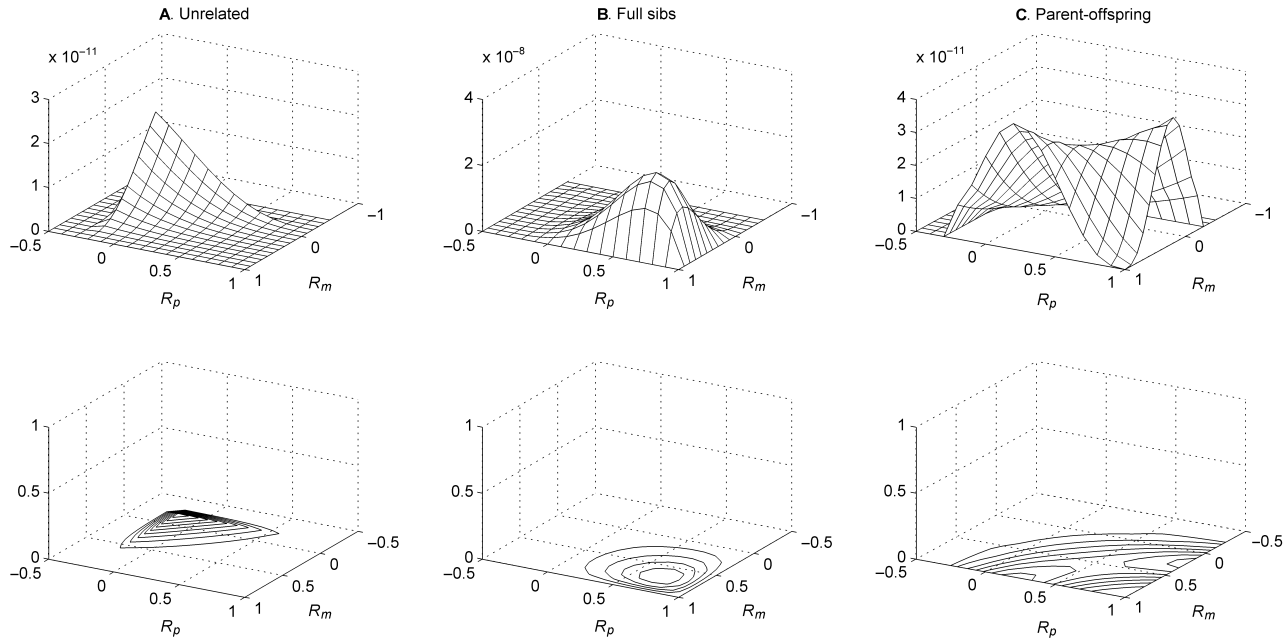

Fig. 1 Typical likelihood surfaces (first row) and contours (second row) for unrelated (A), full-sibling (B) and parent–offspring (C) dyads. Five loci were used, each having five alleles with their population frequencies in a triangular distribution.

coefficient determined by $\sum_{m=1}^{k} p_{ml}=1$. When the Dirichlet distribution was chosen, a freshly generated set of frequencies was used for each data set and pair of individuals. Note that, statistically, the triangular distribution is just an averaged Dirichlet distribution, and hence, mean values obtained with triangular and Dirichlet distributions are very similar while sampling variances are greater with the Dirichlet distribution. The commonly used uniform distribution was not considered since the Dirichlet distribution was more relevant in practice by being the distribution for populations at equilibrium under the joint effects of drift and mutation or migration (Wright 1951). The following four degrees of pairwise kinship were considered: full-siblings (FS, $r = 0.5$), half-siblings (HS, $r = 0.25$), nonrelated (NR, $r = 0$) and parent–offspring (PO, $r = 0.5$). The corresponding genotype dyads were generated as per the simulation procedure described in Goodnight & Queller (1999).

Results

Known allele frequencies

Performance of relatedness estimators was first considered for a biologically artificial case when population allele frequencies were assumed to be known exactly, i.e. the same frequencies were used to generate a sample and to estimate *r* from it. The multilocus tests of Milligan (2003) with five alleles at each locus were performed and results were displayed in Fig. 2. The results of Milligan (2003) from his Figures 3–5 were closely reproduced for the M estimator in our Fig. 2. Note that Milligan (2003) presented his results for the coefficient of coancestry, $\theta = r/2$. To allow for a consistent comparison of the QG, LR and M estimators, Fig. 2 displays the results for the non-negative (NN) versions of the QG and LR estimators (denoted NN-QG and NN-LR), i.e. all negative *r* estimates have been replaced by zeros (see Table 1). Figure 2 confirms the conclusion of Milligan (2003) in that the M estimator delivers the lowest overall estimation error, as measured by the root mean square error (RMSE), for the parent–offspring dyads ($r = 0.5$), where RMSE of an estimator *r* was calculated via $\mathrm{RMSE}(r) = \sqrt{\mathrm{var}(r) + (\mathrm{bias}(r))^2}$. In the

Table 1 Estimators used

Abbreviation	Full name/reference
LR	r_{LR} (Lynch & Ritland 1999)
NN-LR	Non-negative LR defined by $r_{NN\text{-}LR} = \max(0, r_{LR})$, i.e. $r_{NN\text{-}LR} = 0$ if $r_{LR} < 0$, and $r_{NN\text{-}LR} = r_{LR}$ if $r_{LR} \geq 0$
M	The maximum-likelihood estimator of Thompson (1975) and Milligan (2003) based on the non-negativity of *r* estimates
ML	The new maximum-likelihood estimator based on the non-negativity of likelihoods (this study)
QG	r_{QG} from equation 1 (Queller & Goodnight 1989)
NN-QG	Non-negative QG defined by $r_{NN\text{-}QG} = \max(0, r_{QG})$
W	The version of r_W (Wang 2002) bias-corrected for sample size

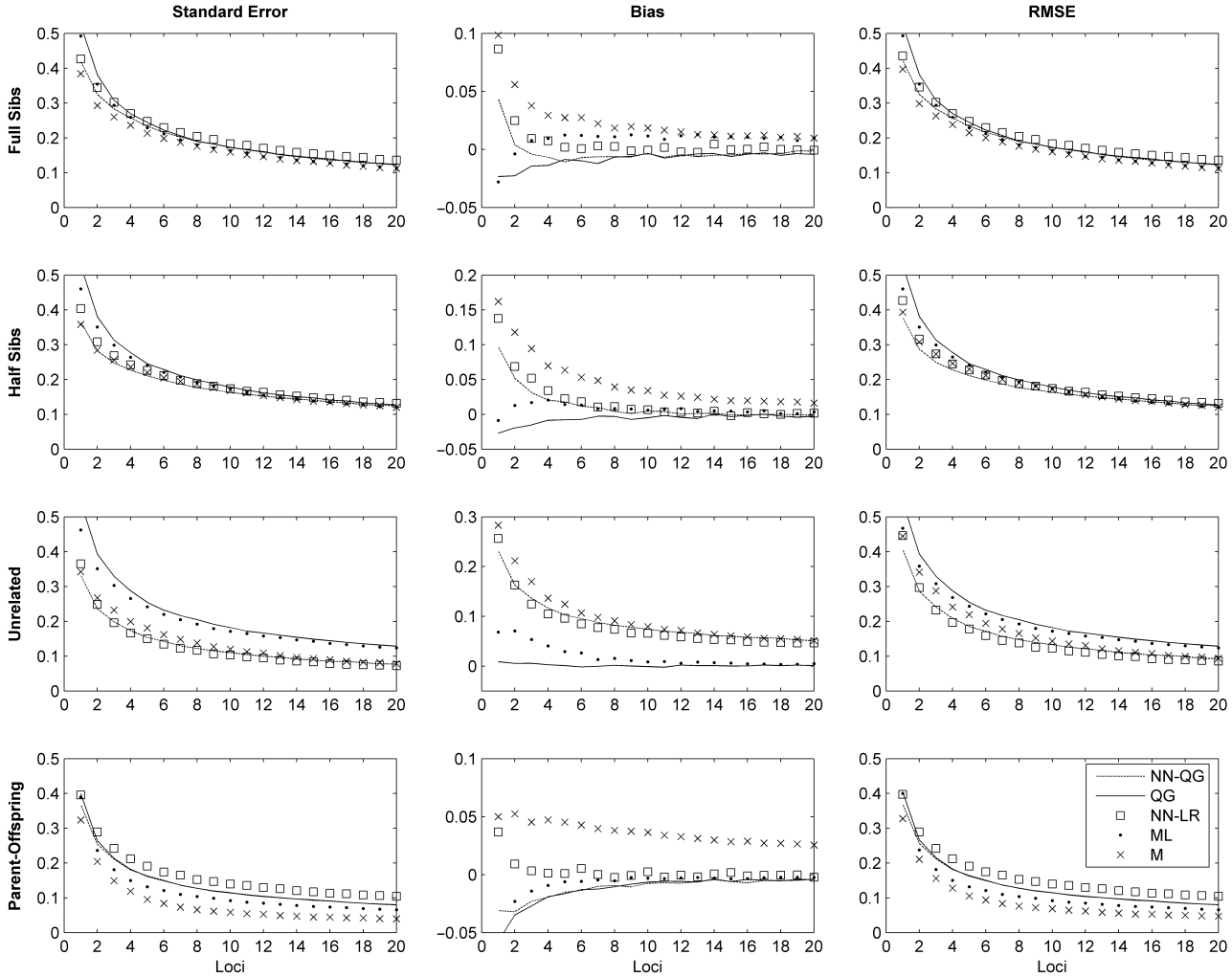

Fig. 2 Comparison of standard error, bias and RMSE among estimators. The curves were based on samples of 10 000 dyads at each point, where dyads were related as shown in the row labels. Each sample dyad was genotyped at a varied number of loci (*x*-axis) with five alleles at each locus. Allele frequencies distributions were drawn from a uniform Dirichlet distribution.

case of the full-sibling dyads ($r = 0.5$), the M estimator's advantage in RMSE is hardly noticeable. In the case of dyads with lower r ($r = 0$ for unrelated and $r = 0.25$ for half-sibling), all three non-negative estimators (NN-QG, NN-LR and M) are virtually identical. Note that the W estimator was not considered for these biologically unrealistic testing conditions. One of the main advantages of the W estimator is being bias- corrected for a limited sample size while the estimates presented in Fig. 2 were calculated with the allele frequencies which were not estimated from the samples.

Figure 2 also displays both the QG and NN-QG estimates to show that the reduction in the standard error of the NN-QG estimates is merely a consequence of making the QG estimator non-negative. The effect could be seen most clearly for the unrelated dyads, where the reduction in the standard error from the QG to NN-QG estimates is almost identical to the reduction from the ML and M estimates. Hence, the lower standard error of the non-negative M estimates for unrelated dyads, when compared to the QG (and LR) estimates, is not a feature of the M estimator. The standard error could be equally reduced by making the QG and LR estimators non-negative. In fact, the standard errors of the NN-QG and NN-LR estimates for the unrelated dyads are even slightly lower than the standard error of the M estimates.

Figure 2 confirms that the new ML estimator becomes less biased significantly faster than the original M estimator as the number of unlinked loci increases.

Estimated allele frequencies

In practice, allele frequencies and relatedness must often be estimated at the same time from a limited population

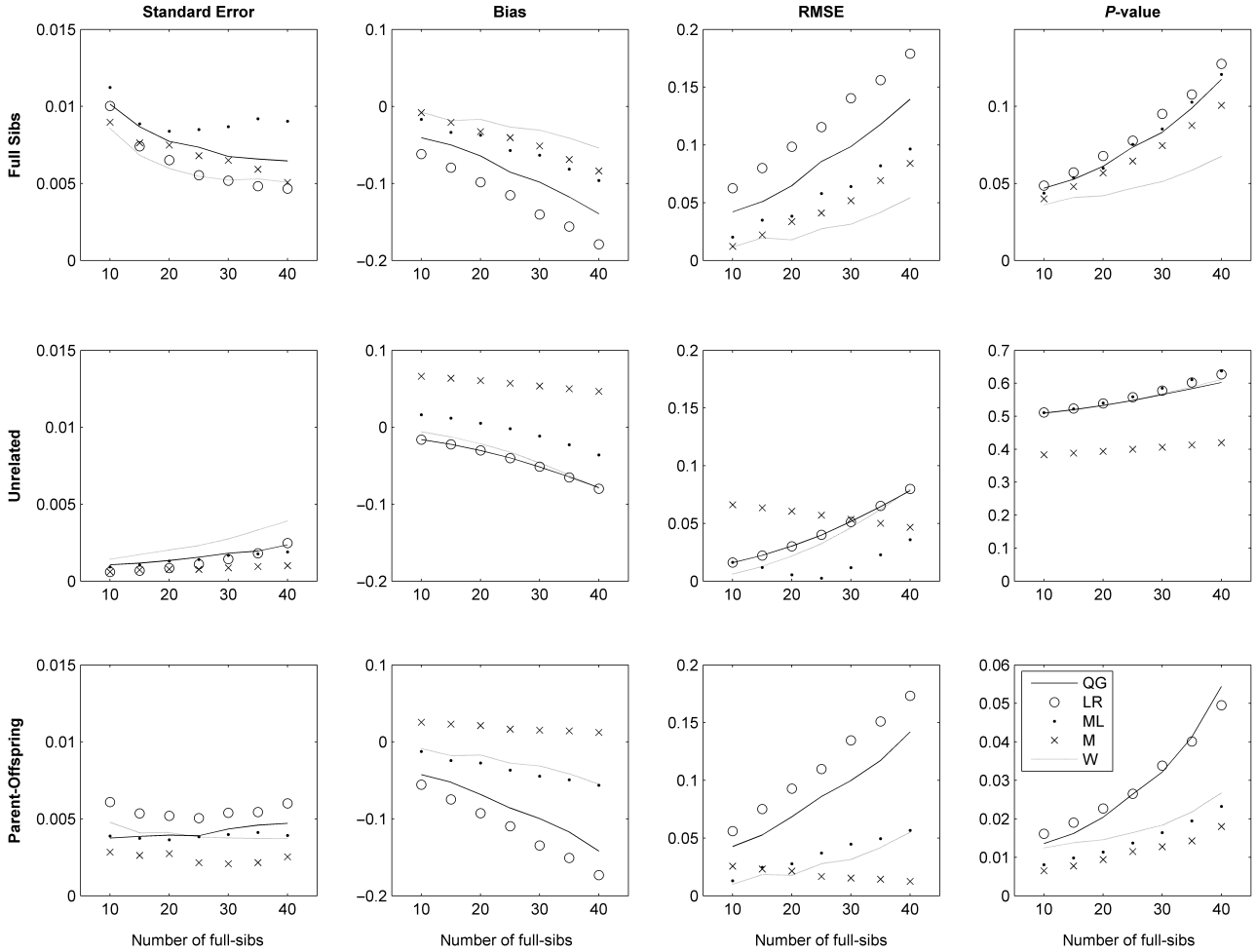

Fig. 3 Standard error, bias, RMSE and *P* values as a function of the number of full-siblings in a sample of 100 individuals. Five loci were used, each having 10 alleles generated according a triangular allele frequency distribution.

sample. Such simultaneous estimation of frequencies and relatedness is not possible with the considered estimators (Wang 2002) because the estimators require the frequencies as an input. Following Wang (2002), Fig. 3 presents results where the frequency estimation errors were introduced by a limited sample size and the presence of a cluster of relatives (two parents together with their full-sibling offspring), where the number of alleles per locus was increased from five (as in Fig. 2) to a more practical (for microsatellites) number of 10 alleles.

Relatedness estimates (obtained with either of the implemented estimators) could be tested for their significance by generating a null distribution (Belkhir *et al*. 2002) and reporting the corresponding *P* values. The resampling procedure of Guo & Thompson (1992) is used to generate null-hypothesis dyads by randomly selecting alleles without replacement from the pool of available sample alleles. It was found that by setting the number of null-hypothesis dyads to 17 000 (as for heterozygosity in Guo & Thompson 1992), the *P* values could be obtained within 0.01 accuracy.

Each simulation trial in Fig. 3 reproduces one biologically feasible experiment where: (i) a population sample is generated based on *given* population allele frequencies which could be predefined (e.g. triangular as per Wang (2002)) or taken from actual biological data; (ii) *sample* allele frequencies are then inferred from the sample; (iii) the sample allele frequencies are used as an input for the relatedness estimators; (iv) standard error, mean, RMSE and *P* values are calculated for each of the pairwise kin relationships present in the sample, e.g. a sample of $n = 100$ individuals contains $n(n-1)/2 = 4950$ unique dyads. Finally, the considered statistics (standard error, bias, RMSE and *P* values) are averaged over a large number of simulation trials. The curves in Fig. 3 were based on 1000 trials (400 trials for the M and ML estimates) at each point.

RMSE captures the overall estimation error (Milligan 2003) safeguarding against optimizing an estimator for variance at the expense of the bias and vice-versa. However, in our case the RMSE values are in qualitative agreement (Fig. 3) with the bias values. All four considered

statistics obtained with the QG, LR and W estimators are very similar for unrelated dyads. Although for the full-sibling (FS) and parent–offspring (PO) dyads the LR estimator has the largest bias and RMSE compared to QG, the corresponding *P* values are very similar, indicating that the inferring powers of the estimators are very similar under the considered conditions. Figure 3 also shows that: (i) the W and ML estimators are more accurate (smaller RMSE) than the QG and LR estimators for all three considered kin relationships; (ii) W outperforms ML for the FS and PO dyads (smaller RMSE), while ML outperforms W for the unrelated dyads; (iii) ML exhibits better inferring power (lower *P* values) than W for the PO dyads. The W estimates are calculated as per Wang (2002) with formulas bias-corrected for a sample size (hence, W does not appear in Fig. 2). The M estimates were also calculated (Fig. 3) for a discussion below.

Discussion

Pairwise relatedness estimates using molecular codominant markers (DNA microsatellite loci in particular) have given us an important toolbox to study patterns of genealogy and kinship in natural populations, where prior pedigree information is typically lacking. The first widely accepted relatedness estimator to appear was developed by Queller & Goodnight (1989) and was implemented in Goodnight's software RELATEDNESS and later in KINSHIP (Goodnight & Queller 1999). The reliability of the QG estimator has been questioned in several theoretical studies, and alternatives have been proposed (Lynch & Ritland 1999; Van de Casteele *et al*. 2001; Wang 2002; Milligan 2003). When the sample allele frequencies were used, our RMSE and *P*-value results (Fig. 3) confirmed the conclusions of Wang (2002) in that the W estimator was a superior estimator compared to QG and LR, although the QG2 estimates were used in Wang (2002) in the comparison.

The recommendation of Lynch & Ritland (1999) to prefer the LR estimator over the QG2 estimator should be clarified. The recommendation may still be valid but it concerns the QG2 estimator which is not used in practice as often as the KINSHIP's version, i.e. the QG estimator. As shown in Fig. 3, the QG (not QG2) estimator is very similar to the LR estimator (by the *P* values) and even performs slightly better (by RMSE) than the LR estimator (at least at the considered circumstances).

Figure 3 shows that bias (and RMSE in this case) of the M estimates may be remarkably small, e.g. for the PO estimates. This is due to the positive *estimation* bias being cancelled out by the negative *sample* (due to the sample allele frequencies) bias. Unfortunately, this interplay of biases is unpredictable and should not be relied upon as a justification of the M estimator or, for that matter, any other non-negative estimators. Figure 2 demonstrates that the positive bias could be artificially introduced to most other estimators by making them non-negative. Then, the sample negative bias (due to overestimated-from-sample allele frequencies) could reduce the overall bias to an arbitrarily low value. In the case of Fig. 3, the effect could not be misjudged as the negative bias is introduced gradually by increasing the family size. In practice, however, a sample of an arbitrary (and unknown) kin structure and/or limited size could not guarantee such cancellation of biases and must be analysed by an unbiased estimator.

The question of which estimator should be preferred at what circumstances is currently a highly debated topic (e.g. Oliehoek *et al*. 2006). Although Fig. 3 showed that the ML and W estimators performed better than the QG and LR estimators, Van de Casteele *et al*. (2001) demonstrated that actual performance of estimators may vary quite considerably and depends on sample and population composition. The main result of this study is in showing that the new ML estimator can now be considered on equal terms with other currently popular unbiased estimators (e.g. QG, QG2, LR and W). On the other hand as shown by Fig. 2, the M estimator could only be consistently compared against the estimators that are non-negative by definition or made non-negative (e.g. NN-QG and NN-LR).

In practice, a given sample may have an unknown composition, and therefore, the simulation assessment of estimators via the RMSE and bias calculations is strictly speaking impossible. Figure 3 illustrates a possible solution where the pairwise relatedness could be first detected with desired confidence (via *P* values) since the calculation of the *P* values does not require the knowledge of the actual kinship. Figure 3 shows that a more accurate estimator (as measured by RMSE) would normally exhibit lower *P* values for a related dyad, e.g. the W and ML estimates for the PO dyads and the W estimates for the FS dyads. Figure 3 also shows that the detection of the full-sibling dyads is a statistically harder task than the detection of the parent–offspring dyads, although both have the same $r = 0.5$, e.g. with better than 95% confidence ($P < 0.05$), only the W estimator detected 20 full-siblings while all considered estimators could detect the parent–offspring dyads in the presence of up to 35 full-siblings. The new ML estimator exhibits lower *P* values than the W estimator for the PO dyads over all considered family sizes, indicating the parent–offspring detection as a possible area of strength of the new ML estimator. From the conservation point of view (Oliehoek *et al*. 2006), an estimator that detects the highest number of related dyads in a sample (e.g. with $P < 0.05$) should be preferred and then used to identify the dyads with the highest *P* values as being the most unrelated.

In summary, it was demonstrated that under the considered biologically realistic conditions (Fig. 3), the QG estimator was comparable to the LR estimator while the W and ML estimators outperformed both QG and LR. A

formulation of a new ML relatedness estimator is possible under the regression interpretation of relatedness, where the previously reported (Milligan 2003) bias is significantly reduced. Although our results demonstrate that ML and W are comparable, ML (being a maximum-likelihood estimator) has a distinct advantage in that it can be further improved to allow for additional field or pedigree information (Wang 2004). For example, users of the KINSHIP (Goodnight & Queller 1999) and KINGROUP (Konovalov *et al*. 2004) programs could specify maternal or paternal relationships between sample genotypes in order to obtain more accurate likelihoods, i.e. the summation in equation 3 is not used or only partially used. Further research could be done to ascertain the extent of such an improvement. Computation of the ML estimates is feasible on a consumer-grade personal computer, e.g. all 4950 pairwise ML estimates (for a sample containing 100 individuals) could be calculated in less than a minute via the KINGROUP version 2 program on a 2GHz Pentium 4 PC, while the corresponding *P* values could be calculated within a few minutes.

The QG, LR, W and ML relatedness estimates (together with the corresponding *P* values) could now be calculated via the KINGROUP version 2 program (Konovalov *et al*. 2004) which is freely available from www.kingroup.org and could be run on all Java-enabled software environments such as Unix/Linux, Windows and Mac-OSX (Java release 1.5 or higher is required). Pairwise and group-wise bias correction procedures (Queller & Goodnight 1989; Goodnight & Queller 1999) were also implemented with additional option for the frequencies of rare alleles, e.g. allowing them to be set to zero (as per KINSHIP) or not if the rare alleles are only present in the focal dyad (or group).

Acknowledgements

This paper was partly written when D.K. was on sabbatical leave at the University of Bern. We would like to thank University of Bern and James Cook University and, in particular, Michael Taborsky and Bruce Litow for supporting this collaborative project, and Peter Stettler for his hospitality. We also would like to thank Laurent Excoffier, Gerald Heckel, Nigel Bajema, Tony Brown and anonymous reviewers for helpful comments on earlier versions of this manuscript. D.H. is supported by SNF Grant 3100A0-108473.

References

Belkhir K, Castric V, Bonhomme F (2002) IDENTIX, a software to test for relatedness in a population using permutation methods. *Molecular Ecology Notes*, **2**, 611–614.

Beyer J, May B (2003) A graph-theoretic approach to the partition of individuals into full-sib families. *Molecular Ecology*, **12**, 2243–2250.

Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution*, **18**, 503–511.

Csillery K, Johnson T, Beraldi D *et al*. (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, **173**, 2091–2101.

Fletcher R, Reeves CM (1964) Function minimization by conjugate gradients. *Computer Journal*, **7**, 149–154.

Goodnight KF, Queller DC (1999) Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology*, **8**, 1231–1234.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Hamilton WD (1972) Altruism and related phenomena, mainly in social insects. *Annual Review of Ecology and Systematics*, **3**, 193–232.

Heg D, van Treuren R (1998) Female-female cooperation in polygynous oystercatchers. *Nature*, **391**, 687–691.

Jacquard A (1974) *The Genetic Structure of Populations*. Springer-Verlag, New York.

Kalinowski ST, Wagner AP, Taper ML (2006) ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*, **6**, 576–579.

Konovalov DA, Manning C, Henshaw MT (2004) KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Molecular Ecology Notes*, **4**, 779–782.

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics*, **152**, 1753–1766.

Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, Massachusetts.

Milligan BG (2003) Maximum-likelihood estimation of relatedness. *Genetics*, **163**, 1153–1167.

Oliehoek PA, Windig JJ, van Arendonk JAM, Bijma P (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. *Genetics*, **173**, 483–496.

Pamilo P (1990) Comparison of relatedness estimators. *Evolution*, **44**, 1378–1382.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258–275.

Rousset F (2002) Inbreeding and relatedness coefficients: what do they measure? *Heredity*, **88**, 371–380.

Thomas SC (2005) The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1457–1467.

Thompson EA (1975) Estimation of pairwise relationships. *Annals of Human Genetics*, **39**, 173–188.

Van de Casteele T, Galbusera P, Matthysen E (2001) A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology*, **10**, 1539–1549.

Wang JL (2002) An estimator for pairwise relatedness using molecular markers. *Genetics*, **160**, 1203–1215.

Wang JL (2004) Sibship reconstruction from genetic data with typing errors. *Genetics*, **166**, 1963–1979.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Wright S (1965) The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.