

# The accuracy of a heritability estimator using molecular information

Silvia Teresa Rodríguez-Ramilo · Miguel Ángel Toro ·  
Armando Caballero · Jesús Fernández

Received: 12 July 2006 / Accepted: 7 December 2006 / Published online: 5 January 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** The heritability of a quantitative trait is a key parameter to quantify the genetic variation present in a population. Although estimates of heritability require accurate information on the genetic relationship among individuals, pedigree data is generally lacking in natural populations. Nowadays, the increasing availability of DNA markers is making possible the estimation of coancestries from neutral molecular information. In 1996, K. Ritland developed an approach to estimate heritability from the regression of the phenotypic similarity on the marker-based coancestry. We carried out simulations to analyze the accuracy of the estimates of heritability obtained by this method using information from a variable number of neutral codominant markers. Because the main application of the estimator is on populations with no family structure, such as natural populations, its accuracy was tested under this scenario. However, the method was also investigated under other scenarios, in order to test the influence of different factors (family structure, assortative mating and phenotypic selection) on the precision. Our results suggest that the main factor causing a directional bias in the estimated

heritability is the presence of phenotypic selection, and that very noisy estimates are obtained in the absence of a familiar structure and for small population sizes. The estimated heritabilities from marker-based coancestries showed lower accuracy than the estimated heritabilities from genealogical coancestries. However, a large amount of bias occurred even in the most favourable situation where genealogical coancestries are known. The results also indicate that the molecular markers are more suitable to infer coancestry than inbreeding.

**Keywords** Molecular markers · Kinship · Coancestry · Quantitative trait · Inbreeding

## Introduction

The knowledge of the genetic architecture of phenotypic traits provides information on the ability of a population to evolve under natural and artificial selection. The amount of genetic variability for quantitative traits, usually summarised by the heritability (the ratio of the additive genetic variance to the phenotypic variance), is a central quantitative genetic parameter of importance in evolutionary, conservation and animal and plant breeding contexts (Falconer and Mackay 1996; Frankham et al. 2002). However, obtaining estimates of heritability from natural or unmanaged populations is not an easy task. One of the main problems is that the conventional techniques available to estimate variance and covariance components require accurate information on the genetic relationship among individuals. This is accomplished when pedigree information of the population under

---

S. T. Rodríguez-Ramilo (✉) · M. Ángel Toro ·  
J. Fernández

Departamento de Mejora Genética Animal, Instituto  
Nacional de Investigación y Tecnología Agraria y  
Alimentaria (INIA), Crta. A Coruña Km. 7.5,  
28040 Madrid, Spain  
e-mail: ramilo@inia.es

A. Caballero  
Departamento de Bioquímica, Genética e Inmunología,  
Facultad de Biología, Universidad de Vigo, Campus  
Universitario, 36310 Vigo, Spain

study is available, but for most natural and captive populations, genealogical information is absent (Lynch and Walsh 1998). Only in special cases, pedigrees can be determined from data of mating activities, but this usually requires long-term intensive observation (see Merilä et al. 2001; Garant et al. 2004) and still may not be entirely reliable.

Throughout the most recent years, the extensive development and application of highly polymorphic molecular markers (especially microsatellites) has proven highly valuable, particularly in the population and conservation genetics fields (Awise 1994; Frankham et al. 2002). Such markers have been used in parentage analyses, providing a high power to infer paternities (Awise et al. 2002; Blouin 2003). Accordingly, several coancestry estimators have been developed to measure pairwise coancestries using the molecular information provided by different types of markers, particularly the codominant microsatellites. Most of them were described and compared by Lynch and Ritland (1999), Wang (2002), Toro et al. (2002) and Oliehoek et al. (2006). Other type of relatedness estimators involving an explicit pedigree reconstruction have also been developed (Butler et al. 2004; Wang 2004; Fernández and Toro 2006).

Once techniques to estimate coancestry from molecular information are being routinely applied, the next movement is to obtain estimates of evolutionary related parameters, including quantitative genetic characteristics as the heritability. Ritland (1996a; 2000) suggested an approach to quantify additive genetic variance from the covariance between pairwise relatedness estimated from molecular markers and phenotypic similarity. The estimator is justified because in a simple setting of independent families the pairwise intraclass correlation, defined as the correlation averaged over all possible pairs of observations, is an estimator of the heritability (Kempthorne 1957; Donner and Kovall 1980). This method seems easily applicable and appealing, as it is based on the determination of this simple relationship between phenotype and genotype. The approach, obviously, does not require specification of an explicit pedigree or prior knowledge of population structure (Garant and Kruuk 2005). In the original procedure, coancestry was calculated using a method of moments estimator (Ritland 1996b).

Alternatively, if prior information is available on population structure, likelihood-based procedures may be implemented, in which pairs are placed into a pre-determined population structure according to the probability of observing their genotype and phenotype (Mousseau et al. 1998; Thomas et al. 2000, 2002). A third approach involves an explicit reconstruction of

groups of a certain coancestry, which can then be used as pedigree information in a standard quantitative genetic analysis. This method is usually performed using Markov chain Monte Carlo (MCMC) procedures to reconstruct sibships within a single generation (Thomas and Hill 2000; Thomas et al. 2002; Blouin 2003).

Ritland's method (1996a, 2000) was well suited for studies of plants, given that they are sedentary and show passive dispersal. It was first applied on a wild population of yellow monkeyflowers (*Mimulus guttatus*) by Ritland and Ritland (1996). In this study, there was available data on 10 quantitative characters and 10 polymorphic allozymic loci showing that, for many characters, the marker-based estimates of heritability were higher than the non-marker estimates, and sometimes outside the parameter space ( $> 1$ ). In contrast, a recent study on the same species failed to find significant heritabilities for similar characters (Van Kleunen and Ritland 2004). Despite its great intuitive appeal, an unexpectedly reduced number of cases applying this method have been published either on plant or animal populations. Most representative examples are in *Oncorhynchus tshawytscha* (Mousseau et al. 1998), *Ovis aries* (Thomas et al. 2002), and *Oncorhynchus mykiss* (Wilson et al. 2003). Klaper et al. (2001), estimating the heritability of phenolic compounds in a population of *Quercus laevis*, showed that the method failed to estimate heritability because of the low and rather uniform coancestry among individuals. A substantial actual variance in coancestry seems to be required for the method to be successful, implying a broad range of relation degrees. This is something problematic in studies of natural populations, suggesting that the limited number of experimental applications of this method could be probably attributed to the low actual variance of coancestry found in most sampled populations (Garant and Kruuk 2005). This could also contribute to the reduced accuracy and unreliable parameter estimation reported in the published empirical studies (Mousseau et al. 1998; Klaper et al. 2001; Thomas et al. 2002; Wilson et al. 2003).

In this paper, we present a simulation study to evaluate the accuracy and behaviour of the narrow sense heritability estimated by the Ritland's regression method considering different population scenarios. These include different numbers of neutral codominant molecular markers available, different population structures and management population regimes (presence or absence of selection). As a collateral result, the accuracy of the estimates of coancestry based on molecular markers is also tested in order to infer

whether it has implications on the performance of the heritability estimator in some scenarios.

**Method and simulations**

**Genealogical coancestry**

The coancestry coefficient between individuals  $i$  and  $j$  ( $f_{g,ij}$ ) is defined as the probability that an allele randomly drawn from individual  $i$  is identical by descent (IBD) to another allele randomly drawn from individual  $j$  at the same autosomal locus (Malécot 1948), and it is calculated from pedigrees.

**Marker-based coancestry**

Accordingly, it is also possible to define the molecular coancestry applying the Malécot’s (1948) definition to the marker loci. Thus, the molecular coancestry between individuals  $i$  and  $j$  ( $f_{M,ij}$ ) is the probability that two alleles taken at random, one from each individual, are equal (identical by state, IBS). This probability can be calculated separately for each allele  $l$  at each locus  $k$  ( $f_{M,ijkl}$ ), comparing each allele of individual  $i$  with those of individual  $j$ , and adding up to obtain  $f_{M,ijk}$ . Throughout  $k$  markers, the molecular coancestry can be obtained as the arithmetic mean over marker loci.

From the above molecular coancestry values, different estimators have been proposed to infer the genealogical coancestries. In the present study, we will use the estimator originally proposed by Ritland (1996b) because, to our knowledge, it is the only one for which a correction for its actual variance is described. This estimator, that will be called marker-based coancestry henceforth, is obtained from

$$\hat{f}_{R,ij} = \frac{\sum_k^L \sum_l^{n_k} \frac{f_{M,ijkl} - p_{kl}^2}{p_{kl}}}{\sum_k^L (n_k - 1)},$$

where  $L$  is the number of marker loci,  $n_k$  is the number of alleles at the locus  $k$ ,  $p_{kl}$  is the frequency of allele  $l$  at locus  $k$  in the base population, and  $(n_k - 1)$  is the weighting factor for each locus. Usually, the allele frequencies of the base population are not available, and then the current population allele frequencies are to be used. This estimator assumes that the correlation among the estimates of different loci is zero.

**Heritability estimation procedure**

The basis of the estimator proposed by Ritland (1996a; 2000) is as follows. The phenotypic similarity for a

given quantitative trait  $X$  between individuals  $i$  and  $j$  ( $i \neq j$ ) is calculated as

$$Z_{ij} = \frac{(X_i - \bar{X})(X_j - \bar{X})}{\sigma_X^2},$$

where  $X_i$  is the phenotypic value for individual  $i$ , and  $\bar{X}$  and  $\sigma_X^2$  are the mean and variance of the phenotypic trait  $X$  in the population. If shared phenotypes are determined by shared genes and environments,  $Z_{ij}$  may also be expressed as

$$Z_{ij} = 2f_{ij}h^2 + r_e + e_{ij},$$

where  $f_{ij}$  is the genealogical coancestry,  $r_e$  is a correlation due to shared environments, and  $e_{ij}$  is the residual error. As the above equation shows a linear relation, estimates of heritability can be obtained from a regression approach using the coancestries (genealogical coancestries or marker-based coancestries) and the phenotypic similarities of pairs of individuals. When relationships are estimated from markers ( $\hat{f}_R$ ) the estimates of heritability are obtained from

$$h_{Mol}^2 = \frac{\text{cov}(Z, \hat{f}_R)}{2 \times \text{var}(\hat{f}_R)},$$

where  $\text{cov}(Z, \hat{f}_R)$  is the covariance between the phenotypic similarity and the estimated marker-based coancestry, and  $\text{var}(\hat{f}_R)$  is the actual variance of the estimated coancestries (described in detail by Ritland 1996a, 2000). The measure of this variance is critical, because usually there is a mixture of different relatives, such as full sibs, half sibs, first cousins, etc., along with unrelated individuals. The marker-based coancestry estimators usually exhibit a high sampling variance (Van de Castele et al. 2001; Rodríguez-Ramilo et al. 2006), thus implying a downwardly biased estimated heritability by the Ritland’s method. Therefore, it is important to use the actual variance to reduce this bias.

**Simulations**

*Trait under study*

A quantitative trait was simulated to be controlled by 120 biallelic loci. The initial allele frequencies were  $p = q = 0.5$ , and the genotypes were generated in Hardy–Weinberg and linkage equilibrium. For each locus, the corresponding genotypic values of the genotypes  $AA$ ,  $Aa$  and  $aa$  were 5.0, 3.5 and 2.0, respectively, and individual genotypic values were

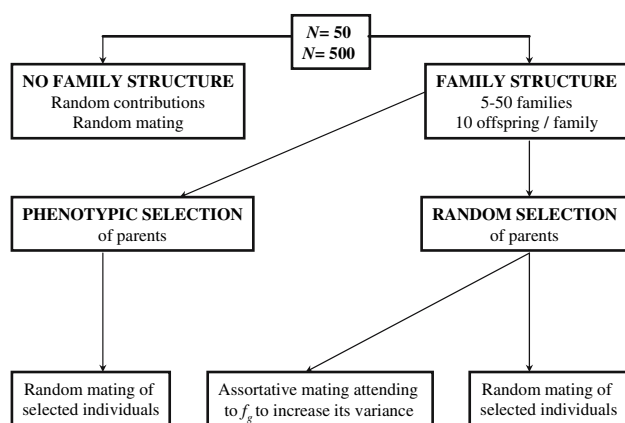
obtained from the sum over loci. The individual phenotypic value was obtained adding, to the genotypic value, an environmental deviation sampled from a normal distribution with mean zero and a variance such that the initial trait heritability was  $h^2 = 0.1$  or  $0.5$ .

### Markers

Ten, 20 or 100 neutral codominant marker loci were simulated. At each marker locus there were initially six alleles at equal frequencies, with genotypes generated in Hardy–Weinberg and linkage equilibrium. Free recombination was assumed along the whole genome.

### Population structure

Two population sizes ( $N = 50$  or  $500$ ) with equal numbers of males and females were considered. Population size was constant along 10 non-overlapping generations in two possible scenarios (see Fig. 1). In the first scenario, the population was not structured in families, contributions of parents to progeny were free, and mating was at random. In the second scenario, the population included a family structure, where 5 or 50 full-sib families (for  $N = 50$  or  $500$ , respectively) were obtained through equal contributions of parents to the next generation (10 offspring per family). The parents of each family were chosen at random or selected according to their phenotypic value (the male and female per family with the largest phenotypic value). The mating procedure in the second scenario, where the population had a family structure, was random or assortative. In this latter case, random selected males and females were sorted according to their genealogical coancestry and mated in sequence, with the



**Fig. 1** Procedure followed to design the different population scenarios where the heritability was evaluated.  $f_g$ : genealogical coancestry

objective of increasing the variance of genealogical coancestries between individuals.

### Measured parameters

In order to assess the accuracy and bias of the heritability estimator, several parameters were evaluated in each analyzed population: (1) the true heritability ( $h_{\text{True}}^2$ ), obtained from the simulations as the ratio of additive genetic variance to the phenotypic variance at the particular generation and scenario considered; (2) the regression heritability estimated from the genealogical coancestries ( $h_{\text{Ped}}^2$ ); and (3) the regression heritability estimated from the marker-based coancestries ( $h_{\text{Mol}}^2$ ).

The indicators of accuracy used were (1) the *bias*, i.e. the difference between the genealogical ( $h_{\text{Ped}}^2$ ) or marker-based heritability ( $h_{\text{Mol}}^2$ ), and the true heritability ( $h_{\text{True}}^2$ ); and (2) the *squared error*, the squared difference between  $h_{\text{Ped}}^2$  or  $h_{\text{Mol}}^2$  and  $h_{\text{True}}^2$ , i.e.  $(h_{\text{Ped}}^2 - h_{\text{True}}^2)^2$  and analogously for  $h_{\text{Mol}}^2$ . Each parameter was evaluated after 1 and 10 generations (short and long-term evaluations, respectively). The number of simulation runs in each situation was always 100.

A full-factorial ANOVA was carried out using SPSS 12.0 to test the significance between factors. This included number of marker loci (10, 20 or 100), initial trait heritability (0.1 or 0.5), type of coancestry (genealogical or marker-based), population size (50 or 500), generation (1 or 10), selection system (random or phenotypic), population structure (structured or not) and mating regime (random or assortative).

The coefficient of determination and the regression coefficient of the genealogical coancestry on the marker-based coancestry were also analysed. These parameters were calculated for the coancestries between pairs of individuals and for the self-coancestries (coancestry of one individual with itself) separately, since the latter is a measure of the inbreeding coefficient ( $F$ ) of the individual ( $F_i = 2f_{ii} - 1$ ).

### Results

The results of the full factorial ANOVA for the bias and the squared error are given in Table 1. The interaction factors between the sources of variation are not shown in the table for clarity. The number of markers, the type of coancestry and the mating regime factors were non-significant for the bias, whereas the generation, and the selection and mating regime factors were non-significant for the squared error. All other factors showed significant effects.

**Table 1** Main factors of the full factorial ANOVA carried out for the bias and the squared error of heritability estimates

Factors	df	SS	MS	F	p
<i>Bias</i>					
Markers	2	0.326	0.163	1.347	0.260
Heritability	1	2.457	2.457	20.283	0.000
Coancestry	1	0.005	0.005	0.038	0.846
Population size	1	3.409	3.409	28.145	0.000
Generation	1	3.119	3.119	25.752	0.000
Selection	1	7.785	7.785	64.272	0.000
Structure	1	0.587	0.587	4.846	0.028
Mating	1	0.023	0.023	0.188	0.665
Error	19014	2303.158	0.121		
Total	19200	2392.069			
<i>Squared error</i>					
Markers	2	25.235	12.618	89.053	0.000
Heritability	1	11.224	11.224	79.218	0.000
Coancestry	1	53.280	53.280	376.038	0.000
Population size	1	64.063	64.063	452.141	0.000
Generation	1	0.219	0.219	1.548	0.213
Selection	1	0.047	0.047	0.335	0.563
Structure	1	142.090	142.090	1002.841	0.000
Mating	1	0.024	0.024	0.171	0.680
Error	19014	2694.045	0.142		
Total	19200	3697.977			

The average bias and the average squared error of the estimated heritabilities for the diverse population scenarios considered are shown in Tables 2 and 3, respectively, and the distributions of the biases for the different factors are illustrated in Fig. 2. A clear consistent downward bias in the estimates of heritability is

observed under phenotypic selection and high initial trait heritability, irrespective of the heritability estimator, the population size or the generation number (Table 2). This bias is also apparent in the box plot corresponding to the selection factor in Fig. 2. Other significant average upward biases refer to the case for  $h^2_{Ped}$ ,  $N = 50$ , generation 10 and random selection, particularly for high initial trait heritability, as well as for  $h^2_{Mol}$  with 20 markers and generation 10 in unstructured populations (Table 2). These average biases are responsible for the significance of the bias for different factors in the ANOVA of Table 1. Nevertheless, there is not a clear overall tendency for a positive or negative sign of the biases, except for the aforementioned phenotypic selection factor (see Fig. 2).

The largest noises in the estimates were detected for unstructured populations and reduced populations sizes (Table 3 and Fig. 2). In addition, a high initial trait heritability, a reduced number of molecular markers, and marker-based estimated heritabilities implied more variable estimates than a low initial heritability, a larger number of markers and genealogical estimated heritabilities, respectively. The generation number, and the selection and mating regimes did not affect the variance of the bias, as evidenced by the lack of significance in the ANOVA results of Table 1. Note that the variance of the biases obtained from the marker-based estimated heritabilities was larger than those from pedigree-based heritabilities, as expected, but

**Table 2** Average bias of the estimated heritabilities ( $\times 10^2$ )

Scenario	N	50								500													
		Structure				Selection				Mating				Structure				Selection				Mating	
Generation	Markers	$h^2$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$			
1	10	0.1	2	-7	0	0	0	-1	0	1	1	11	0	2	0	2	-1	-2					
		0.5	-1	-16	-2	-3	-1	-4	-10*	-10*	0	4	2	4	0	4	-11*	-12*					
	20	0.1	3	14	-2	-2	1	1	-1	-1	-1	-9	1	0	1	2	0	0					
		0.5	-3	11	-2	-3	0	1	-14*	-13*	-1	-5	1	1	0	2	-11*	-12*					
	100	0.1	3	5	3	3	1	1	-1	-1	-1	-3	0	1	0	0	0	0					
		0.5	3	9	3	2	4	3	-14*	-13*	-2	-3	1	2	0	1	-10*	-11*					
10	10	0.1	-2	0	8*	3	4	1	9*	6	1	8	1	0	1	1	0	0					
		0.5	1	8	21*	-2	15*	-1	-1	-5	1	12	2	1	1	2	-14*	-14*					
	20	0.1	10	21*	3*	-1	6	2	-2	-3	0	1	1	0	-1	-1	0	0					
		0.5	11	28*	18*	-7	18*	3	3	-10*	2	7	3	1	-5	-3	-12*	-12*					
	100	0.1	2	3	1	-5	2	0	4	0	0	2	1	1	0	0	-1	-1					
		0.5	-1	4	18*	-27	13	-1	11	-3	-1	3	2	1	-1	0	-12*	-13*					

N: population size;  $h^2$ : simulated initial trait heritability;  $h^2_{Ped}$ : estimated heritability from genealogical coancestry;  $h^2_{Mol}$ : estimated heritability from marker-based coancestry; \* statistically significant difference between the estimated heritability and the true one ( $h^2_{True}$ ) ( $p < 0.05$ )

**Table 3** Average squared error of the estimated heritabilities ( $\times 10^2$ )

Scenario			N		50								500							
			Structure		NO				YES				NO				YES			
Selection			Random		Random		Random		Phenotype		Random		Random		Random		Phenotype			
			Mating		Random		Random		Assortative		Random		Random		Random		Assortative		Random	
Generation	Markers	$h^2$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$	$h^2_{Ped}$	$h^2_{Mol}$		
1	10	0.1	16	72	3	5	4	5	4	6	2	78	0	1	0	2	0	2		
		0.5	20	94	9	15	13	16	9	11	3	89	1	4	1	5	2	5		
	20	0.1	26	94	4	4	3	4	4	4	2	38	0	1	0	1	0	1		
		0.5	28	69	11	12	8	8	9	10	3	61	1	3	1	3	2	3		
	100	0.1	18	26	4	4	4	4	4	4	2	8	0	1	0	0	0	0		
		0.5	18	26	11	11	10	10	10	10	1	10	1	1	1	1	2	2		
10	10	0.1	13	68	10	9	9	7	10	16	1	38	0	1	1	1	0	1		
		0.5	25	87	31	20	35	18	20	19	2	68	2	4	5	8	3	4		
	20	0.1	21	73	8	5	9	6	7	5	1	16	0	1	0	1	0	1		
		0.5	24	77	31	16	32	19	19	12	2	22	2	3	3	4	2	3		
	100	0.1	22	32	6	1	7	4	11	5	1	6	0	0	0	0	0	0		
		0.5	27	38	28	14	27	13	28	12	2	8	2	2	4	4	2	3		

N: population size;  $h^2$ : simulated initial trait heritability;  $h^2_{Ped}$ : estimated heritability from genealogical coancestry;  $h^2_{Mol}$ : estimated heritability from marker-based coancestry

these latter had also a substantial amount of bias (Fig. 2). Thus, the biases are not only due to the errors in the estimates of marker-based coancestries but also to the regression method itself applied on true (genealogical) coancestries. Note also that an increasing number of markers improved the estimation, but the benefits were not too large (see Fig. 2).

The slope of the regression of the genealogical coancestry (or inbreeding) on the marker-based coancestry (or inbreeding) estimate is shown in Fig. 3. The expected value of the regression coefficient for a perfect estimator should be one. The observed values for inbreeding were lower than those for coancestry. The regression coefficient was, in general, small but increased with an increase in the number of molecular markers. This increase was more pronounced in populations with a family structure and assortative mating. The coefficient of determination showed the same behaviour as the regression coefficient and therefore is not shown.

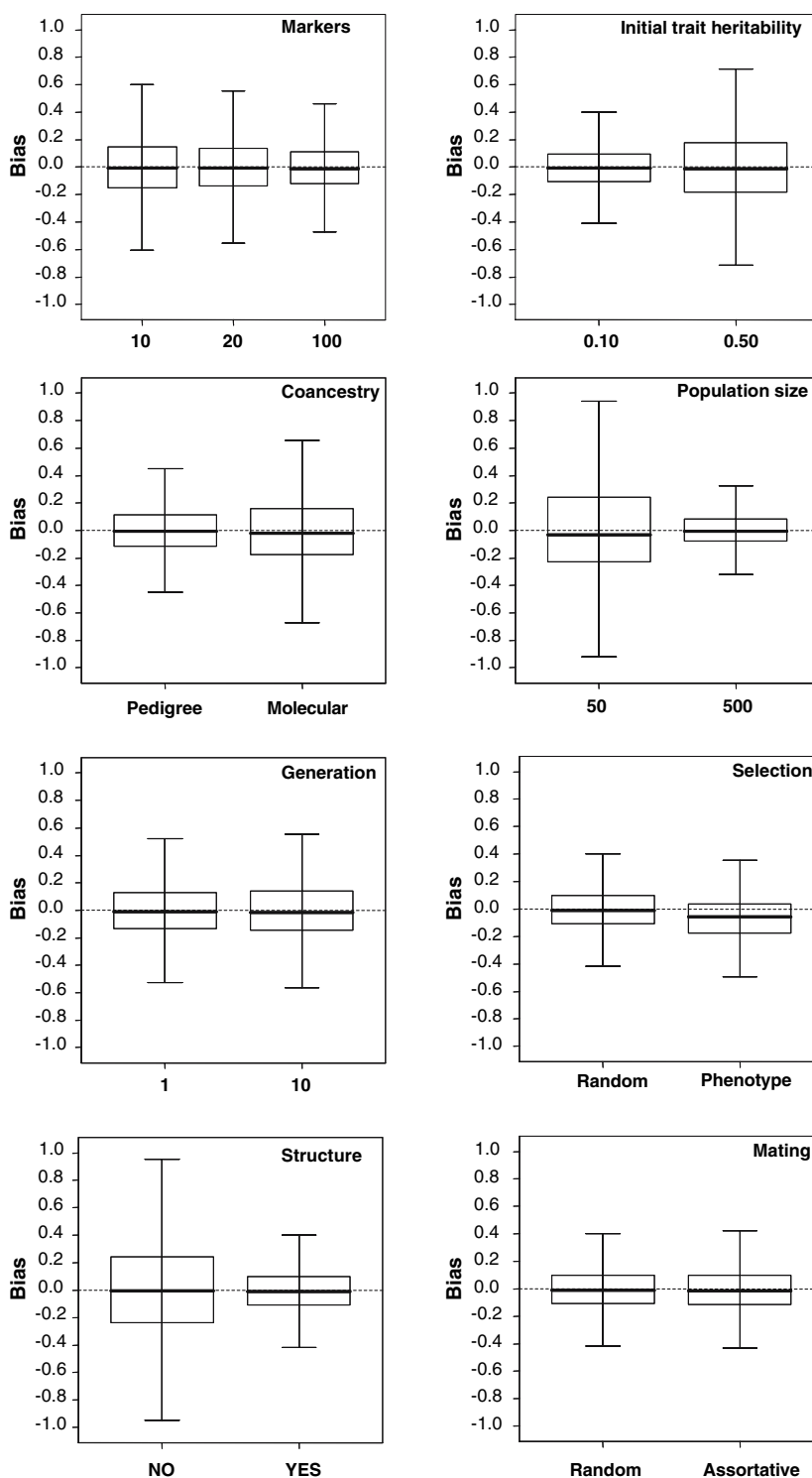
**Discussion**

In the last years there has been a growing interest in the estimation of genetic parameters (such as the heritability and genetic correlations) in populations where pedigree information is not available and must be estimated from molecular markers. One of the marker-based methods to infer quantitative inheritance was the regression method proposed by Ritland (1996a,

2000). This method has been applied in a reduced number of experimental studies (Ritland and Ritland 1996; Mousseau et al. 1998; Klaper et al. 2001; Thomas et al. 2002; Wilson et al. 2003) providing unreliable heritability estimates. Other proposed marker-based approaches implement likelihood-based procedures (Mousseau et al. 1998; Thomas et al. 2000, 2002) and techniques involving an explicit reconstruction of sibling groups were also developed (Thomas and Hill 2000; Thomas et al. 2002; Blouin 2003). However, it remains unclear the most suitable way of exploiting the information obtained from molecular markers to achieve reliable estimates of evolutionary related parameters (see Garant and Kruuk 2005 for a review).

We carried out simulations to evaluate the population characteristics which influence the bias and accuracy of the marker-based method proposed by Ritland (1996a, 2000). Our simulation results suggest that the analyzed regression method supply unreliable heritability estimates in most evaluated scenarios, specially in non structured populations (like natural ones) where the utilisation of the method would be more appealing. The poor performance depends both on the accuracy of the estimation of the relationship between individuals and on the accuracy of the regression approach itself. The latter was tested by implementing the method with the coancestries calculated from the pedigree ( $h^2_{Ped}$ ), which would imply the most favourable situation for the estimation method. However, biased estimates were found even when using these genealogical coancestries.

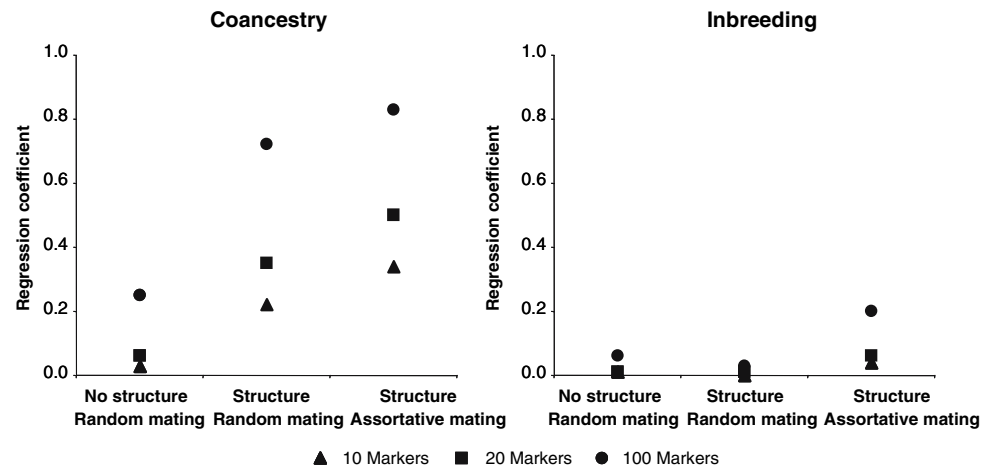
**Fig. 2** Distribution of the bias of the studied factors. The thick line within the boxes denotes the median. The boxes height represents the range between the 25th and 75th percentiles of the bias distribution across scenarios. The vertical bars include all cases with biases up to 1.5 times the interquartile range. This allows us to see whether or not the bias distribution is centred around zero as well as the range of biases across scenarios. All available information, except selection data, was applied to analyse the number of markers, initial trait heritability, type of coancestry, population size and generation factors. For the selection factor analysis, only data from structured populations and random mating were evaluated. In the structure factor analysis, only data from random selection and random mating were used. For the mating factor analysis, only information from structured populations and random selection was examined



The most determining factor in the performance of the heritability estimator appears to be the selection regime. As a consequence of the presence of selection on phenotype, neutral and selective variation could follow different pathways. This is in agreement with the lack of significance for the interaction factor in the

ANOVA between the number of molecular markers and the selection regime (data not shown). A second consequence of the selection on phenotype could be the reduction in the phenotypic similarity variance. As a result, the expected covariance between the phenotypic resemblance and the relationship tends to zero,

**Fig. 3** Slope of the regression of the genealogical coancestry (or inbreeding) on the marker-based coancestry (or inbreeding). Generation 10, 500 individuals



inducing biased heritability estimates. This result was also observed in the estimated heritabilities from genealogical coancestries. The reason could be that the genealogical coancestry, as the marker-based coancestry, refers to neutral loci variation by definition (Malécot 1948).

The second factor explaining the differences in the accuracy of the heritability estimator is the population structure. The influence of this element may be due to the reduced values for the variance of coancestry observed in non-structured populations (Csilléry et al. 2006), leading to low accuracy levels. Possibly, the poor reliability of the heritability estimates obtained in the empirical studies (Ritland and Ritland 1996; Mousseau et al. 1998; Klaper et al. 2001; Thomas et al. 2002; Wilson et al. 2003) could be due to the lack of a family structure and/or the low numbers of markers used. Notwithstanding, small structured populations also showed reduced values for the variance of coancestry in the long-term evaluation, as the reduced number of individuals effectively contributing with progeny led to populations with uniformly and highly related individuals.

Assortative mating according to the genealogical coancestry was carried out in our simulations in order to increase the mean variance of the genealogical coancestry. For example, in populations with a family structure,  $N = 500$ , 10 molecular markers and a simulated initial trait heritability  $h^2 = 0.10$ , the genealogical coancestry variance took values up to 0.0013 at generation 10 for random mating, but under assortative mating the values were increased up to 0.0027 at the same generation. However, this increase in the variance of the genealogical coancestries had little effect, as the estimated heritabilities under assortative mating revealed similar biases and squared errors of estimates (see Fig. 2).

In general, estimates from  $h_{\text{Mol}}^2$  showed lower accuracy than  $h_{\text{Ped}}^2$ , as deduced from the higher average squared error detected for  $h_{\text{Mol}}^2$ . Note, however, that a large amount of bias was also observed in the most favourable situation where the genealogical coancestry is known (i.e. using  $h_{\text{Ped}}^2$ ). This illustrates the deficiencies of the regression method irrespective of the problems in the estimation of the coancestries. However, in cases of reduced variation in genealogical coancestry, molecular information could be useful to discriminate between individuals equally related in genealogical terms but sharing different proportions of alleles. Consequently, in small structured populations with randomly selected parents estimates were better when using marker-based coancestries in a long-term evaluation. There are other aspects that contribute to the improvement of the accuracy of  $h_{\text{Mol}}^2$ . The use of a considerable number of molecular markers and larger population sizes could improve moderately the precision of the estimator (see Fig. 2; Thomas et al. 2000).

The method proposed by Ritland (1996a, 2000) requires the use of the actual variance of the estimated marker-based coancestry, because the marker-based coancestry estimators exhibit a high sampling variance (Van de Casteele et al. 2001; Rodríguez-Ramilo et al. 2006). For example, in populations with a family structure,  $N = 500$ , 10 molecular markers and a simulated initial trait heritability  $h^2 = 0.10$  in a long-term evaluation,  $h_{\text{True}}^2 = h_{\text{Mol}}^2 = 0.10$ . However, if the sampling variance from molecular markers were considered, a substantial reduction in the estimated heritability would be obtained, increasing the bias of the heritability estimator ( $h_{\text{Mol}}^2 = 0.02$ ). Garant and Kruuk (2005) already pointed out that an actual variance for the coancestry is fundamental to apply the regression method.

Although it has been suggested that the knowledge of the population structure to apply the heritability estimator analyzed is not necessary (Garant and Kruuk 2005), our simulation results indicate that this aspect (beside other factors) improves the performance of the heritability estimator. Obviously, when pedigrees are available, reliable heritability values can be estimated by an animal model with a restricted maximum-likelihood estimation procedure (e.g. VCE: Groeneveld and García-Cortés 1998; Lynch and Walsh 1998) instead of the regression method. Moreover, the regression procedure examines relationship at a pairwise level. A restricted maximum-likelihood based estimator accounts for higher order relationship groups from known genealogical coancestries. The animal model takes into account both the inbreeding and the selection, although it requires including all information used to make selection decisions (Fernando and Gianola 1990). Therefore, it is expected that the discrepancies between  $h^2_{\text{True}}$  and  $h^2_{\text{Ped}}$  in small structured populations or in phenotypically selected populations will decrease if an animal model is considered. A more sophisticated statistical approach where molecular coancestry estimates are integrated in the animal model could be valuable (Thomas and Hill 2000).

As a collateral result of this simulations study, it has been analysed the behaviour of the marker-based coancestry estimator implemented (Ritland 1996b) in order to infer the genealogical coancestry. In general, there is not a good agreement between the genealogical and the marker-based coancestry, and the marker-based estimator considered is likely to be more suitable to infer coancestry than inbreeding. The regression coefficient increased with the number of molecular markers, but always below unity. This fact could have several consequences. For example, the amount of genetic diversity will be probably overestimated when selecting least-related individuals on the basis of the estimated coancestry from molecular information (Oliehoek et al. 2006). Our results showed that using marker-based coancestries in structured populations would be more adequate than in panmictic populations (Oliehoek et al. 2006). This latter situation is common in populations of species in need of conservation and, the former circumstance frequently corresponds to domestic species that usually are kept in herds. Nevertheless, the regression and correlation parameters between the genealogical and marker-based coancestries could not be the optimal criterion to test the quality of a coancestry estimator. As suggested by Fernández and Toro (2006) and Oliehoek et al. (2006), a better criterion would be the particular performance of each estimator when coancestries are used for a

particular purpose, for example, the optimal methods of genetic management when different estimators are used.

**Acknowledgments** We thank two anonymous referees for helpful comments on the manuscript. This work was supported by a project of the Plan Estratégico del INIA (CPE03-004-C2).

## References

- Avise JC (1994) Molecular markers, natural history and evolution. Chapman and Hall, New York
- Avise JC, Jones AG, Walker D, DeWoody JA (2002) Genetic mating systems and reproductive natural histories of fishes: lessons for ecology and evolution. *Annu Rev Genet* 36:19–45
- Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *TREE* 18:503–511
- Butler K, Field C, Herbinger CM, Smith BR (2004) Accuracy, efficiency and robustness of four algorithms allowing full sibship reconstruction from DNA marker data. *Mol Ecol* 13:1589–1600
- Csilléry K, Johnson T, Beraldi D, Clutton-Brock T, Coltman D, Hansson B, Spong G, Pemberton JM (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics* 173:2091–2101
- Donner A, Koval JJ (1980) The estimation of intraclass correlation in the analysis of family data. *Biometrics* 36:19–25
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Harlow, UK
- Fernández J, Toro MA (2006) A new method to estimate relatedness from molecular markers. *Mol Ecol* 15:1657–1667
- Fernando RL, Gianola D (1990) Statistical inferences in populations undergoing selection or non-random mating. In: Gianola D, Hammond K (eds) *Advances in statistical methods for genetic improvement of livestock*. Springer-Verlag, Berlin, pp 437–453
- Frankham R, Ballou JD, Briscoe DA (2002) Introduction to conservation genetics. Cambridge University Press, Cambridge, UK
- Garant D, Kruuk LEB (2005) How to use molecular marker data to measure evolutionary parameters in wild populations. *Mol Ecol* 14:1843–1859
- Garant D, Kruuk LEB, McCleery RH, Sheldon BC (2004) Evolution in a changing environment: a case study with great tit fledging mass. *Am Nat* 164:E115–E129
- Groeneveld E, García Cortés LA (1998) VCE 4.0: A (Co)variance component package for frequentists and bayesians. *Proc 6th World Cong Genet. Appl Livest Prod* 27:455–456
- Kempthorne O (1957) An introduction to genetic statistics. John Wiley and Sons, New York
- Klaper R, Ritland K, Mousseau TA, Hunter MD (2001) Heritability of phenolics in *Quercus laevis* inferred using molecular markers. *J Hered* 92:421–426
- 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, MA
- Malécot G (1948) *Les Mathématiques de l'hérédité*. Masson et Cie. Paris, France

- Merilä J, Kruuk LEB, Sheldon BC (2001) Cryptic evolution in a wild bird population. *Nature* 412:76–79
- Mousseau TA, Ritland K, Heath DD (1998) A novel method for estimating heritability using molecular markers. *Heredity* 80:218–224
- 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
- Ritland K (1996a) A marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* 50:1062–1073
- Ritland K (1996b) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genet Res* 67:175–185
- Ritland K (2000) Marker-inferred relatedness as a tool for detecting heritability in nature. *Mol Ecol* 9:1195–1204
- Ritland K, Ritland C (1996) Inferences about quantitative inheritance based upon natural population structure in the common monkeyflower, *Mimulus guttatus*. *Evolution* 50:1074–1082
- Rodríguez-Ramilo ST, Toro MA, Caballero A, Fernández J (2006) Análisis de estimadores basados en información molecular para inferir el parentesco genealógico. *ITEA* 102:192–202
- Thomas SC, Coltman DW, Pemberton JM (2002) The use of marker-based relationship information to estimate the heritability of body weight in a natural population: a cautionary tale. *J Evolution Biol* 15: 92–99
- Thomas SC, Hill WG (2000) Estimating quantitative genetic parameters using sibships reconstructed from marker data. *Genetics* 155:1961–1972
- Thomas SC, Pemberton JM, Hill WG (2000) Estimating variance components in natural populations using inferred relationships. *Heredity* 84:427–436
- Toro MA, Barragán C, Óvilo C, Rodríguez J, Rodríguez C, Silió L (2002) Estimation of coancestry in Iberian pigs using molecular markers. *Conserv Genet* 3:309–320
- Van Kleunen M, Ritland K (2004) Predicting evolution of floral traits associated with mating system in a natural plant population. *J Evol Biol* 17:1389–1399
- Van de Casteele T, Galbusera P, Matthysen E (2001) A comparison of microsatellite-based pairwise relatedness estimators. *Mol Ecol* 10:1539–1549
- Wang J (2002) An estimator for pairwise relatedness using molecular markers. *Genetics* 160:1203–1215
- Wang J (2004) Sibship reconstruction from genetic data with typing errors. *Genetics* 166:1963–1979
- Wilson AJ, McDonald G, Moghadam HK, Herbinge CM, Ferguson MM (2003) Marker-assisted estimation of quantitative genetic parameters in rainbow trout, *Oncorhynchus mykiss*. *Genet Res* 81:145–156