Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits

Pierre de Villemereuil[1]*, Olivier Gimenez[2] and Blandine Doligez[1]

[1]*CNRS; Université de Lyon, 69000 Lyon; Université Lyon 1, France. Department of Biometry and Evolutionary Biology, CNRS, UMR 5558, Bâtiment Gregor Mendel, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France; and* [2]*Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, Campus CNRS, Montpellier Cedex 5, France*

Summary

1. Estimating heritability of traits in wild populations is a major prerequisite to understand their evolution. Until recently, most heritability estimates had been obtained using parent-offspring regressions. However, the popularity of animal models, that is, (generalized) linear mixed models assessing the genetic variance component based on population pedigree information, has markedly increased in the past few years. Animal models are claimed to perform better than parent–offspring regressions mainly because they use full between-individual relatedness information and they allow explicit modelling of the environmental effects shared by individuals. However, the differences between heritability estimates obtained using both approaches are not straight forward, and the factors influencing these differences remain unclear.

2. We performed a simulation study to evaluate and compare the accuracy and precision of estimates obtained from parent–offspring regressions and animal models using both Frequentist (REML, PQL) and Bayesian (MCMC) estimation methods. We explored the influence of (i) the presence and type of shared environmental effects (non-transgenerational or transgenerational), (ii) the distribution of the phenotypic trait considered (Gaussian or binary trait) and (iii) data quantity and quality (sample size, pedigree connectivity) on heritability estimates obtained from the two approaches for different levels of true heritability.

3. In the absence of shared environmental effects, the animal model using the REML method performed best for a Gaussian trait, while the animal model using MCMC was more appropriate for a binary trait. For low quantity and quality data, and a binary trait, the parent–offspring regression yielded very imprecise estimates.

4. Estimates from the parent–offspring regression were not influenced by a non-transgenerational shared environmental effect, whereas estimates from animal models in which environmental effects are ignored were affected by both non-transgenerational and transgenerational effects.

5. We discuss the relevance of each approach and estimation method for estimating heritability in wild populations. Importantly, because most effects fitted in animal models are, in fact, non-transgenerational (including environmental maternal effects), we advocate a systematic comparison between parent–offspring regression and animal model estimates to detect potentially missing non-transgenerational environmental effects.

Key-words: Quantitative genetics < Population Genetics, Bayesian methods < Statistics, Statistics

Introduction

Natural selection is acting solely on individuals' phenotype, whereas individuals mainly pass on genotypes to their offspring. Understanding how genotype shapes phenotype is therefore an essential issue to understand evolution in nature (Ridley 2003). Quantitative genetics is a powerful framework to explore the complex genetic architecture of phenotypic traits (Kruuk, Slate & Pemberton 2008). In wild populations, how much of the observed phenotypic variation on a trait can be transmitted to the next generations is a frequently asked question because it affects the speed and magnitude of trait evolution. The fraction of variability in the phenotypic trait that is of transmittable genetic origin is called heritability (Falconer & Mackay 1996; Roff 1997; Lynch & Walsh 1998). Because heritability is a key genetic parameter in regard to whether natural selection is able to generate evolution on a trait or not, it has been the focus of many studies in various species (Mousseau & Roff 1987; Falconer & Mackay 1996; Kruuk 2004; Roffn, 2007; Hill & Kirkpatrick 2010).

*Correspondence author. E-mail: bonamy@horus.ens.fr

Until recently, two approaches were available and classically used to estimate heritability in wild populations: half-sibling design and parent–offspring regression (Falconer & Mackay 1996; Roff 1997). The half-sibling design method operates by comparing intra- and inter-family variance for half-sibs (the full-sibling design is to be avoided because of its sensitivity to dominance effects, see Lynch & Walsh 1998). In the parent–offspring regression method, the heritability (resp. half of the heritability) of a trait is given by the slope of the regression between the mid-parent phenotype (resp. the phenotypes of one of the parents) and the mean offspring phenotype. Advantages and disadvantages of these two methods are well known and reviewed in Falconer & Mackay (1996), Roff (1997) and Lynch & Walsh (1998). The parent–offspring regression has been more frequently used than the half-sibling design in the wild because it is easier to set up and requires less offspring per individual (Roff 1997) and less information about family structure (e.g. molecularly assigned paternities). In wild populations, the presence of environmental effects shared by related individuals (Wilson *et al.* 2010) and issues related to data quality (Quinn *et al.* 2006; Postma & Charmantier 2007), misassigned paternities (Charmantier & Réale 2005) or imperfect detection (Cam 2009; Papaïx *et al.* 2010) can, however, generate biases or decrease statistical power when estimating heritability. The estimation of heritability in wild populations therefore requires accounting for these specificities, in particular unbalanced sampling designs (Kruuk 2004). Over the last decade, an increasing number of studies (e.g. Réale, Festa-Bianchet & Jorgenson 1999; Milner *et al.* 2000; Merilä, Kruuk & Sheldon 2001; Kruuk, Merilä, Sheldon 2001, Kruuk *et al.* 2002; Sheldon, Kruuk & Merilä 2003; McCleery *et al.* 2004; Charmantier, Keyser & Promislow 2007; Nilsson, Åkesson & Nilsson 2009; Morales *et al.* 2010; Lane *et al.* 2011) have estimated the heritability of traits in the wild using the animal model approach (Kruuk 2004; Postma & Charmantier 2007; Visscher, Hill & Wray 2008). This model was developed in the 1950s (e.g Henderson, 1950, 1976) for animal (and plant) breeding studies, from which it owes its name. The animal model is a (possibly generalized) linear mixed model that uses a pedigree of the population to estimate the additive genetic variance component (and potentially other kinds of genetic effects). The advantages of this approach over the parent–offspring regression and half-sibling design are twofold. First, the animal model is not restricted to specific types of relationships between individuals. Therefore, it maximizes statistical power (Sorensen & Kennedy 1984; Kruuk 2004) and is more robust to inbreeding and selection (Sorensen & Kennedy 1984; van der Werf & de Boer 1990; Sillanpää 2011). Second, the animal model can explicitly account for many confounding effects such as dominance, common environment and parental identity (Kruuk 2004; Wilson *et al.* 2010). Because of its flexibility when dealing with such unbalanced sampling design, the animal model approach has been strongly promoted for estimating heritability of traits in wild populations (Kruuk 2004; Postma & Charmantier 2007; Wilson *et al.* 2010). However, animal models also suffer from the classical pitfalls of mixed models, which are notoriously computationally demanding and sometimes difficult to handle correctly (Bolker *et al.* 2009; Zuur *et al.* 2009).

As a practical demonstration of the advantages of the animal model over the parent–offspring regression, Kruuk (2004) reviewed heritability estimates obtained with both methods in wild populations. She showed that parent–offspring regression estimates were on average 30% higher than those from animal models. Yet, this comparison was criticized by Åkesson *et al.* (2008) who argued that the data sets were too different for the estimates obtained with both methods in different studies to be comparable. Indeed, when restricting to the four studies comparing parent–offspring regressions and animal models for the same data sets, yielding 22 different heritability estimates (Réale, Festa-Bianchet & Jorgenson 1999; MacColl & Hatchwell 2003; Åkesson *et al.* 2008; Hadfield *et al.* 2006), we could find no bias anymore: the heritability estimate was higher for the animal model than for the parent–offspring regression in 13 cases, and lower in eight cases (see Table 1). In a simulation study, in which related individuals shared an environmental effect, Kruuk & Hadfield (2007) showed that the parent–offspring regression performed better in estimating heritability than an animal model in which this environmental effect had not been specified, that is, a 'naive' animal model, and almost as well as an animal model incorporating this effect, that is, an 'informed animal' model. These results may be due to the simulation by the authors of a 'non-transgenerational' environmental effect shared by related individuals (Rossiter, 1996). These non-transgenerational effects are shared by related individuals within the same generation only (e.g. sibs). By contrast, transgenerational effects, that is, effects shared by related individuals between generations (e.g. parents and their offspring), increase the resemblance between parent and offspring and may thus artificially inflate heritability estimates given by parent–offspring regressions. In other words, parent–offspring regressions are expected to give higher estimates than animal models only in the presence of transgenerational environmental effects if relevant information is provided to the animal model (i.e. additional random effect(s) in the model). When referring to a transgenerational effect here, we will exclusively consider parents and their offspring, excluding other potential levels of relatedness between individuals (e.g. for grandmaternal effects), which have not been commonly investigated in wild populations using animal models so far.

The previous reasoning is valid for any kind of data distribution. Yet, new difficulties arise when using non-Gaussian distribution, especially for all-or-none (i.e. binary) data, as standard methods become irrelevant, because of intrinsic normality assumptions (e.g. classical parent–offspring regression) or issues in likelihood computation (e.g. REML). To relax normality assumptions, methods based on threshold models are used (Wright, 1934; Dempster & Lerner1950; Elston, Hill & Smith 1977; Gianola, 1982; Lynch & Walsh, 1998). These models assume an underlying continuous character and a threshold value triggering the presence of the all-or-none trait. The statistical properties of parent–offspring regressions using threshold models are well understood (van Vleck, 1972; Elston,

Table 1. Difference between estimates of heritability of continuous traits obtained from the parent–offspring regressions (RegPO) and animal models (AM) on the exact same data sets in the four studies showing both estimates published so far (we included only studies based on observed pedigree). Positive values indicate higher values for the animal model. Significance of differences could not be tested, because information on sample size is not always available. Marked (*) heritability are considered as zero for the differences computation. Additionnal random effects ('Effects' column) are either maternal effects (M) or brood-litter effects (BL, also called 'nest effect' in the cited articles)

Phenotypic trait	Effects	RegPO	AM	Differences	Refs
Body mass (lambs, June)	—	0·00	0·31	0·31	1
Body mass (lambs, September)	—	0·02	0·29	0·27	1
Body mass (yearling, June)	—	0·12	0·43	0·31	1
Body mass (yearling, September)	—	0·07	0·24	0·17	1
Body mass (2-year-old, June)	—	−0·15*	0·03	0·03	1
Body mass (2 years old, September)	—	−0·09*	0·00	0·00	1
Body mass (3 years old, June)	—	0·28	0·27	−0·01	1
Body mass (3 years old, September)	—	0·49	0·51	0·02	1
Body mass (4 years old, June)	—	0·26	0·23	−0·03	1
Body mass (4 years old, September)	—	0·59	0·34	−0·25	1
Body mass (adult, June)	M	0·39	0·28	−0·11	1
Body mass (adult, September)	M	0·57	0·81	0·24	1
Parental effort	BL	0·59	0·43	−0·16	2
Cap color	BL	0·06	0·10	0·04	3
Wing length	M, BL	0·76	0·72	−0·04	4
Wing projection	M, BL	0·47	0·48	0·01	4
Tail length	M, BL	0·68	0·81	0·13	4
Bill depth	M, BL	0·07	0·07	0·00	4
Bill width	M, BL	0·39	0·46	0·07	4
Bill length	M, BL	0·97	0·84	−0·13	4
Skull length	M, BL	0·44	0·32	−0·12	4
Tarsus length	M, BL	0·72	0·73	0·01	4

[1](Réale *et al.* 1999), [2](MacColl & Hatchwell 2003), [3](Hadfield *et al.* 2006), [4](Åkesson *et al.* (2008)). For the latter study, we used estimates from the animal model called 'mean animal model', instead of the 'repeated measures animal model', for the relevance of the comparison with the parent–offspring regression (e.g. same phenotypic variance).

Hill & Smith 1977; Lynch & Walsh, 1998; Roff, 1997). On the contrary, the behaviour of some estimation methods based on animal models needs further investigation for binary data (e.g.

Charmantier, Keyser & Promislow 2007). For instance, Charmantier *et al.* (2011) obtained contradictory results on the heritability of natal dispersal behaviour in the wandering albatross when comparing a parent–offspring regression and an animal model and between different estimation methods used to fit the animal model. Numerous studies of heritability of binary traits in the wild have been published using only parent–offspring regression (Hansson, Bensch & Hasselquist 2003; Doligez, Gustafsson & Pärt 2009), animal model (Thériault *et al.* 2007; Wilson *et al.* 2011; Reid *et al.* 2011a, b) or both approaches (Charmantier, Keyser & Promislow 2007; Doligez *et al.* 2012; Charmantier *et al.* 2011). Yet, despite the growing number of heritability estimates for binary traits using animal models, we are still lacking statistical studies comparing different estimation methods.

In order to address these different issues, we conducted a comprehensive simulation study carrying out a statistical comparison of the performance of parent–offspring regressions and animal models in estimating heritability. We assessed the influence of different factors on this comparison. First, we simulated contrasted conditions to assess the influence of environmental effects shared by individuals: (i) no shared environment, (ii) share of a non-transgenerational environmental effect and (iii) share of a transgenerational environmental effect. Second, we investigated the heritability estimation for both a continuous and a binary trait, using several popular estimation methods to fit the animal model to Gaussian or binomial data. We also investigated in each case the influence of data quality and quantity (Quinn *et al.* 2006) on the bias and precision of heritability estimates by simulating (i) a large and a small-size data set with (ii) a high and a low level of knowledge about the genetic relationships between individuals (i.e. a fully connected pedigree and a sparsely connected pedigree with many missing relationships). Finally, we investigated estimates for low, medium and high true heritability levels of the traits, in particular because of possible boundary effects for low heritability level, which could lead to shifts in accuracy or precision. Our results are discussed along some results from other studies, and conclusions are drawn about the relevance of each approach and method in the form of advice to the practitioner.

Material and methods

SCENARIOS FOR SIMULATION OF PEDIGREES AND PHENOTYPES

We investigated the influence of three parameters on the estimation of heritability by different methods: (i) the true heritability level of the trait, using three levels: 0.5 (high heritability), 0.3 (moderate heritability) or 0.1 (low heritability); (ii) sample size, using two levels of population size: 125 individuals per generation ('large' sample) or 25 individuals per generation ('small' sample); and (iii) pedigree connectivity level, using two levels: full connectivity (all parental relationships were known) or sparse connectivity (a fraction of relationships were missing, see below). The values of these parameters were chosen to be

realistic in regard to field studies. By combining these levels, we obtained 12 different scenarios. For each of these scenarios, we considered three different situations: (i) related individuals shared no common environmental effect, (ii) related individuals shared a common but non-transgenerational environmental effect, here a resemblance between individuals from the same mother with no link to the mother's phenotype or genotype, and (iii) related individuals shared a common transgenerational effect, here through a common breeding patch for a given fraction of parents and offspring (see below). For each scenario and situation, we simulated 1 000 pedigrees and the associated phenotypes on which we estimated heritability. The phenotypes were either normally distributed or binary. In total, we simulated 72 000 different pedigrees and associated phenotypic data sets.

To simulate pedigrees, we considered a closed and saturated population (i.e. no immigration) with nonoverlapping generations (i.e. the whole population was replaced by local recruits at each generation), breeding on 30 different patches for the large population and six patches for the small population (approximately four individuals per patch). Pedigree depth was kept constant to eight generations. First, we randomly assigned a breeding patch to each individual of the first generation. For each of the following generations, we again randomly assigned a breeding patch to each individual and randomly drawn its parents within the patch. We then considered a natal dispersal rate of 0.5, that is, offspring have a probability of 0.5 to remain associated to the same breeding patch as their parents. The resulting average number of offspring per pair was 1.5, with a maximum between 5 and 7 (depending on scenarios). As there is no juvenile mortality in our simulations, this is the actual number of recruited offspring. When generating sparsely connected pedigrees, we first generated a fully connected pedigree to simulate individual phenotypes (see below). Then, we randomly selected 50% of sires and 20% of dams in the pedigree and reported them as missing values when estimating heritability. To include breeding patches in the simulation process generating fully connected pedigrees, we modified the R function `generatePedigree()` from the GeneticsPed package (Gorjanc *et al.* 2007, see Appendix A). We used the `fpederr()` function from the pedantics package (Morrissey & Wilson 2009), which randomly generated missing data in a full pedigree, to make them sparse when needed.

Using these pedigrees, we simulated either a normally distributed phenotypic trait or a binary phenotypic trait. For the normally distributed trait, the phenotype y_i of individual i was obtained using the following equation:

$$y_i = \mu + a_i + e_i \qquad \text{eqn 1}$$

where μ was the mean phenotype in the population, whose value was arbitrarily set to 10, a_i was the breeding value for the individual i, normally distributed assuming genetic additive variance V_A and e_i was a residual ('environmental') variation, normally distributed with variance one. We used the `rbv` function from the R package MCMCglmm (Hadfield 2010a) to

compute the breeding value a_i according to the simulated pedigree and V_A. This function includes a Mendelian random deviation for each offspring. This corresponds to the simplest version of the animal model (Kruuk 2004; Wilson *et al.* 2010). The value V_A varied according to the true heritability level investigated and was the only variance component to vary across scenarios. When individuals shared a common environmental non-transgenerational maternal effect, we added the effect of the identity of the mother k on the phenotype of the offspring i:

$$y_i = \mu + a_i + m_k + e_i \qquad \text{eqn 2}$$

This effect was non-transgenerational because by adding the identity of the mother as a random effect, we assumed a resemblance among the offspring of the same female, but not between the female and her offspring. When individuals shared a common transgenerational breeding patch effect, we added the effect of the breeding patch j on the phenotype of the individual i:

$$y_i = \mu + a_i + p_j + e_i \qquad \text{eqn 3}$$

The variance of both the maternal and breeding patch effects was set to one. Because V_A was the only variance component that could be changed to reach the desired true level of heritability, the total phenotypic variance V_P varied across scenarios. The computation of V_P was always based on all variance components fitted in the model.

For the binary distributed phenotypic trait, we adjusted the value of μ to obtain an arbitrary proportion of 1/3 of individuals with a phenotype 1 (and 2/3 with phenotype 0) in the population; this proportion was reached for $\mu = -0.41$. We used the same equations as for a normally distributed phenotype and subsequently defined the binary phenotype using a threshold as follows:

$$y_i^{0,1} = \begin{cases} 1 & \text{if} \quad y_i > 0 \\ 0 & \text{if} \quad y_i < 0 \end{cases} \qquad \text{eqn 4}$$

ESTIMATION OF HERITABILITY

Heritability was estimated using either a mid-parent–mean-offspring regression (Falconer & Mackay 1996; Roff 1997; Lynch & Walsh 1998), hereafter abbreviated RegPO in tables and figures, or an animal model (Lynch & Walsh 1998; Kruuk 2004). For the continuous phenotypic trait and using the animal model, we computed estimates using either the widely used restricted maximum likelihood method (REML, Patterson & Thompson 1971; Knott *et al.* 1995) or Markov Chain Monte-Carlo method (MCMC, Sorensen & Gianola 2002; Hadfield 2010b; Gelman *et al.* 2004). These animal models follow the formalism described in eqns 1–3, respectively, for situations where individuals shared no common environmental effect, a non-transgenerational maternal effect and a transgenerational breeding patch effect.

For the binary phenotypic trait in the parent–offspring regression, we used the method described in Roff (1997):

heritability was first estimated as if the trait was normally distributed (noted $\hat{h}^2_{\text{obs}}$), and a correction was then applied to the estimate obtained (Dempster & Lerner 1950; Elston, Hill & Smith 1977; Lynch & Walsh 1998) based on the threshold model hypothesis. To account for the binary distribution, this correction uses the proportion p of phenotypes 1 (e.g. presence of the character) to compute the value of the normal standard curve z at the threshold corresponding to p. The heritability $\hat{h}^2_{\text{obs}}$ was then corrected as follows:

$$\hat{h}^2 \simeq \frac{p(1-p)}{z^2} \hat{h}^2_{\text{obs}} \qquad \text{eqn 5}$$

For the binary phenotypic trait in animal models, we used three different estimation methods. (i) Based on the correcting method described above, we estimated heritability using REML as if the trait was normally distributed and then applied the same correction (Charmantier, Keyser & Promislow 2007). This method is thereafter abbreviated REMLc (for 'corrected REML'). As far as we know, this method has never been properly validated. (ii) We used penalized quasi-likelihood (PQL, Breslow & Clayton 1993; Breslow & Lin 1995). This method is known for generally underestimating variance components for binary data (Rodriguez & Goldman 2001; Callens & Croux 2005; Gilmour *et al.* 2006). However, it is currently the default estimation method of the ASReml software (Gilmour *et al.* 2006). (iii) We also resorted to MCMC methods (Hadfield 2010a). Animal models for the binary trait were built using the following equation (here for the situation where individuals shared no common environmental effect):

$$l_i = \mu + a_i + e_i \qquad \text{eqn 6}$$

where l_i is a normally distributed hypothetical trait for individual i, called liability. Because binary data do not provide enough information to infer liability variance, we fixed the residual variance to 1. The probability of displaying phenotype 1 was linked to the liability through a probit link function, such that:

$$P(y_i^{0,1} = 1) = \text{probit}^{-1}(l_i) \qquad \text{eqn 7}$$

Fitting this model allowed us to get an estimate $\hat{V}_A$ of the additive genetic variance. We then calculated the heritability as:

$$\hat{h}^2 = \frac{\hat{V}_A}{\hat{V}_A + 1 + 1} \qquad \text{eqn 8}$$

where the first 1 in the denominator stood for the residual variance and the second 1 for the 'probit link' variance. This last term was needed to estimate heritability on the liability scale (Nakagawa & Schielzeth 2010) corresponding to the estimate given by the parent–offspring regression (eqn 5). For the MCMC estimation of the heritability of a continuous trait, we used the usual inverse-Gamma(0.001,0.001) distribution as the prior distribution for variance components. After pilot studies, we let the MCMC run for 500 000 iterations with a thinning interval of 10 after a burn-in of 10 000 in order to obtain effective sampling sizes around 20 000. For the MCMC estimation in the case of a binary trait, we used a χ^2 distribution with one degree of freedom as the prior

distribution rather than the inverse-Gamma distribution. This choice was motivated both by software constraints and by the fact that the inverse-Gamma resulted in a prior distribution for heritability in which too much weight was put on the value 1 (see Appendix B), whereas the χ^2 resulted in a more balanced distribution. For a binary trait, we let the MCMC run for 1 million iterations with a thinning interval of 100 after a burn-in of 10 000 in order to obtain effective sampling sizes around 3 000. We used the posterior median as point estimate of the posterior distribution.

We performed all computations using the R statistical software (R Development Core Team 2011), except for REML, REMLc and PQL, which were implemented using the ASReml software (Gilmour *et al.* 2006). MCMC estimations were made using the MCMCglmm R package (Hadfield, 2010a). Results are presented here for the heritability solely, but results about additive genetic variance are presented in Appendix C.

ASSESSMENT OF THE METHODS FOR ESTIMATING HERITABILITY

For each set of 1 000 simulations, we computed the mean and 2.5, 25, 75 and 97.5% quantiles to quantify the bias and the dispersion of heritability estimates. We also calculated the root mean square error (RMSE) defined as the square root of the mean square difference between the estimate and its true value: $\sqrt{\text{E}[(\hat{h}^2 - h^2)^2]}$. This statistics quantifies both bias and precision and is a measure of the quality of an estimator (Bolker 2008). A small RMSE indicates that the estimator is close to its true value both in terms of bias (systematic error) and precision (random error). We also calculated coverage (Bolker, 2008) of 95% confidence intervals (for regression, REML and REMLc) and 95% credible intervals (for MCMC). Coverages were computed as the proportion of times these intervals contained the true value of heritability. By definition, for a $1-\alpha$ level of certainty, the coverage of confidence or credible intervals should be $1-\alpha$. A coverage above (resp. below) this predicted value indicates a conservative (resp. anti-conservative) confidence or credible interval. Confidence intervals were calculated assuming a normal distribution of estimates ($\hat{h}^2 \pm 1.96$ s.e.). Credible intervals were calculated as highest posterior density intervals using the `HPDinterval()` function from the coda R package (Plummer *et al.* 2006).

Results

HERITABILITY ESTIMATION IN THE ABSENCE OF SHARED ENVIRONMENTAL EFFECTS

In the case of a continuous phenotypic trait, both parent–offspring regression and animal model performed similarly in estimating heritability (Fig. 1) although parent–offspring regression tended to have a larger dispersion for small sample size and sparsely connected pedigree (Fig. 1, last column; the 95% interquantile interval is systematically larger for the parent–offspring regression than for the two other methods).

Heritability was accurately estimated (i.e. mean values close to the true value), with a bias usually less than the second digit order, except for small sample sizes, for which the estimates of the parent–offspring regression and animal model using the MCMC estimation method were slightly biased downwards (Fig. 1, last two columns; bias up to 0.25 for the parent–offspring regression and 0.03 for the MCMC estimation method). Because animal model estimates are bounded to positive values, this model yielded lower dispersion of estimates for low level of true heritability, while estimates of the parent–offspring regression could be negative (Fig. 1, last column, last row). Even when constraining heritability estimates from parent–offspring regressions to be positive (e.g. by setting negative values to zero), the precision of estimates would remain lower for the parent–offspring regression than the animal model. RMSE values were the lowest for the animal model using the REML estimation method, showing that it was the best estimation method. For low level of true heritability and small sample size, the MCMC estimation method, however, performed better (Table 2), which reflects its better precision for low heritability level (Fig. 1, last row, $h^2 = 0.1$). The parent–offspring regression approach performed poorly when compared with animal model–based estimation methods, but the difference was lower when data quantity and/or quality was high (i.e. large sample size and/or full pedigree connectivity) and increased when data quantity and/or quality decreased.

When estimating heritability of a binary phenotypic trait, the dispersion of heritability estimates increased in most cases (Fig. 2), especially for small sample size and/or sparsely connected pedigree (Fig. 2, last three columns). This was due to the low information contents of binary data,
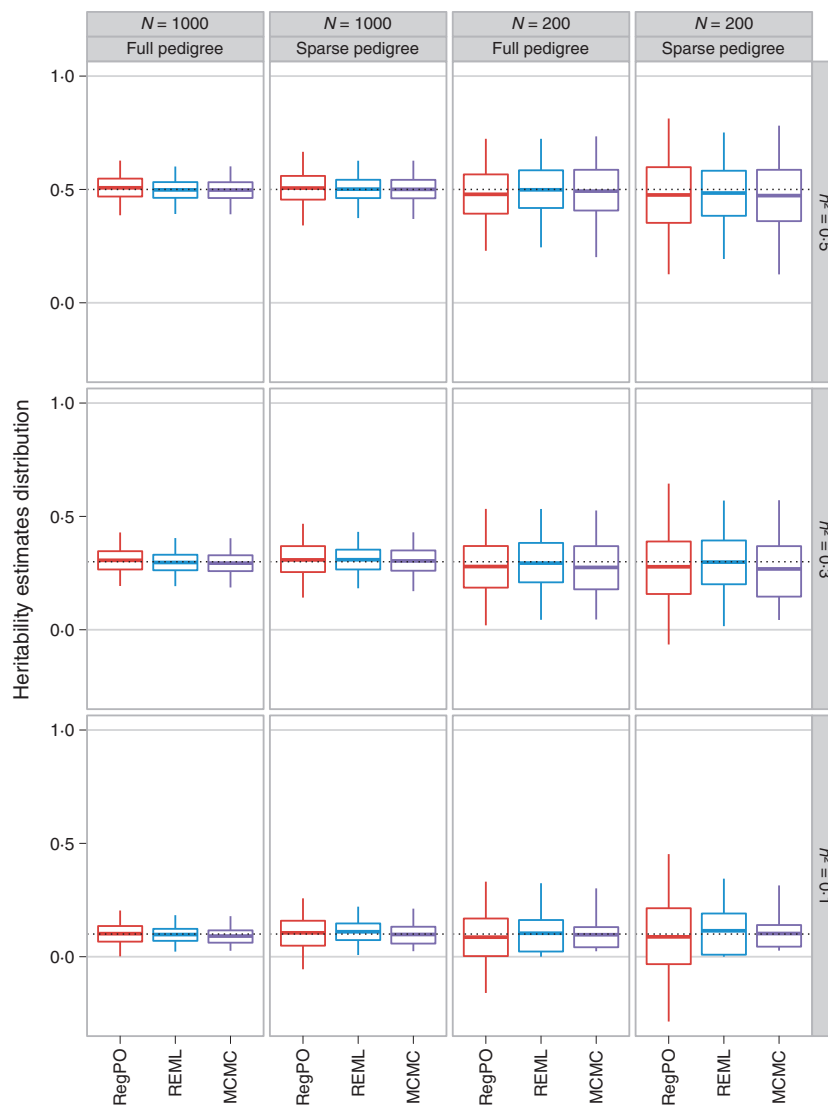
Fig. 1. Distribution of heritability estimates of a continuous trait for the different approaches and estimation methods depending on the true level of heritability (h^2, dotted lines), pedigree connectivity (full and sparse) and sample size (N=1 000 or N=200). The simulations did not include environmental effect shared by related individuals. Boxes represent 25 and 75% quantiles, and the middle line represents the mean. Whiskers show 2.5% (bottom) and 97.5% (top) quantiles. Red boxes show distribution of estimates from the parent–offspring regression (RegPO), bluish boxes show distributions of estimates from the animal model using either REML (light blue) or MCMC (violet) estimation methods.

Table 2. Root mean square error (RMSE) of heritability estimates of a continuous trait for the different approaches and estimation methods depending on the true level of heritability (h^2), sample size (N) and level of pedigree connectivity (full or sparse). The simulations included no environmental effect shared by related individuals

| h^2	0·5				0·3				0·1			
N	1 000		200		1 000		200		1 000		200	
Pedigree	Full	Sparse	Full	Sparse	Full	Sparse	Full	Sparse	Full	Sparse	Full	Sparse
RegPO	**0·061**	0·081	**0·131**	0·177	**0·06**	0·084	**0·133**	0·179	0·052	0·081	0·122	0·187
REML	**0·053**	**0·064**	**0·122**	**0·144**	**0·054**	**0·064**	**0·125**	**0·138**	**0·04**	**0·054**	0·093	0·108
MCMC	**0·054**	**0·065**	**0·132**	0·168	**0·055**	**0·066**	**0·133**	0·15	**0·041**	**0·051**	**0·075**	**0·082**

Estimation methods: RegPO, parent–offspring regression; REMLc, animal model with corrected restricted maximum likelihood; PQL, animal model with penalized quasi-likelihood; MCMC, animal model with Monte-Carlo Markov Chain estimation. The smallest RMSE are shown in bold for each set of simulations. Because of the low precision expected for heritability estimates, we decided that a difference with the smallest RMSE of less than 0·01 was nonsignificant.

which amplified sampling variance. The parent–offspring regression approach and the animal model using REMLc yielded a particularly large dispersion of estimates (e.g. Fig. 2, last column, first row: the 95% interquartile interval covers the whole [0,1] interval). The animal model using PQL strongly underestimated heritability in all cases, even though the bias decreased with decreasing true level of heritability. This was consistent with the known tendency of PQL to underestimate large variance components (Gilmour et al.2006). With small sample size (Fig. 2, last two columns), the estimates given by the animal model using MCMC were more biased, or as biased as, compared with the parent–offspring regression or animal model using REMLc (towards lower and higher values for high and low levels of heritability respectively). However, the animal model using MCMC yielded a lower level of estimate dispersion (i.e. smaller 95% interquartile intervals). This was confirmed by the comparison of RMSE (Table 3), for which the animal model using MCMC had the lowest values, except for small sample size and low true heritability level (Table 3, $h^2 = 0.1$, last column). In this case, PQL yielded higher-quality estimates (i.e. smallest RMSE) because of a smaller dispersion and a decrease in bias with decreasing heritability level. The bias observed for the MCMC estimation method for small sample size was most likely due to prior sensitivity issues (see Appendix B). For both continuous and binary phenotypic traits, all approaches and estimation methods led to anti-conservative coverage in most cases, with 95% confidence/credible intervals excluding the true heritability value in more than 5% of the simulations (Fig. 3; the animal model with PQL is not shown, because the bias caused coverage to be consistently wrong in this case). The animal model performed better for high-quantity and quality data (i.e. large sample size and fully connected pedigree), as may be expected. Conversely, the parent–offspring regression led to a better coverage for a sparsely connected pedigree. For low level of heritability ($h^2 = 0.1$) and low sample size, the animal model with MCMC was conservative (i.e. above expected 95% coverage).

HERITABILITY ESTIMATION IN THE PRESENCE OF SHARED NON-TRANSGENERATIONAL AND TRANSGENERATIONAL EFFECTS

The comparison of heritabilityestimates in the presence of a shared non-transgenerational environmental effect (i.e. common mother for siblings) and a shared transgenerational environmental effect (i.e. common breeding patch for parents and offspring) showed that parent–offspring regression yielded biased estimates only when the common environment shared by individuals was transgenerational (Figs 4 & 5: here and further, results are discussed for high-quantity and quality data, i.e. $N=1$ 000 and fully connected pedigree). This result confirmed that estimates from parent–offspring regression were not sensitive to non-transgenerational effects (see Introduction).

For the continuous phenotypic trait, the animal model yielded unbiased heritability estimates only when the shared environmental effect was specified in the model (i.e. informed animal model, Fig. 4). The bias of estimates from the animal model in which no shared environmental effect had been specified (i.e. naive animal model) increased when the true level of heritability decreased (Fig. 4). This is a consequence of the values of variance used for simulations: for low true heritability, true V_A was lower whereas the variance for additional effects remained fixed to 1, hence increasing the relative importance of the latter. The bias of estimates from the naive animal model was also lower in the presence of a non-transgenerational (maternal) compared with a transgenerational (breeding patch) effect (Fig. 4). This last result was due to the difficulty in generating a strong non-transgenerational maternal effect in our simulations, as nonoverlapping generations led to a reduced total number of offspring per mother. Indeed, a preliminary study showed the strength of a non-transgenerational maternal effect to be very sensitive to the number of offspring per mother. Results for small sample size and/or sparsely connected results were qualitatively similar, although the estimation of the additive genetic variance was biased downward (up to 20%) for the informed animal model. Furthermore, the variance of the maternal effect was overestimated (leading to an additional underestimation of the additive genetic variance)
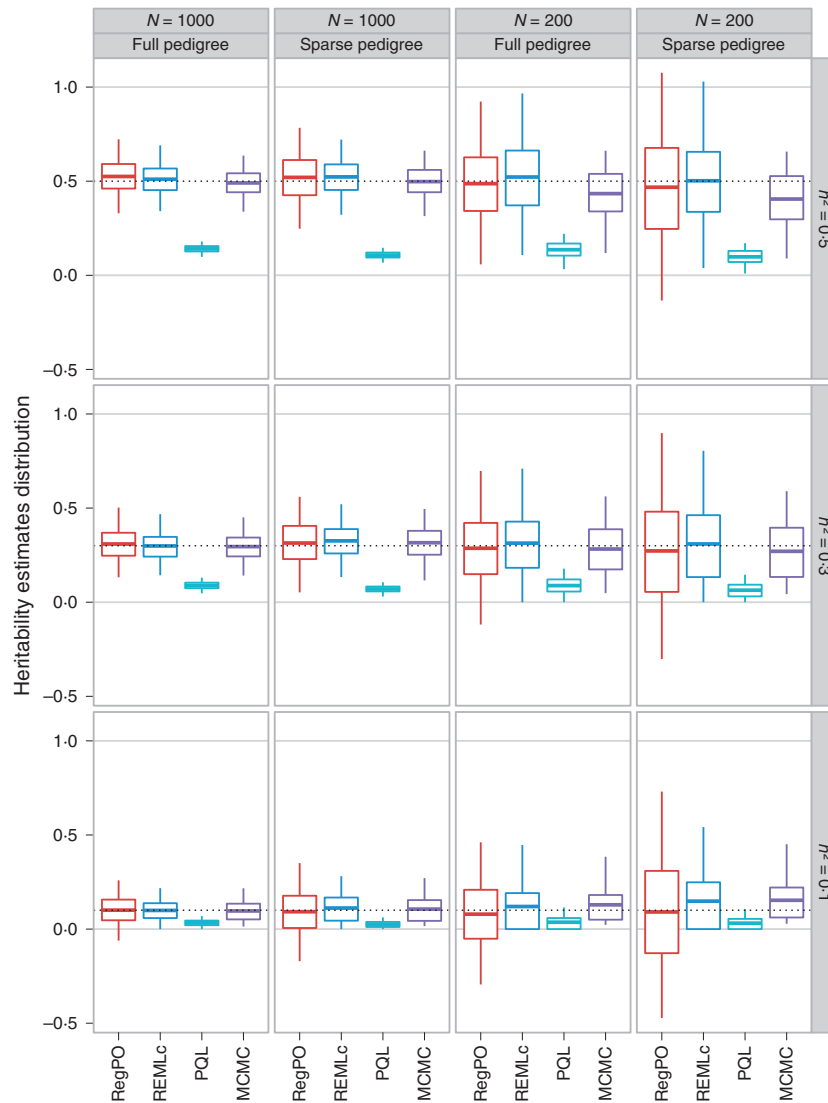
Fig. 2. Distribution of heritability estimates of a binary trait for the different approaches and estimation methods depending on the true level of heritability (h^2, dotted lines), pedigree connectivity (full and sparse) and sample size ($N=1\,000$ or $N=200$). The simulations did not include environmental effect shared by related individuals. Boxes represent 25 and 75% quantiles, and the middle line represents the mean. Whiskers show 2.5% (bottom) and 97.5% (top) quantiles. Red boxes show distribution of estimates from the parent–offspring regression (RegPO), bluish boxes show distributions of estimates from the animal model using REMLc (light blue), PQL (turquoise) or MCMC (violet) estimation methods.

Table 3. Root mean square error (RMSE) of heritability estimates of a binary trait for the different approaches and estimation methods depending on the true level of heritability (h^2), sample size (N) and level of pedigree connectivity (full or sparse). The simulations included no environmental effect shared by related individuals

h^2	0·5				0·3				0·1			
N	1 000		200		1 000		200		1 000		200	
Pedigree	Full	Sparse	Full	Sparse	Full	Sparse	Full	Sparse	Full	Sparse	Full	Sparse
RegPO	0·101	0·137	0·221	0·315	0·093	0·134	0·208	0·308	0·084	0·131	0·194	0·315
REMLc	0·09	0·107	0·216	0·245	**0·08**	**0·102**	0·186	0·221	**0·057**	**0·081**	0·131	0·172
PQL	0·36	0·393	0·367	0·404	0·212	0·231	0·217	0·24	0·069	**0·076**	**0·073**	**0·076**
MCMC	**0·078**	**0·09**	**0·156**	**0·183**	**0·077**	**0·097**	**0·142**	**0·159**	0·056	**0·073**	0·106	0·127

Estimation methods: RegPO, parent–offspring regression; REMLc, animal model with corrected restricted maximum likelihood; PQL, animal model with penalized quasi-likelihood; MCMC, animal model with Monte-Carlo Markov Chain estimation. The smallest RMSE are shown in bold for each set of simulations. Because of the low precision expected for heritability estimates, we decided that a difference with the smallest RMSE of less than 0·01 was nonsignificant.

when using a sparsely connected pedigree (see Appendix D for more details).

For the binary phenotypic trait, the patterns were similar to the continuous trait, except that biases appeared for estimates from the informed animal model when the true level of heritability was high or medium (Fig. 5, $h^2 = 0.5$ or $h^2 = 0.3$). Surprisingly, the direction of the bias differed among estimation methods (downward using MCMC, upward using REMLc). For the MCMC estimation method, because both the additive genetic variance and the variance of the other random (maternal or environmental) effect were biased downward, and because the bias was stronger for small sample size (see Appendix D), we expect the prior sensitivity to be at the origin of the bias in heritability estimates. Unfortunately, we could not test this hypothesis, due to the impossibility to use uniform priors in the MCMCglmm package. Concerning the REMLc

estimation method, the upward bias in estimates was observed only in the presence of shared environmental effects and was due to an underestimation of the variance of the environmental effect combined with an overestimation of the additive genetic variance (see Appendix D). The animal model using PQL largely underestimated heritability, as found above in the absence of shared environment effects.

Discussion

WHICH APPROACH AND ESTIMATION METHOD FOR ESTIMATING HERITABILITY IN THE ABSENCE OF SHARED ENVIRONMENTAL EFFECTS?

In the case of a continuous trait in the absence of shared environmental effects, our simulations showed little difference
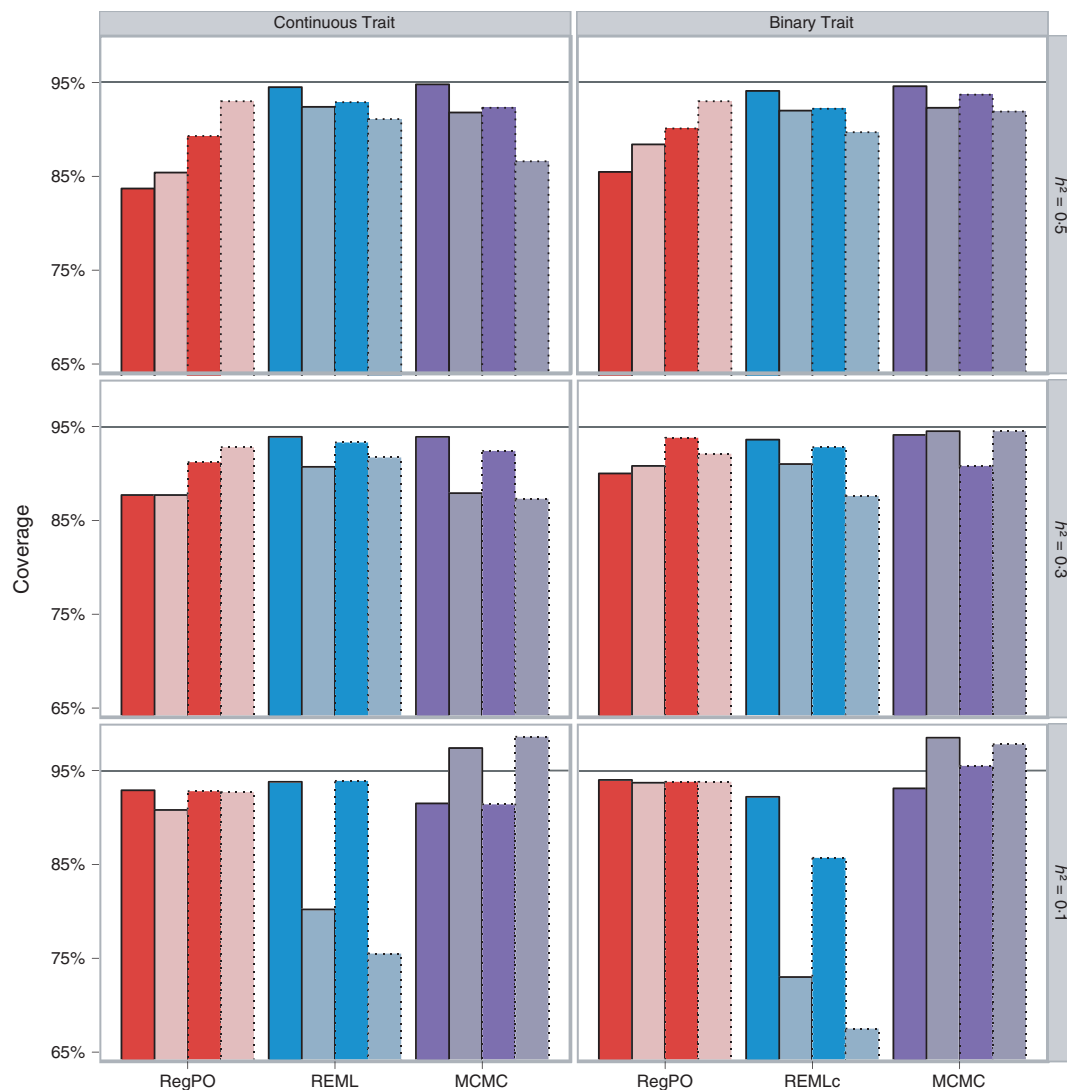

Fig. 3. Coverage of the 95% confidence (for parent–offspring regression (RegPO) and animal model using REML and REMLc) or credible (for animal model using MCMC) interval of heritability estimates over the simulations for each set of conditions and for the continuous (left) and binary (right) phenotypic trait. The simulations included no environmental effect shared by individuals. For the binary trait, the coverage of 95% confidence interval of estimates from the animal model using PQL is not shown. Bright bars: large sample size; grey tint bars: small sample size. Solid borders: fully connected pedigree; dotted borders: sparsely connected pedigree. Expected coverage is 95% (horizontal line).
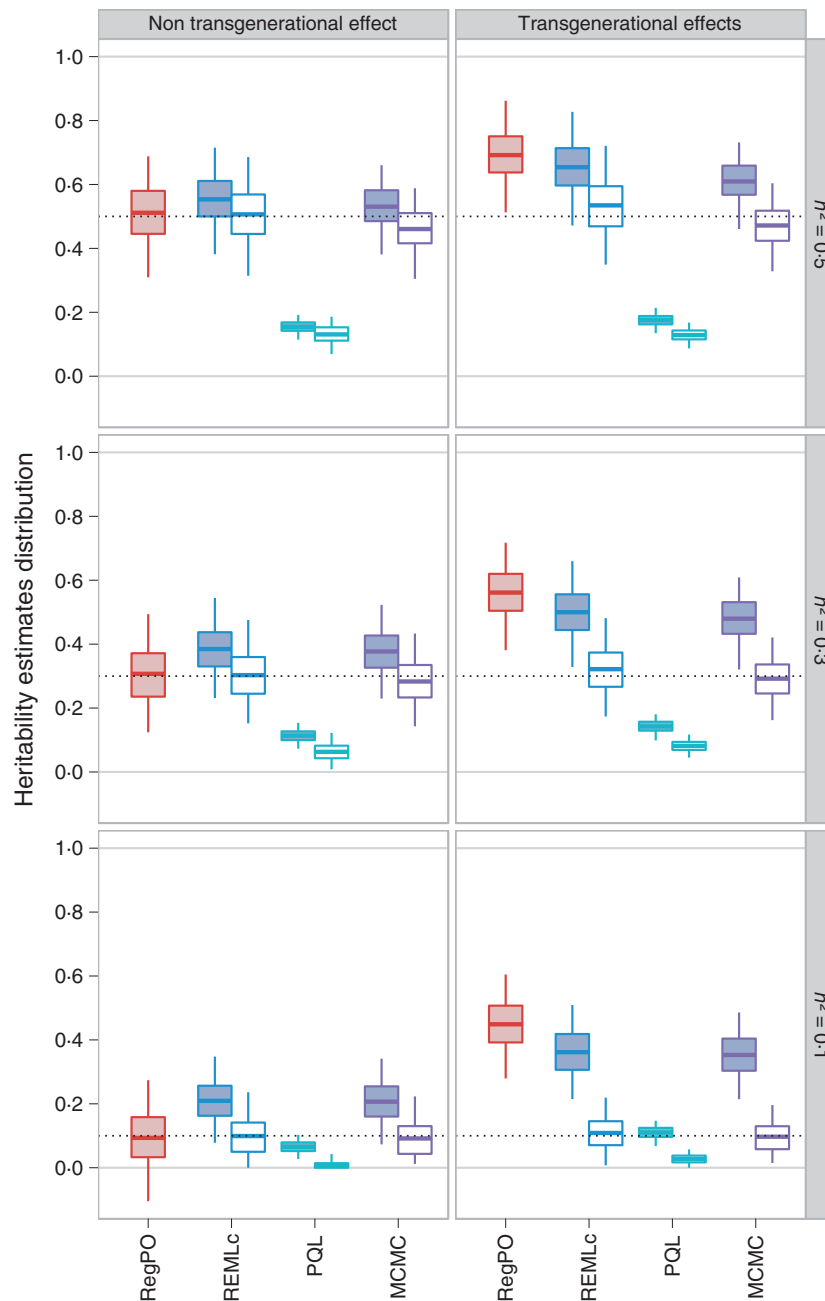
Fig. 4. Distribution of heritability estimates of a continuous trait for the different approaches and estimation methods depending on the true level of heritability (h^2, dotted lines) and the environmental effect shared by individuals (non-transgenerational (left) or transgenerational (right) effect); results obtained for a fully connected pedigree and a large sample size (1 000 individuals). Boxes represent 25 and 75% quantiles, and the middle line represents the mean. Whiskers show 2.5% (bottom) and 97.5% (top) quantiles. Filled boxes: naive models (including parent–offspring regression); open boxes: informed animal models (see text).

in heritability estimates between the parent–offspring regression and animal model, and little difference between the REML and MCMC estimation methods used in the animal model. However, the animal model using REML seemed the most accurate and precise estimation method. The parent–offspring regression showed more variability in heritability estimates for low quantity and/or quality data, that is, here for low sample size and/or low pedigree connectivity. This increased variability originates from the partial use of the information about the relatedness between individuals, leading to a smaller statistical power. Importantly, in the presence of a sparsely connected pedigree, the sample size is impacted as well: for our large sample size simulations (N=1 000), the actual sample size for parent–offspring regressions varied between 302 and 400 and was 350 on average; for our small sample size simulations (N=200), the actual sample size varied between 49 and 92 and was 70 on average.
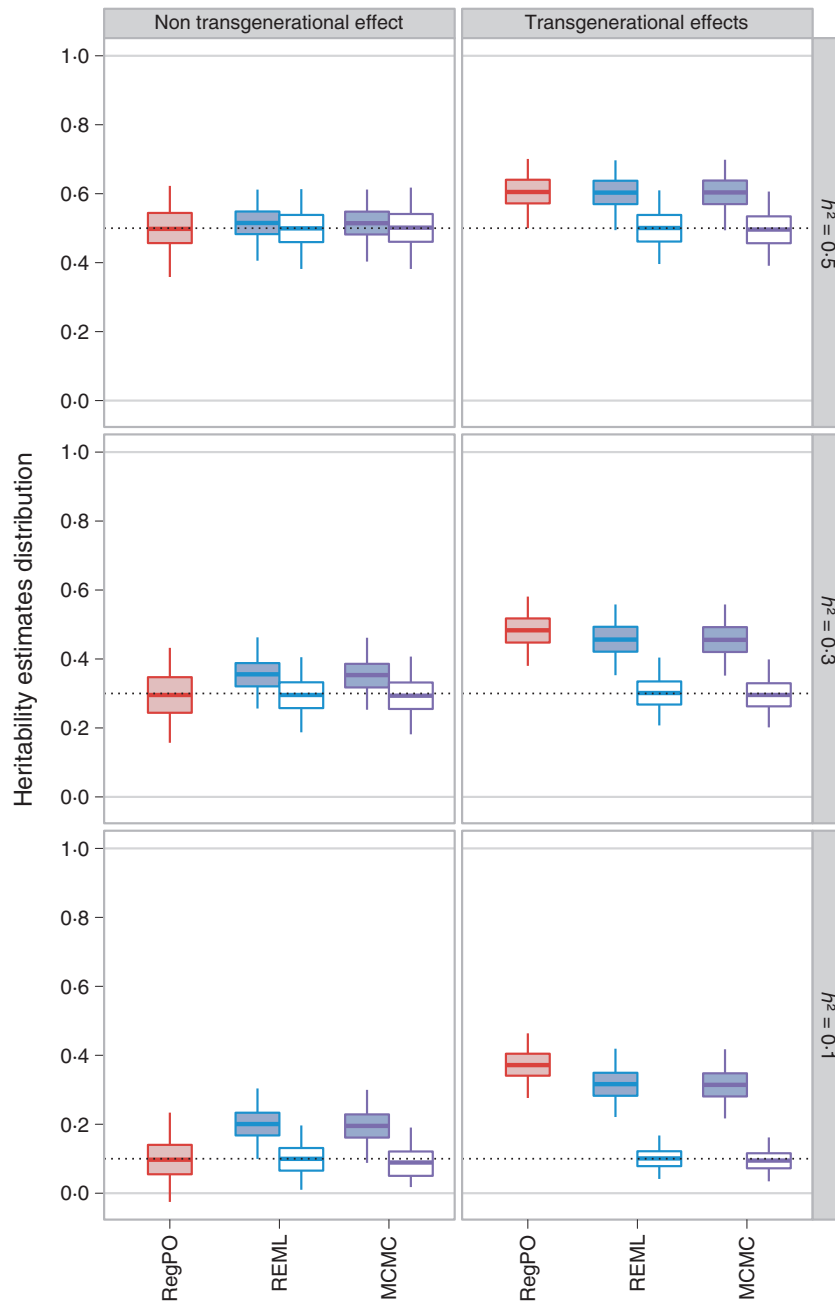
Fig. 5. Distribution of heritability estimates of a binary trait for the different approaches and estimation methods depending on the true level of heritability (h^2, dotted lines) and the environmental effect shared by individuals (non-transgenerational (left) or transgenerational (right) effect); results obtained for a fully connected pedigree and a large sample size (1 000 individuals). Boxes represent 25 and 75% quantiles, and the middle line represents the mean. Whiskers show 2.5% (bottom) and 97.5% (top) quantiles. Filled boxes: naive models (including parent–offspring regression); open boxes: informed animal models (see text).

In the case of a binary trait, the simulations showed that the animal model using MCMC was the most precise method and yields the estimates of the best 'quality' (i.e. with the smallest RMSE). The corrected REML (REMLc) method yielded valid estimates in terms of accuracy, but suffered from a low precision. The same result held for the parent–offspring regression: it also proved accurate but suffered from a disastrous precision for low quantity/quality data, in which case heritability estimates covered the entire [0,1] range and showed very large

50% interquartile intervals (up to 0.4). Admittedly, the bias and precision of the estimation methods for binary traits will depend on the incidence of the phenotype under study. Our results are based on an incidence of the phenotype of $\frac{1}{3}$. However, in many studies, the phenotypes under consideration will show an incidence closer to one or zero, in which case the estimation methods could yield biased estimates (e.g. see van Vleck 1972, for the parent–offspring regression on binary traits). Besides, as mentioned earlier, accuracy is a systematic

error leading to downward or upward biases, whereas imprecision is a stochastic source of error (i.e. a consequence of sampling stochasticity). In most cases, practitioners will analyse only one data set, that is, one particular sampling associated with the true heritability value to be estimated. The influence of inaccuracy on the estimate obtained based on a particular data set can be assessed (e.g. by referring to simulations such as ours or by performing its own preliminary simulation study by using the R package pedantics, Morrissey *et al.* 2007). Conversely, the influence of imprecision remains unknown for a particular data set. The standard error of the estimate is estimated from this one data set, but our coverage analysis shows that this estimate is not trustful when considering heritability. Therefore, we emphasize that a method with a limited inacccuracy but rather high precision, such as MCMC for binary traits, or even PQL if the expected heritability is small enough (less than 0.1), should be preferred over a more accurate but highly imprecise method, such as the REMLc. The RMSE allows summarizing both accuracy and precision to decide which estimator should be preferred. Despite the inability to assess imprecision, the strong impact of low sample size on the precision of estimates given by parent–offspring regressions seems to have been overlooked so far in the literature. Of the four studies estimating the heritability of binary traits using parent–offspring regressions in the wild, providing 13 heritability estimates in total (Hansson, Bensch & Hasselquist 2003; Charmantier, Keyser & Promislow 2007; Doligez *et al.* 2009; Charmantier *et al.* 2011), we, indeed, found five estimates based on a sample size lower than 100 parent–offspring pairs. We therefore recommend making use of a sufficiently high sample size when estimating heritability of a binary trait in the wild using the parent–offspring regression. Furthermore, the low level of precision suggests that comparing results of different regressions (for example father–offspring and mother–offspring regressions) is relevant only if estimates are computed from the exact same data set (i.e. same sample size from the same data set). This, however, means a loss of information for at least one of the regressions, because the sample to be used for all regressions should correspond to the smallest sample over all regressions (e.g. the father–offspring regression if more fathers than mothers are missing). If the regressions compared are not conducted on the same sample, the high imprecision for small sample size (say below 100 pairs) may mask high differences between estimates. Conversely, for high-quantity and quality data (i.e. large sample size and high pedigree connectivity), all approaches and methods behaved similarly, except for the animal model using PQL, which always yielded inaccurate (though precise) estimates. Yet, the PQL estimation method is not necessarily to be avoided because, for small variance components (i.e. low heritability), the animal model using PQL remains a valid estimation method (see Table 3, last column). In line with Wilson *et al.* (2011), we recommend carrying out preliminary simulations as a validation step (see Morrissey *et al.* 2007, for more details) for any particular data set before the use of PQL estimation method. Note also that Engel & Buist (1998) developed a bias correction for sire models, which might be extended to animal models with a minimum of effort.

ESTIMATING HERITABILITY IN THE PRESENCE OF SHARED ENVIRONMENTAL EFFECTS AND CONSEQUENCES ON BIAS

Our simulations showed that the parent–offspring regression is sensitive to shared environmental effects only when they are transgenerational. This result was expected but it has far from trivial implications, as most environmental effects fitted in animal models are non-transgenerational, including parental effects. Indeed, fitting an environmental maternal effect in animal models in the way recommended in the literature (Kruuk 2004; Wilson *et al.* 2010) is carried by including the identity of the mother as a random effect. Including this effect in the animal model allows one to control for additional sibling resemblance due to sharing the same mother, but not for dependency between mother and offspring's phenotypes. Hence, we do not expect this type of non-transgenerational parental effects to result in differences between heritability estimates provided by animal models and mid-parent–offspring regressions. Estimating trait heritability separately for mothers and fathers (using e.g. father–offspring and mother–offspring regressions) allows one to detect transgenerational maternal or paternal effects that are not accounted for. In order to fit transgenerational maternal effects, more general models are available, such as the Kirkpatrick & Lande model (Kirkpatrick & Lande 1989; Räsänen & Kruuk 2007; Hill & Kirkpatrick 2010; Day & Bonduriansky 2011). However, it might be very complex to handle in the context of wild population studies, because it requires a fine understanding of parental traits involved in the offspring phenotype. A simpler model (Willham model : Willham 1963, , 1972; Thompson 1976) can be used, allowing the estimation of a *genetic* maternal effect (Wilson *et al.* 2004, , 2010), that is, an effect assuming a hypothetical 'maternal performance' trait, which summarizes maternal traits potentially affecting the offspring phenotype. It also assumes that this performance is heritable. This effect is fitted by adding the 'maternal pedigree' to the animal model (i.e. an additive genetic effect based on the relatedness computed using only the matrilineage of the individuals). This kind of transgenerational maternal effects can account for any maternal trait that has an effect on the offspring's phenotype under study given that the acting maternal trait has a genetic basis (Wilson *et al.* 2010). It is necessary to point out that a poorly comprehensive (i.e. naive) animal model is likely to be more biased than a parent–offspring regression. It is, thus, highly important to cautiously define shared environment effects when using animal model. Only when properly conceived does the animal model show high advantages over parent–offspring regression, because of its higher statistical power and greater flexibility in taking possible transgenerational environmental effects into account, but also because it allows using repeated measurements and accounting for dominance effects (Lynch & Walsh 1998).

BAYESIAN INFERENCE FOR ANIMAL MODELS

Heritability estimates from animal models can be obtained using either a Frequentist (here, REML or REMLc) or a

Bayesian (here, MCMC) method. The Bayesian framework has two main advantages. First, it allows fitting a great variety of non-Gaussian data distributions. Second, the calculation of transformed estimates (like heritability) and their associated standard error is straightforward, through the use of posterior distributions that do not rely on first order approximations (Fischer, Gilmour & Werf 2004). Moreover, it is also possible to use an alternative to MCMC estimation method, the integrated nested laplace approximation (INLA). The INLA has recently been shown to provide accurate heritability estimates for Gaussian and several non-Gaussian distributions, and is faster than the MCMC (Holand *et al.* 2011, introducing the R package animalINLA). Unfortunately, the INLA performed much poorer than MCMC when estimating heritability based on binary data (Holand *et al.* 2011). Note that the pedigreemm R package (Vazquez *et al.* 2010) uses the Gaussian-Hermite quadrature for non-Gaussian distribution. Laplacian approximations (e.g. the PQL) are simple particular cases of Gaussian-Hermite quadrature, which can generate less approximate estimates given sufficient computation time and therefore are less biased than PQL (Rodriguez & Goldman 2001).

The drawbacks of Bayesian methods mostly relate to practical issues. First, concerning computation time, Frequentist methods (REML and REMLc) are far quicker (almost instantaneous) than the MCMC method, which could take up to two hours depending on sample size and model complexity (e.g. linear vs. generalized linear model) in our study. Second, the Bayesian framework in general and MCMC in particular are also less user-friendly. In addition to defining the model, one needs to set a relevant prior on variance components (see below) and appropriate MCMC checking (Monte-Carlo error, convergence, autocorrelation). Third, in terms of pedigree-handling software, REML is implemented in a wide choice of packages, including the (non free) ASReml software (Gilmour *et al.* 2006), the (free) WOMBAT software (Meyer 2007) and the (free) R package pedigreemm (Vazquez *et al.* 2010), while MCMC is only implemented in the (free) R package MCMCglmm (Hadfield 2010a).

Despite their flexibility, Bayesian methods for animal models (implemented through either MCMC or INLA) still require the specification of a prior distribution for variance components, which can be tricky especially for binary traits. Indeed, the value of the residual variance has to be fixed in that case (see Material & methods), and this happens to break the symmetry in the prior distribution for heritability, putting too much of the prior probalistic weight on 0 or 1. We showed in Appendix B that the commonly used inverse-Gamma prior is no longer a suitable option in this case, and we introduced the use of a χ^2 distribution following Gelman (2006)'s advice. However, for small sample size, the heritability estimates still suffer from a high sensitivity to this χ^2 prior distribution yielding a tedious downward bias (see Fig. 2 and Appendix B). A probably more convenient solution would be to assign directly a uniform prior on the heritability parameter (e.g. Charmantier *et al.* 2011). Unfortunately, this is currently not possible using existing pedigree-handling packages (MCMCglmm or animalINLA), unless one resorts to programming (Damgaard 2007; Waldmann, 2009; Papaïx *et al.* 2010; Authier, Cam & Guinet 2011; Charmantier *et al.* 2011).

FUTURE CHALLENGES IN HERITABILITY ESTIMATION

Whatever the estimation approach and method used, two main issues have been overlooked and remain to be addressed when estimating heritability in the wild. First, heritability estimates obtained so far relied on the infinite locus assumption, which assumes an infinite number of unlinked loci involved in the phenotype, each one having a small impact (no strong trait locus). This assumption ignores genetic constraints such as linkage and oligogenic determinism. The multilocus association model (Sillanpää 2011) allows overcoming this infinite locus assumption and considering oligogenic traits. This model takes into account more genetic effects than animal model (e.g. epistasis and linkage), but requires dense genetic markers. Given the rapid development of molecular techniques implemented to explore the genomes of various species in the wild, for example, the use of SNPs banks (Ellegren & Sheldon 2008; Backström *et al.* 2008; Bers *et al.* 2010; Slate *et al.* 2010) we have little doubt that such detailed genetic information will be available in the coming years to efficiently use the multilocus association model.

Second, contrary to captive or domestic populations, wild populations are usually characterized by imperfect individual detection. If individual capture or sighting probability is linked to the trait of interest, in particular in a transgenerational way, heritability estimates may be biased (Cam 2009; Doligez *et al.* 2012). Improving the estimation of heritability of traits in wild populations may therefore require the development of integrated capture–recapture animal model (CRAM), which would account for full pedigree information and imperfect individual detection simultaneously (O'Hara *et al.* 2008; Papaïx *et al.* 2010). Such models have already been developed (Papaïx *et al.* 2010), but user-friendly software implementation is still missing. Making a thorough list of challenges regarding heritability estimation in the wild is not the point of the present article. However, considering the work that still needs to be performed, we urge field biologists to remain cautious when drawing conclusions from heritability estimation results.

Conclusions

Our simulations showed that the animal model was the best approach to estimate heritability, using REML for Gaussian phenotypic traits and MCMC for binary traits. We join previous authors (Quinn *et al.* 2006; Postma & Charmantier 2007) in pointing out the importance of data quantity and quality: our simulations revealed the high level of imprecision for estimates given by some approaches and estimation methods for a sample size of 200 individuals, which may be considered already a large sample in studies on wild populations.

Importantly, this is especially true for binary traits. To best describe and account for the influence of shared environmental effects on heritability estimates, we advocate the systematic use of (and comparison between) mid-parent–offspring, mother–offspring and father–offspring regressions in addition to animal models to compute heritability estimates. Comparing the different estimates obtained would allow detecting overlooked non-transgenerational environmental effects, which would generate low heritability estimates for the parent–offspring regression but high estimates for the animal model. However, given the high imprecision of parent–offspring regression, this comparison is not likely to be significant. This comparison would also allow detecting sex-dependent transgenerational effects, such as maternal transgenerational effects, which would generate higher estimates for mother–offspring compared with father–offspring regressions. Therefore, we still encourage both parent–offspring and animal model estimates to be reported simultaneously. Although improvements in estimating heritability are still required, especially regarding binary traits and implementation of transgenerational parental effects, we hope that our study contributes to a better guidance and use of models and methods to estimate heritability of traits in wild populations.

Acknowledgments

References

Authier, M., Cam, E. & Guinet, C. (2011) Selection for increased body length in subantarctic fur seals on amsterdam island. *Journal of Evolutionary Biology*, **24** (3), 607–616.

Backström, N., Karaiskou, N., Leder, E.H., Gustafsson, L., Primmer, C.R., Qvarnström, A. & Ellegren, H. (2008) A gene-based genetic linkage map of the collared flycatcher (Ficedula albicollis) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics*, **179** (3), 1479–1495.

Bers, N.E.M.V., Oers, K.V., Kerstens, H.H.D., Dibbits, B.W., Crooijmans, R.P.M.A., Visser, M.E. & Groenen, M.a.M. (2010) Genome-wide SNP detection in the great tit parus major using high throughput sequencing. *Molecular Ecology*, **19**, 89–99.

Bolker, B. (2008) *Ecological Models and Data in R*. Princeton University Press, Princeton, New Jersey, USA.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.S.S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, **24**(3), 127–135.

Breslow, N.E. & Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88** (421), 9–25.

Breslow, N.E. & Lin, X. (1995) Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, **82**(1), 81–91.

Callens, M. & Croux, C. (2005) Performance of likelihood-based estimation methods for multilevel binary regression models. *Journal of Statistical Computation and Simulation*, **75**(12), 1003–1017.

Cam, E. (2009) Contribution of capture-mark-recapture modeling to studies of evolution by natural selection. *Modeling Demographic Processes In Marked Populations*, (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 83–129. Springer, Boston, Massachusetts, USA.

Charmantier, A. & Réale, D. (2005) How do misassigned paternities affect the estimation of heritability in the wild? *Molecular Ecology*, **14**(9), 2839–2850.

Charmantier, A., Keyser, A.J. & Promislow, D.E. (2007) First evidence for heritable variation in cooperative breeding behaviour. *Proceedings of the Royal Society B: Biological Sciences*, **274**(1619), 1757–1761.

Charmantier, A., Buoro, M., Gimenez, O. & Weimerskirch, H. (2011) Heritability of short-scale natal dispersal in a large-scale foraging bird, the wandering albatross. *Journal of Evolutionary Biology*, **24**, 1487–1496.

Damgaard, L.H. (2007) Technical note: how to use winbugs to draw inferences in animal models. *Journal of Animal Science*, **85**(6), 1363–1368.

Day, T. & Bonduriansky, R. (2011) A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. *The American Naturalist*, **178** (2), E18–E36.

Dempster, E.R. & Lerner, I.M. (1950) Heritability of threshold characters. *Genetics*, **35**(2), 212–236.

Doligez, B., Daniel, G., Warin, P., Pärt, T., Gustafsson, L. & Réale, D. (2012) Estimation and comparison of heritability and parent–offspring resemblance in dispersal probability from capture-recapture data using different methods: the collared flycatcher as a case study. *Journal of Ornithology*, **152**, Issue 2 supplement, 539–554.

Doligez, B., Gustafsson, L. & Pärt, T. (2009) 'Heritability' of dispersal propensity in a patchy population. *Proceedings of the Royal Society B: Biological Sciences* **276**(1668), 2829–2836.

Ellegren, H. & Sheldon, B.C. (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**(7184), 169–175.

Elston, R.C., Hill, W.G. & Smith, C. (1977) Query: Estimating "Heritability" of a dichotomous trait. *Biometrics*, **33**(1), 231–236.

Engel, B. & Buist, W. (1998) Bias reduction of approximate maximum likelihood estimates for heritability in threshold models. *Biometrics*, **54**(3), 1155–1164.

Falconer, D.S. & Mackay, T.F. (1996) *Introduction to Quantitative Genetics*. 4th edn, Benjamin Cummings, Harlow, Essex, UK.

Fischer, T.M., Gilmour, A.R. & Werf, J. (2004) Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. *Genetics Selection Evolution*, **36**(3), 363–369.

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**(3), 515–533.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004) *Bayesian Data Analysis. Text in Statisctical Science*. 2nd edn, Chapman & Hall/CRC Press, Boca Raton, Florida, USA.

Gianola, D. (1982) Theory and analysis of threshold characters. *Journal of Animal Science*, **54**(5), 1079–1096.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2006) ASReml user guide release 2.0. URL: http://www.vsni.co.uk/software/asreml/.

Gorjanc, G., Henderson, D.A., Kinghorn, B. & Percy, A. (2007) GeneticsPed: pedigree and genetic relationship functions. R package version 1.13.0. URL: http://www.bioconductor.org/packages/2.0/bioc/html/GeneticsPed.html.

Hadfield, J.D. (2010a) MCMC methods for multi-response generalised linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, **33** (2), 1-22.

Hadfield, J.D. (2010b) MCMCglmm course notes. URL: http://cran.r-project.org/web/packages/MCMCglmm/index.html.

Hadfield, J.D., Burgess, M.D., Lord, A., Phillimore, A.B., Clegg, S.M. & Owens, I.P.F. (2006) Direct versus indirect sexual selection: Genetic basis of colour, size and recruitment in a wild bird. *Proceedings of the Royal Society B: Biological Sciences*, **273**(1592), 1347–1353.

Hansson, B., Bensch, S. & Hasselquist, D. (2003) Heritability of dispersal in the great reed warbler. *Ecology Letters*, **6**(4), 290–294.

Henderson, C.R. (1950) Estimation of genetic parameters. *Annals of Mathematical Statistics*, **21**, 309–310.

Henderson, C.R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, **32**(1), 69–83.

Hill, W.G. & Kirkpatrick, M. (2010) What animal breeding has taught us about evolution. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 1–19.

Holand, A.M., Steinsland, I., Martino, S. & Jensen, H. (2011) Animal models and integrated nested laplace approximations. *Statistics*. Preprints (4/2011).

Kirkpatrick, M. & Lande, R. (1989) The evolution of maternal characters. *Evolution*, **43**(3), 485–503.

Knott, S.A., Sibly, R.M., Smith, R.H. & Moller, H. (1995) Maximum likelihood estimation of genetic parameters in life-history studies using the 'Animal model'. *Functional Ecology*, **9**(1), 122–126.

Kruuk, L.E., Merilä, J. & Sheldon, B.C. (2001) Phenotypic selection on a heritable size trait revisited. *The American Naturalist*, **158**(6), 557–571.

Kruuk, L.E.B. (2004) Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **359**(1446), 873–890.

Kruuk, L.E.B. & Hadfield, J.D. (2007) How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology*, **20**(5), 1890–1903.

Kruuk, L.E., Slate, J. & Wilson, A.J. (2008) New answers for old questions: The evolutionary quantitative genetics of wild animal populations. *Annual Review of Ecology, Evolution, and Systematics*, **39**(1), 525–548.

Kruuk, L.E.B., Slate, J., Pemberton, J.M., Brotherstone, S., Guinness, F. & Clutton-Brock, T. (2002) Antler size in red deer: Heritability and selection but no evolution. *Evolution*, **56**(8), 1683–1695.

Lane, J.E., Kruuk, L.E.B., Charmantier, A., Murie, J.O., Coltman, D.W., Buoro, M., Raveh, S. & Dobson, F.S. (2011) A quantitative genetic analysis of hibernation emergence date in a wild population of columbian ground squirrels. *Journal of Evolutionary Biology*, **24**(9), 1949–1959.

Lynch, M. & Walsh, B. (1998) *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland, Massachussets, USA.

MacColl, A.D.C. & Hatchwell, B.J. (2003) Heritability of parental effort in a passerine bird. *Evolution*, **57**(9), 2191–2195.

McCleery, R.H., Pettifor, R.A., Armbruster, P., Meyer, K., Sheldon, B.C. & Perrins, C.M. (2004) Components of variance underlying fitness in a natural population of the great tit parus major. *The American Naturalist*, **164**(3), E62–E72.

Merilä, J., Kruuk, L.E.B. & Sheldon, B.C. (2001) Natural selection on the genetical component of variance in body condition in a wild bird population. *Journal of Evolutionary Biology*, **14**(6), 918–929.

Meyer, K. (2007) WOMBAT-A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University. Science. B*, **8**(11), 815–821.

Milner, J.M., Pemberton, J.M., Brotherstone, S. & Albon, S.D., (2000) Estimating variance components and heritabilities in the wild: a case study using the 'animal model' approach. *Journal of Evolutionary Biology*, **13**(5), 804–813.

Morales, J., Kim, S., Lobato, E., Merino, S., Tomás, G., Martínez de la Puente, J. & Moreno, J. (2010) On the heritability of blue-green eggshell colouration. *Journal of Evolutionary Biology*, **23**(8), 1783–1791.

Morrissey, M.B. & Wilson, A.J. (2009) Pedantics: an R package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Molecular Ecology Resources*, **10**(4), 711–719.

Morrissey, M.B., Wilson, A.J., Pemberton, J.M. & Ferguson, M.M. (2007) A framework for power and sensitivity analyses for quantitative genetic studies of natural populations, and case studies in soay sheep (Ovis aries). *Journal of Evolutionary Biology*, **20**(6), 2309–2321.

Mousseau, T.A. & Roff, D.A. (1987) Natural selection and the heritability of fitness components. *Heredity*, **59**(2), 181–197.

Nakagawa, S. & Schielzeth, H. (2010) Repeatability for gaussian and non gaussian data: a practical guide for biologists. *Biological Reviews*, **85**(4), 935–956.

Nilsson, J.A., Åkesson, M. Nilsson, J.F. (2009) Heritability of resting metabolic rate in a wild population of blue tits. *Journal of Evolutionary Biology*, **22**(9), 1867–1874.

O'Hara, R.B., Cano, J.M., Ovaskainen, O., Teplitsky, C. & Alho, J.S. (2008) Bayesian approaches in evolutionary quantitative genetics. *Journal of Evolutionary Biology*, **21**(4), 949–957.

Papaïx, J., Cubaynes, S., Buoro, M., Charmantier, A., Perret, P. & Gimenez, O. (2010) Combining capture-recapture data and pedigree information to assess heritability of demographic parameters in the wild. *Journal of Evolutionary Biology*, **23**(10), 2176–2184.

Patterson, H.D. & Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**(3), 545–554.

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News*, **6**(1), 7–11.

Postma, E. & Charmantier, A. (2007) What 'animal models' can and cannot tell ornithologists about the genetics of wild populations. *Journal of Ornithology*, **148**(S2), 633–642.

Quinn, J.L., Charmantier, A., Garant, D. & Sheldon, B.C. (2006) Data depth, data completeness, and their influence on quantitative genetic estimation in two contrasting bird populations. *Journal of Evolutionary Biology*, **19**(3), 994–1002.

R Development Core Team (2011) R: a language and environment for statistical computing. URL: http://www.R-project.org/.

Reid, J.M., Arcese, P., Sardell, R.J. & Keller, L.F. (2011a) Additive genetic variance, heritability and inbreeding depression in male extra-pair reproductive success. *The American Naturalist*, **177**, 177–187.

Reid, J.M., Arcese, P., Sardell, R.J. & Keller, L.F. (2011b) Heritability of female extra-pair paternity rate in song sparrows (Melospiza melodia). *Proceedings of the Royal Society B: Biological Sciences*, **278**(1708), 1114–1120.

Ridley, M. (2003) *Evolution. 3rd edn*. Wiley-Blackwell, Hoboken, New-Jersey, USA.

Rodriguez, G. & Goldman, N. (2001) Improved estimation procedures for multilevel models with binary response: a case-study. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **164**(2), 339–355.

Roff, D.A. (1997) *Evolutionary quantitative genetics*. Chapman & Hall, New York, USA.

Roff, D.A. (2007) A centennial celebration for quantitative genetics. *Evolution*, **61**(5), 1017–1032.

Rossiter, M. (1996) Incidence and consequences of inherited environmental effects. *Annual Review of Ecology and Systematics*, **27**, 451–476.

Räsänen, K. & Kruuk, L.E.B. (2007) Maternal effects and evolution at ecological time-scales. *Functional Ecology*, **21**(3), 408–421.

Réale, D., Festa-Bianchet, M. & Jorgenson, J.T. (1999) Heritability of body mass varies with age and season in wild bighorn sheep. *Heredity*, **83**(5), 526–532.

Sheldon, B.C., Kruuk, L.E.B. & Merilä, J. (2003) Natural selection and inheritance of breeding time and clutch size in the collared flycatcher. *Evolution*, **57**(2), 406–420.

Sillanpää, M.J. (2011) On statistical methods for estimating heritability in wild populations. *Molecular Ecology*, **20**(7), 1324–1332.

Slate, J., Santure, A.W., Feulner, P.G., Brown, E.A., Ball, A.D., Johnston, S.E. & Gratten, J. (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics*, **26**(6), 275–284.

Sorensen, D. & Gianola, D. (2002) *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. Statistics for Biology and Health, Springer, New York, USA.

Sorensen, D.A. & Kennedy, B.W. (1984) Estimation of genetic variances from unselected and selected populations. *Journal of Animal Science*, **59**(5), 1213-1223.

Thompson, R. (1976) The estimation of maternal genetic variances. *Biometrics*, **32**(4), 903–917.

Thériault, V., Garant, D., Bernatchez, L. & Dodson, J.J. (2007) Heritability of life-history tactics and genetic correlation with body size in a natural population of brook charr (Salvelinus fontinalis). *Journal of Evolutionary Biology*, **20**(6), 2266–2277.

Vazquez, A.I., Bates, D.M., Rosa, G.J.M., Gianola, D. & Weigel, K.A. (2010) Technical note: an r package for fitting generalized linear mixed models in animal breeding1. *Journal of Animal Science*, **88**(2), 497–504.

Visscher, P.M., Hill, W.G. & Wray, N.R. (2008) Heritability in the genomics era-concepts and misconceptions. *Nature Reviews Genetics*, **9**(4), 255–266.

van Vleck, L. (1972) Estimation of heritability of threshold characters. *Journal of Dairy Science*, **55**(2), 218–225.

Waldmann, P. (2009) Easy and flexible Bayesian inference of quantitative genetic parameters. *Evolution*, **63**(6), 1640–1643.

van der Werf, J.H. & de Boer, I.J. (1990) Estimation of additive genetic variance when base populations are selected. *Journal of Animal Science*, **68**(10), 3124–3132.

Willham, R.L. (1963) The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics*, **19**(1), 18.

Willham, R.L. (1972) The role of maternal effects in animal breeding: III. Biometrical aspects of maternal effects in animals. *Journal of Animal Science*, **35**(6), 1288–1293.

Wilson, A.J., Coltman, D.W., Pemberton, J.M., Overall, A.D.J., Byrne, K.A. & Kruuk, L.E.B. (2004) Maternal genetic effects set the potential for evolution in a free-living vertebrate population. *Journal of Evolutionary Biology*, **18**(2), 405–414.

Wilson, A.J., Morrissey, M.B., Adams, M.J., Walling, C.A., Guinness, F.E., Pemberton, J.M., Clutton-Brock, T.H. & Kruuk, L.E.B. (2011) Indirect genetics effects and evolutionary constraint: an analysis of social dominance in red deer, cervus elaphus. *Journal of Evolutionary Biology*, **24**(4), 772–783.

Wilson, A.J., Réale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A., Kruuk, L.E.B. & Nussey, D.H. (2010) An ecologist's guide to the animal model. *Journal of Animal Ecology*, **79**(1), 13–26.

Wright, S. (1934) An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics*, **19**(6), 506–536.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009) *Mixed effects models and extensions in ecology with R*. Statistics for Biology and Health, Springer, New York, USA.

Åkesson, M., Bensch, S., Hasselquist, D., Tarka, M. & Hansson, B. (2008) Estimating heritabilities and genetic correlations: comparing the 'Animal model'

with parent–offspring regression using data from a natural population. *PLoS ONE*, **3**(3), e1739.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix A. Function for pedigree simulation

Appendix B. Analysis of prior distributions for the estimation of heritability of binary traits

Appendix C. Results on the additive genetic variance V_A

Appendix D. Comprehensive results of simulations including a non-transgenerational or transgenerational shared environmental effect