SPECIAL 'HOW TO...' PAPER

An ecologist's guide to the animal model

Alastair J. Wilson[1]*, Denis Réale[2], Michelle N. Clements[1], Michael M. Morrissey[1], Erik Postma[3], Craig A. Walling[1], Loeske E. B. Kruuk[1] and Daniel H. Nussey[1]

[1]*Wild Evolution Group, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3JT, UK;* [2]*Département des Sciences Biologiques Université du Québec à Montréal, Montréal, QC, Canada; and* [3]*Zoologisches Museum, Universität Zürich-Irchel, Zurich, Switzerland*

Summary

1. Efforts to understand the links between evolutionary and ecological dynamics hinge on our ability to measure and understand how genes influence phenotypes, fitness and population dynamics. Quantitative genetics provides a range of theoretical and empirical tools with which to achieve this when the relatedness between individuals within a population is known.

2. A number of recent studies have used a type of mixed-effects model, known as the animal model, to estimate the genetic component of phenotypic variation using data collected in the field. Here, we provide a practical guide for ecologists interested in exploring the potential to apply this quantitative genetic method in their research.

3. We begin by outlining, in simple terms, key concepts in quantitative genetics and how an animal model estimates relevant quantitative genetic parameters, such as heritabilities or genetic correlations.

4. We then provide three detailed example tutorials, for implementation in a variety of software packages, for some basic applications of the animal model. We discuss several important statistical issues relating to best practice when fitting different kinds of mixed models.

5. We conclude by briefly summarizing more complex applications of the animal model, and by highlighting key pitfalls and dangers for the researcher wanting to begin using quantitative genetic tools to address ecological and evolutionary questions.

Key-words: animal model, genetic correlation, heritability, mixed-effect model, natural selection, pedigree, quantitative genetics

Introduction

The role of natural selection and micro-evolution in the ecological dynamics of naturally occurring populations has become the focus of an increasing number of studies (Hairston *et al.* 2005; Saccheri & Hanski 2006; Carroll *et al.* 2007; Pelletier *et al.* 2007). However, we can say nothing about evolutionary processes without a means of measuring and understanding the way that genetics underpins variation in demographic rates and fitness (Ellegren & Sheldon 2008; Kruuk, Slate & Wilson 2008). Quantitative genetics, a discipline with a long and distinguished history within evolutionary biology and animal breeding, provides a potentially powerful means for estimating the genetic architecture and predicting the evolutionary potential of phenotypic traits (Falconer & Mackay 1996; Lynch & Walsh 1998). The recent application of quantitative genetic methodology to long-term

field studies of vertebrate populations has yielded new insight into the complexities of and constraints on evolutionary dynamics under realistic ecological conditions (Kruuk 2004; Ellegren & Sheldon 2008; Kruuk *et al.* 2008). These studies have typically used a form of mixed-effects models known as the 'animal model' to decompose phenotypic variance into different genetic and environmental sources and to estimate key parameters such as the heritability of a trait or the genetic correlations between traits (e.g. Réale, Festa-Bianchet & Jorgenson 1999; Kruuk *et al.* 2000; Milner *et al.* 2000; Kruuk, Merila & Sheldon 2001; Garant *et al.* 2004; Wilson *et al.* 2005; Gienapp, Postma & Visser 2006). This approach, as with any in the field of quantitative genetics, requires knowledge of the relatedness of individuals in a population. Such information, although challenging to come by in field populations, is increasingly available for studies of a range of taxa (Pemberton 2008), fuelling a growing interest in the application of quantitative genetics to studies of natural, rather than laboratory or domestic, populations. Animal models are not

*Correspondence author. E-mail: alastair.wilson@ed.ac.uk

difficult to implement, given appropriate data, but correctly specifying and interpreting them is a potentially fraught business.

In this paper, we present a practical guide aimed at the ecologist wishing to use the animal model for the first time. Our aim is neither to provide a comprehensive treatment of the theoretical and statistical models used in quantitative genetics, nor to review the empirical results of their application to ecological data sets. Rather our goal is to provide a practical guide for ecologists interested in exploring the potential to apply quantitative genetic methods to their research. In what follows we briefly lay out some of the key concepts involved. We describe the parameters that quantitative genetic methods attempt to estimate, why these parameters are of interest to ecologists, and how we can use statistical models – particularly the animal model – to estimate them. We also provide code (and example data sets) to run models with several common software applications. Whilst we have consequently tried to avoid technical terminology and issues as far as possible, in a mathematical and statistical subject like quantitative genetics technical details do assume vital importance. We have therefore tried to highlight some of the most likely pitfalls to be wary of, whilst referring the reader to the original literature and, where appropriate, more detailed reviews on specific topics. Thus this paper is intended to serve as a useful starting point and a way into the literature for the uninitiated. We do not wish this paper to be treated as a replacement for or an excuse to skip over the original quantitative genetics literature and we would always advise against a 'black box' approach when using complex statistical models to analyse data. However, we recognize that for many empiricists, ourselves included, grappling with the theory and mathematics underlying a technique is much more rewarding given a clear sense of the end goal. We very much hope this paper will provide a useful clarification of both the ultimate goals and the key considerations and pitfalls for any field ecologist interested in applying quantitative genetic analyses to their data.

The basics

The premise of classical quantitative genetics is that, given knowledge about the relationships among individuals in a population and data on phenotypic traits, we can make useful inferences about the inheritance and evolutionary potential of those traits without explicit knowledge of the genetic loci involved. If individuals that are closely related (and therefore share lots of genes) are phenotypically more similar to one another than individuals that are unrelated (and therefore share fewer genes), we can infer that genes make an important contribution to phenotypic variance. For most ecologically relevant traits with continuous or discrete distribution quantitative geneticists assume that phenotypic differences observed among individuals are related to differences in a large number of genes, each of them having a minor effect (the so called infinitesimal model; Falconer & Mackay 1996; Lynch & Walsh 1998).

For a single trait, we can estimate the amount of phenotypic variance (V_P) that is due to genetic differences among individuals (V_G) (Falconer & Mackay 1996). Genotypic differences among individuals are composed of additive (V_A), dominance (V_D) and interaction or epistatic (V_I) genetic sources of variance. However, V_D and V_I are extremely difficult to estimate in non-experimental settings and both animal breeders and field ecologists have tended to focus on measuring additive genetic variance by estimating the phenotypic similarity of relatives (Falconer & Mackay 1996; Kruuk 2004). In the simplest case, this involves statistically partitioning the phenotypic variance into two parts such that $V_P = V_A + V_R$ where V_R is the residual variance. V_R is normally interpreted as arising from environmental effects which entails the assumption that dominance and epistasis make negligible contributions to V_P.

The narrow-sense heritability of a trait (h^2) is then defined as the proportion of phenotypic variance explained by additive genetic variance (i.e. V_A/V_P) and describes the degree of resemblance between relatives. This idea of partitioning variance extends to multiple traits. Thus, for a pair of covarying traits, we can ask how much of the phenotypic covariance (COV_P) is due to additive genetic effects (COV_A). Genetic covariance between traits is expected to arise through linkage or pleiotropy (a single locus influencing multiple traits) and is often expressed as a genetic correlation (r_G).

The goal of any empirical study will be to estimate these parameters to answer biological questions, and this can be done using various statistical methods. Different approaches differ with respect to their implicit assumptions and the type of data required, but the common principle underlying all methods is that of comparing phenotypic similarity among individuals of known relationship to one another to quantify the (additive) genetic basis of trait (co)variance. Probably the most familiar technique is parent–offspring regression in which, for example, a trait's heritability can be estimated as the slope of the regression of offspring phenotype on mid-parent phenotype (Falconer & Mackay 1996). This method has often been used in wild systems, particularly in studies of passerine birds (e.g. Perrins & Jones 1974; Van Noordwijk, van Balen & Scharloo 1981; Flux & Flux 1982; Gustafsson 1986). The application of ANOVA-based analyses of full-sib and half-sib families to field data has been relatively limited (compared to experimental studies), due to a general requirement for controlled and balanced breeding designs (Falconer & Mackay 1996; Lynch & Walsh 1998); its most frequent use has been in analyses of juvenile traits in avian studies, sometimes combined with cross-fostering (e.g. van Noordwijk 1984; Smith & Wettermark 1995; Merila & Sheldon 2001).

The technique that has come to be known as the 'animal model' has a long history of development and use within the animal breeding and statistical genetics literature (Henderson 1953, 1976, 1984; Meyer 1985; Thompson 2008). Evolutionary ecologists were generally slower than animal breeders to recognize the potential applications of this and related techniques (but see Shaw 1987). However, over the last decade this method has both superseded the alternatives for, and

facilitated a great upsurge of interest in, estimating quantitative genetic parameters in natural populations (for reviews see Kruuk 2004; Kruuk *et al.* 2008; Postma & Charmantier 2007). Perhaps the primary reason for this is that the animal model makes use of information from all types of relationship within the complex, unbalanced pedigrees we expect to find in natural populations. Additionally, it is flexible enough to cope with variable but non-trivial amounts of missing data (e.g. unknown paternities, unmeasured phenotypes), although missing data will obviously reduce estimate precision and in some circumstances can cause bias. Furthermore, while the primary goal may be to estimate genetic parameters, any known or hypothesized non-genetic influences on phenotype (e.g. effects of age, sex, cohort, territory) can also be explored within an animal model. This is useful for two reasons: first, if not modelled, these effects may bias estimates of the genetic parameters; and secondly, understanding the environmental influences on phenotype is an important part of most ecological studies.

How does the animal model work?

An animal model is a particular form of model in which an individual's 'breeding value' (or 'genetic merit') is included as an explanatory variable for a phenotypic trait of interest. The breeding value is simply the additive effect of an individual's genotype on the trait expressed relative to the population mean phenotype. This means that, in the simplest case, we might model a single trait (y) in an individual (i) as:

$$y_i = \mu + a_i + e_i \qquad \text{(model 1)}$$

where μ is the population mean, a_i is the breeding value (i.e. effects of i's genotype relative to μ) and e_i is a residual term. This type of model will be intuitive to anyone familiar with regression and linear models. However, there is a catch here in that, as we do not actually know what each individual's breeding value is, we cannot fit the model to see whether this term is significant and how much variance it explains. The solution lies in specifying model 1 as a linear mixed effects model – a type of model that contains both fixed and random effects (Galwey 2006) – in which the breeding value is treated as a random effect.

Ecologists often use random effects to account for sources of non-independence among data points or to avoid pseudoreplication (Milner, Elston & Albon 1999; Bolker *et al.* 2009; van de Pol & Wright 2009). For instance, if a researcher is testing the effect of a variable (x) on a trait (y) but some individuals in the population have been measured on multiple occasions, they may chose to discard all but one record per individual, or take an average, or fit identity as a random effect to avoid getting an erroneous picture of the significance of x on y. However, random terms also allow us to make inferences about the distribution of effects in a wider population. This is because including identity as a random effect also yields an estimate of the among-individual variance for y in the population. In this example, additional random effects could be fitted if other sources of non-independence between

data points were suspected (e.g. habitat patch, year of birth, mother), and for each additional random effect a corresponding component of the total phenotypic variance would be estimated. In other words, the use of mixed effects models allows us to partition variance. An animal model then is simply a linear mixed effect model in which we treat the breeding value as a random effect. The idea behind this is identical to the rationale for any other mixed model – data points are non-independent (because individuals measured share genes) and we want to estimate the amount of variance explained by a source of non-independence (genes) in the population. By fitting breeding value as a random effect, we obtain an estimate of the variance in breeding values which is defined as the additive genetic variance V_A. In addition, variation from numerous other environmental and indirect genetic sources can be estimated using a mixed model approach, often simultaneously if the right pedigree and phenotypic data is available. In Table 1, we list the main sources of variance that can be estimated in a mixed model analysis, highlighting the key considerations and sources of bias associated with each.

The interested reader should most definitely consult more detailed treatments of the animal model (Lynch & Walsh 1998; Kruuk 2004) as well as the primary work leading to its development (Henderson 1976; Meyer 1985; Shaw 1987). Our deliberate avoidance of an algebraic presentation limits the depth of the current description and hides the mathematical complexity, but the key point to grasp is that population pedigree data gives us an expectation of the way breeding values should covary among individuals, and this in turn allows us to solve for genetic parameters including V_A and, in the case of multivariate models, COV_A. For any pair of individuals i and j, the expected additive genetic covariance between them is equal to $2\theta_{ij}V_A$ where θ_{ij} – a parameter normally called the coefficient of coancestry – is the probability that an allele drawn at random from individual i is identical by descent to one drawn at random from individual j. Doubling the coefficient of coancestry yields the more familiar values of 'relatedness' (i.e. 0·5 for parent–offspring pairs and full-siblings, 0·25 for half siblings, 0·125 for first cousins etc). The higher the relatedness, and the more V_A underlying the trait, the greater the expected covariance between two individuals. Among all the n individuals in a pedigree, the matrix of additive genetic covariance for a trait is given as $\mathbf{A}V_A$ where $\mathbf{A}$ is the additive genetic relationship matrix. This is an $n \times n$ matrix that contains all the pairwise values of relatedness. Figure 1 shows how a simple pedigree structure can be represented in the form of an $\mathbf{A}$ matrix. Note that the matrix is symmetrical about the diagonal (since $2\theta_{ij} = 2\theta_{ji}$) with values of 1 on the diagonal (i.e. $2\theta_{ii} = 1$ as individuals are perfectly related to themselves), assuming no inbreeding. If an individual is inbred, the diagonal elements for this individual will be equal to $1 + F$ (the coefficient of inbreeding). Here we show a pedigree containing only 14 individuals (Fig. 1a) and it is clear that working out the corresponding $\mathbf{A}$ matrix (Fig. 1b) for a sample of hundreds or thousands would be a tedious and difficult task. Fortunately it is unnecessary for the researcher to do so because all the information contained

Table 1. Table showing the key variance components that an evolutionary biologist might wish to estimate using an 'animal model'

Component/Name	Data required	May also include	Notes
V_I / Individual	At least two records per individual	V_A, V_{PE}, V_M, V_C, V_D	V_I differs from V_{PE} because it includes V_A. The ratio of V_I o V_P is referred to 'repeatability' and models between-individual differences caused by unspecified genetic and non-genetic factors
V_A/Additive genetic	Pedigree information (half-sib/full sib structure)	V_{CE}, V_{PE}, V_M, V_C, V_D	See main text and supporting tutorials
V_{CE}/Common environment	Pedigree information + mothers with at least two offspring	V_{PE}, V_M, V_C, V_D	V_{CE} generally corresponds to environmental effects shared by the members of a family (e.g. nest effects) that affect each individual permanently
V_{PE}/Permanent environment	Pedigree information + at least two records per individual	V_{CE}, V_M, V_C, V_D and maternal effects specific to a particular individual and not considered in V_M, V_C	See main text and supporting tutorials
V_M/Maternal environment	Pedigree information (half-sib/full sib structure) + mothers with at least two offspring	V_{PE} if not specified	An equivalent effect could be fitted for paternal effects. V_M /V_C differs from V_{CE} in cases where mothers produce different clutches (otherwise the information of the two matrices is redundant) Information from cross-fostering experiments can improve the partitioning of V_A, V_M and V_C
V_C/Maternal genetic	Pedigree information (half-sib/full sib structure) over at least three generations	V_A (if $COV_{A,C}$ different from zero and not specified)	An equivalent effect could be fitted for paternal effects. V_M/V_C differ from V_{CE} in cases where mothers produce different clutches (otherwise the information of the two matrices is redundant) Information from cross-fostering experiments can improve the partitioning of V_A, V_M and V_C Differs from V_M because it represents any environment effects of maternal origin with a genetic basis (i.e. a mother inherits some additive genetic effects affecting her ability to raise her offspring) In principle a model including V_A and V_C can also allow direct genetic (i.e. V_A) and maternal genetic effects (i.e. V_C) to covary (i.e. genes affecting directly the phenotype of an individual also affect, or are linked, to genes affecting its phenotype indirectly though the performance of its mother)
V_D/Dominance	Pedigree information (half-sib/full sib structure) over several generations + matrix of dominance	V_{PE} if common environmental effects are correlated with dominance effects	Estimating dominance variance in a mixed model framework requires large amounts of data, a near complete pedigree and lots of phenotypic information on full- and half-sibs Variation in inbreeding coefficients within the population is also required so that the dominance matrix is not redundant with the additive genetic or the common environment matrix

The data required for partitioning the phenotypic variance (V_P) in each case is listed along with other sources of variance that may be included/confounded with each estimated component if these other sources are not specified in the model. Unspecified variance components can appear in other components if: (i) they are nested in another variance component (i.e. V_A and V_{PE} are nested within V_I); (ii) redundancy between the information of two matrices can be a source of bias in variance partitioning (i.e. V_A can be upward biased if V_{PE} is unspecified; see text).

in **A** can be expressed, and fed to the software being used, in a much simpler format, requiring only that the parents of each individual be specified (Fig. 1c). Note that if no parentage information is known for an individual, a situation necessarily true for members of the first generation, the default assumption is that that individual is unrelated to all others.

As with most linear models, we typically assume that the residual terms (e_i) are normally distributed with a mean of zero and a variance to be estimated (V_R). Of course departures from normality often occur and major violations of this assumption should always be acknowledged and preferably dealt with (e.g. by appropriate data transformations, or use of generalized models discussed further below). We also assume that residual terms are uncorrelated among individuals. In the simplest case of a model containing only the fixed effect of trait mean and a random effect of breeding
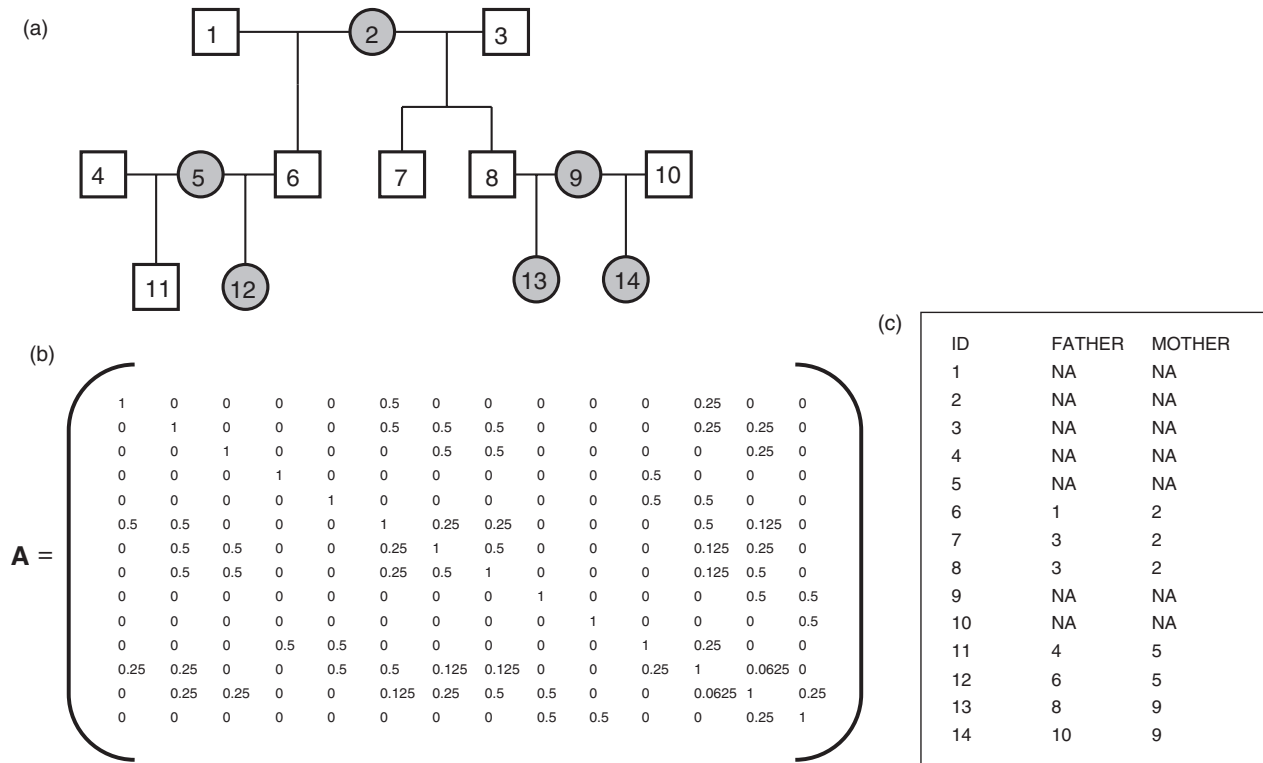
Fig. 1. (a): An example pedigree with males as squares and females as circles, with illustrations of (b) a relatedness matrix, typically denoted 'a', derived from this pedigree and (c) pedigree data structure as typically required by software packages corresponding to this pedigree.

value (i.e. model 1 above), this means that any covariance among observations must arise from sharing of genes as determined by the pedigree structure. In reality of course there are likely to be other sources of phenotypic similarity among records. These may include intrinsic variables (e.g. sex), and extrinsic variables (e.g. climate, population density, prey abundance). If we know, or hypothesize, that such effects are important then it might be sensible to expand this model accordingly. For instance, if the trait was body size and observations were made on individuals of varying ages in a sexually dimorphic organism we might prefer a model something like:

$$y_i = \mu + \text{sex}_i + \text{age}_i + a_i + e_i \qquad \text{(model 2)}$$

in which sex and age are included as additional fixed effects. Model building or reduction can proceed according to normal practice, with the researcher often choosing to add or remove explanatory terms based on effect size and/or statistical significance.

If additional explanatory covariates or factors are not associated with the pedigree structure then their inclusion should not systematically change the estimate of V_A. That is to say violating the assumption of uncorrelated residuals does not necessarily induce bias. However, a particular concern is that relatives are often clustered in time and/or space so that they tend to share environmental effects more often than unrelated individuals. For example, nest effects arising from patch quality or parental provisioning will often be shared among siblings within a clutch, while maternal effects may also cause within-family similarity (Table 1; Kruuk & Hadfield 2007; Wilson *et al.* 2005). These 'common environment' effects are confounded with the pedigree structure and, if not controlled for, cause upward bias in estimates of genetic parameters. On the contrary, failure to control for some fixed effects (e.g. specific environmental conditions or age) can cause upward bias in residual variance when related individuals are measured in different environmental conditions or at different ages (i.e. sampling bias). Consequently, additional random and/or fixed effects are often fitted specifically to try and disentangle genetic from common environment effects (Kruuk & Hadfield 2007). We illustrate this in the worked examples to follow.

Considerations before you start

There are two key steps to ensuring success that should be undertaken prior to embarking on a quantitative genetic analysis of field data. The first is to know what your biological hypothesis is and to think carefully about how to formulate a statistical model to test it. The second is to recognize that empirical quantitative genetics is data-hungry: statistical power will always limit what is possible, but it will also limit what is sensible. While seemingly obvious, the second point is overlooked with surprising frequency and unfamiliarity with quantitative genetics can sometimes lead to erroneous expectations about what is or isn't possible. More generally, clear

biological questions are needed to formulate a sensible modelling strategy with which to test statistical hypotheses. For instance, if clutch size is under selection but you want to know whether it can evolve, then the question is what is h^2 for clutch size? If you suspect the evolution of larger clutches is constrained by a life-history trade-off with egg size then you should test for a strong negative genetic correlation between these traits. Of course more sophisticated questions and hypotheses are possible as well but the key is to know what they are before you start modelling.

These considerations should be followed by a realistic appraisal of the data. No advanced statistical techniques can compensate for inadequate data (a point that should actually reassure field ecologists everywhere) and obtaining accurate and precise estimates of genetic parameters requires a lot of data. Even the largest data sets compiled from ecological studies are small by comparison to many available to animal breeders, and the power of a quantitative genetic analysis also depends crucially on the pedigree structure. In other words, one can do very little with an enormous data set in which few individuals are related, but fewer individuals in a well-connected pedigree can be very informative. Obviously the ideal situation will always be a large number of individuals in a well-connected pedigree!

From a pragmatic point of view how much data is enough, and how do we know if a pedigree structure is suitable? Unfortunately rules of thumb are hard to come by: a useful estimate of heritability can certainly be obtained from a hundred records or less in some circumstances, but not in all. If hypotheses are to be tested using genetic correlations then an order of magnitude more data will usually be required for a similar level of statistical power. If the data are already in hand then the simplest way forward is probably to run some models to get an idea of the power (as indicated by standard errors or confidence interval around your heritability estimates). Given an estimate of h^2 ($\pm$ SE) of 0·5 $\pm$ 0·24 we might be able to conclude that a trait is significantly heritable, but we should also recognize that a 95% confidence interval of $0\cdot02 \le h^2 \le 0\cdot98$ does not convey a lot more information that the 100% confidence interval which is necessarily $0 \le h^2 \le 1$.

If a project is in the planning stages then simulations offer a useful way to assess how much data will need to be collected. Tools are available for this purpose (see for example, Morrissey *et al.* 2007) although simulations will always require assumptions to be made about the sort of pedigree structures likely to be obtained. A point to reiterate is that sampling strategies should be designed with the specific aim of making sure that pedigree information can be obtained (whether through behavioural observation or molecular pedigree analysis) and that close relatives are sampled, where possible across different environmental conditions,. Achieving this aim will range from the perfectly feasible (e.g. studying early growth traits in nestling passerine birds) to the virtually impossible (e.g. studying adult traits in dispersive marine fishes) depending on the system under study.

While we have emphasized the importance of data quantity, there may also be issues of data quality. Certainly we acknowledge that pedigree data from wild populations is unlikely to be perfect and that errors in the pedigree will bias quantitative genetic parameters. If extra-pair paternity (EPP) or clutch parasitism occurs then relationships determined by observation can be incorrect, while methods of molecular pedigree analysis can never give perfect results unless there is complete sampling with either no genotyping error or a very large number of variable loci. There are at least two sources of comfort though. First, the effects of pedigree error are, at least for simple scenarios, predictable. Both the failure to recognize a true relationship and erroneously assigning a relationship between unrelated individuals will result in downward bias of genetic variance and hence heritability. Secondly, simulation studies suggest that the level of bias for realistic levels of EPP or paternity assignment error will often be low (Charmantier & Reale 2005). As a partial caveat to these findings, bias may be less predictable for more complex models (e.g. those including maternal effects) and could be in either direction for genetic correlations (i.e. there will be downward bias in the genetic covariance but also in the genetic variances of each trait; Morrissey *et al.* 2007). Consequently it should not be generally concluded that pedigree error can be ignored, and tailored-simulations to explore the impact on particular analyses are very useful. Software is available for this purpose (Morrissey *et al.* 2007).

Available software

Having thought though the hypotheses to be tested and collected the required data, the last requirement is to obtain suitable software. As outlined above, the animal model is simply a special case of a linear mixed effects model, but unfortunately not all generic statistical software packages are able to fit the random effect structure associated with the pedigree information. Nevertheless, there are a number of possible options (Table 2) that differ in cost, capability, and method of inference. All should be equally suitable for basic models of the type discussed here, while some offer greater flexibility to fit more complex models (e.g. incorporating spatial structure: ASReml) or offer particular advantages for non-Gaussian distributions (e.g. MCMCglmm package in ʀ). None are 'point and click' based programs so the investigator will likely have to invest some time and effort learning the syntax. This may well be factor in deciding which software to use: for example, researchers already proficient in the use of R should find ASReml-R and MCMCglmm intuitive.

We present tutorials in four software packages: ASReml, ASReml-R, WOMBAT, and MCMCglmm (see Files S1–S5, Supporting Information; Table 2). MCMCglmm is the only one of these packages that uses Bayesian inference; the others listed employ restricted (or residual) maximum likelihood (REML). The relative merits of frequentist and Bayesian inference philosophies are a source of endless discussion and argument which we will not enter into here. On a purely practical note REML is faster and more widely used, but it also

Table 2. A list of some available software packages that can be used to run animal models, with details of whether the software is freely available, the method of statistical inference implemented (REML: restricted maximum likelihood; MCMC: Markov Chain Monte Carlo) and on-line sources of further information. This is not an exhaustive list and merely reflects the software the authors are familiar with

Software	Free to download/ use?	Inference	Notes/Website	Documentation
ASReml	No	REML	Owned and licensed by VSN International Ltd http://www.vsni.co.uk/software/asreml/	http://www.vsni.co.uk/resources/doc/ asreml2/UserGuide.pdf
ASReml-R	No	REML	Commercially available R interface for ASReml http://www.vsni.co.uk/software/asreml	http://www.vsni.co.uk/resources/doc/ asreml-R.pdf
DMU	Yes	REML or MCMC	http://www.dmu.agrsci.dk/	http://www.dmu.agrsci.dk/ dmuv6_guide-R4-6-7.pdf
MCMCglmm	Yes	MCMC	R package http://cran.r-project.org/web/packages/ MCMCglmm/index.html	http://cran.r-project.org/web/packages/ MCMCglmm/MCMCglmm.pdf
WOMBAT	Yes	REML	Replaces DFREML http://agbu.une.edu.au/˜kmeyer/wombat.html	http://agbu.une.edu.au/˜kmeyer/ WOMBAT/WWW/manual.html
VCE	Yes	REML or MCMC	http://vce.tzv.fal.de/software	ftp://ftp.tzv.fal.de/pub/latest_vce/ doc/vce6-manual-3·1-A4.pdf

has drawbacks. For instance, there are some difficulties and uncertainties associated with both parameter estimation and hypothesis testing for non-Gaussian traits (see later discussion and Bolker *et al*. 2009 for a good introduction to Generalized Linear Mixed Models). Bayesian inference via MCMC offers a useful way around some of these difficulties (Sorensen & Gianola 2002; Ovaskainen, Cano & Merila 2008; Hadfield *et al*. 2010; see also 'MCMC methods for multi-response generalised linear mixed models: the MCMCglmm R package' available from the author, J.D. Hadfield, Institute of Evolutionary Biology, University of Edinburgh, UK) but is slower (although not necessarily appreciably so for small problems) and requires specification of prior distributions for unknown parameters. In large informative data sets the prior should have little impact on parameter estimates, but it is important to check that results are robust to prior specification. We do not advocate any one program package but in the worked examples to follow we provide sample code to run animal models using ASReml, ASReml-R, WOMBAT and the R package 'MCMCglmm' (reflecting the programs with which we are most familiar). These programs offer wide capabilities for fitting mixed effect models and readers wishing to become generally proficient in their use should consult the extensive documentation that accompanies software distributions.

Getting stuck in

The best way to learn how to use the animal model is obviously to use it. To this end, we have provided some sample data and tutorials to get people started (see Files S1–S5, Supporting Information). These tutorials describe a series of quantitative genetic analyses on a population of gryphons (reflecting a compromise between the avian and mammalian biases of the authors). As the gryphon is a mythical beast the data provided were necessarily simulated. Phenotypes were simulated over an arbitrary pedigree structure using the program 'Pedantics' (Morrissey *et al*. 2007). The tutorial materials are designed to be self-contained and have been provided in four software-specific versions. Here, we give only a short description of their contents but we also highlight the salient points that should emerge from the tutorials. Thus, aside perhaps from the brief data file description below, this section will also be of relevance to readers who are not yet ready to start the accompanying exercises.

PREPARATION OF DATA FILES

Although software applications may differ slightly in formatting requirements (e.g. limits on the number of data columns, or characters in a field), they typically require data and pedigree files to be provided in plain text format, delimited by white space. In the examples provided, we use 'NA' to denote missing data and the tutorial and its required data and pedigree files is presented separately for each software package (see File S1, Supporting Information for further details). The pedigree file required for analysis comprises three columns of data, each line corresponding to an individual's own identity, its father and its mother. We have used numerical codes for the individuals but in general any alphanumeric code is normally acceptable (i.e. individual 23 could as easily be called 'Bob' or 'G17a', but avoid any special characters – although note special coding requirements for WOMBAT detailed in the relevant tutorial). The pedigree file is usually ordered such that the line specifying the parents of an individual animal appears before any line in which that individual is present as a parent (Fig. 1c). This ordering is a requirement for most software and in practice is often most simply achieved by sorting the file according to generation or cohort starting with the earliest (i.e. parents are always born before their offspring). Note that all individuals must have a record in the pedigree file but only those with phenotypic data need be present in the data file. The data file also has a first column of ID while subsequent columns include phenotypic traits and any additional variables that may be fit in the model.

TUTORIAL 1 – ESTIMATING THE HERITABILITY OF BIRTH WEIGHT

The first tutorial is designed to estimate the heritability of a single trait (birth weight in gryphons). We start with a very simple animal model containing effects of population mean and breeding value only to estimate the additive genetic variance and the heritability of the trait. We then explore the consequences of adding additional effects to the model. Two key points emerge. First, if we follow the common practice of defining V_P as the sum of estimated variance components (i.e. $V_A + V_R$) in the simple case then we generally expect the heritability of a trait to increase with inclusion of more fixed effects. This dependence of h^2 on the model structure (as well as the actual biology) may initially seem alarming but is in fact perfectly sensible. This is because by defining V_P in this way, when we add a fixed effect (in this case sex) we changed the interpretation of h^2 from the proportion of variance explained by additive effects to the proportion of variance left after accounting for sex that is explained by additive effects. See Wilson (2008) for more detailed discussion of this issue.

The second point that emerges is a practical demonstration of the bias that can be induced by common environment effects. Here, we simulated a non-genetic maternal effect in the data such that offspring of the same mother are more similar to one another than offspring from different mothers. Biologically this type of effect could occur if mothers differ in the levels of resource available to them during gestation (perhaps due to spatial heterogeneity in the environment). When we add an additional random effect of maternal identity, the bias is reduced and estimates of V_A (and hence h^2) are therefore lower. We also obtain an estimate of the variance caused by maternal effects (V_M). Other examples of potential biases when estimating V_A related to unspecified variance components can be seen in Table 1. Adding an additional random effect that is not confounded with the pedigree structure (here, year of birth) results in a further partition of variance (V_{BY}) but there is minimal change in V_A since, in our example, year of birth is not confounded with pedigree. See Kruuk & Hadfield (2007) for further discussion on this issue.

TUTORIAL 2 – A BIVARIATE ANIMAL MODEL

The second tutorial extends the first situation to the bivariate case where we are interested in the genetic covariance (and correlation) between two traits (birth weight and tarsus length at fledging) as well as the genetic variance for each. For multivariate models, it is generally useful to start thinking in terms of variance–covariance matrices. So, for this two trait model, we would consider the phenotypic matrix **P** as comprising phenotypic variances in birth weight (V_{P1}) and tarsus length (V_{P2}) and the phenotypic covariance between the two traits ($COV_{P1,P2}$). **P** is then initially decomposed into

the additive genetic matrix **G** and a residual (or environmental) matrix **R** where, for two traits:

$$\mathbf{P} = \mathbf{G} + \mathbf{R}$$

$$\begin{bmatrix} V_{P1} & COV_{P1,P2} \\ COV_{P1,P2} & V_{P2} \end{bmatrix} = \begin{bmatrix} V_{A1} & COV_{A1,A2} \\ COV_{A1,A2} & V_{A2} \end{bmatrix} + \begin{bmatrix} V_{R1} & COV_{R1,R2} \\ COV_{R1,R2} & V_{R2} \end{bmatrix}$$

Or, in the more complex situation where maternal and year of birth effects are included and where **M** and **BY** are the matrices corresponding to those additional random effects:

$$\mathbf{P} = \mathbf{G} + \mathbf{M} + \mathbf{BY} + \mathbf{R}$$

$$\begin{bmatrix} V_{P1} & COV_{P1,P2} \\ COV_{P1,P2} & V_{P2} \end{bmatrix} = \begin{bmatrix} V_{A1} & COV_{A1,A2} \\ COV_{A1,A2} & V_{A2} \end{bmatrix} + \begin{bmatrix} V_{M1} & COV_{M1,M2} \\ COV_{M1,M2} & V_{M2} \end{bmatrix} + \begin{bmatrix} V_{BY1} & COV_{BY1,BY2} \\ COV_{BY1,BY2} & V_{BY2} \end{bmatrix} + \begin{bmatrix} V_{R1} & COV_{R1,R2} \\ COV_{R1,R2} & V_{R2} \end{bmatrix}$$

So in multivariate analyses what we are really doing is fitting models of these variance-covariance structures to generate estimates of the elements within each. It is possible to impose certain constraints on one or more of the matrices in order to test particular hypotheses. For instance to test the significance of the genetic covariance we can compare the full model to one in which **G** is fitted with the condition that $COV_A = 0$, i.e.:

$$\mathbf{G} = \begin{bmatrix} V_{A1} & 0 \\ 0 & V_{A2} \end{bmatrix}$$

Similarly, if we suspect there really are no maternal effects on the second trait (tarsus length) then we might try a model where **M** is fitted as:

$$\mathbf{M} = \begin{bmatrix} V_{M1} & 0 \\ 0 & 0 \end{bmatrix}$$

Of course the two trait example presented here can be extended in principle to any number of traits. However, as the dimension of each matrix increases, the number of parameters to be estimated rises very quickly and you can soon run into difficulties getting your models to converge. The solution to this is to use simpler models, at least to start with. For instance, if you having trouble getting a bivariate model to converge then try modelling each trait in a univariate model first. This will give you a good idea of the variance components for each trait and these can be used as starting values in the bivariate analysis. If you want to estimate a full **G** matrix among a large number of traits then ultimately you may find that you cannot fit a full model but rather you will need to run a series of bivariate models to estimate each of the pairwise genetic covariances.

In the first two tutorials, we use a data set in which the traits of interest (birth weight and tarsus length at fledging) are measured only once per offspring. However, in many cases ecological studies generate multiple observations per individual. For example, in iteroparous organisms we may have repeated measures of reproductive traits. Here, we use the example of litter size, treating it as a female reproductive trait for which we have repeated measures.

With repeated measures on individuals a common starting point in both ecology and quantitative genetics is to partition the phenotypic variance into within- vs. between-individual components and this can be done here by fitting individual identity as a random effect without associating it with the pedigree (V_I in Table 1). The among-individual variance expressed as a proportion of the trait is the repeatability. As repeatability must equal heritability in the extreme case that all differences among individuals are caused by additive genetic effects, it is generally considered to set the upper limit for h^2 (Falconer & Mackay 1996), although there are actually some situations in which this need not hold true (see Dohm 2002). More generally, we might expect fixed or 'permanent' differences between individuals to arise through environmental and/or non-additive genetic effects as well (Table 1). As an individual is perfectly related to itself and completely shares its own environment, permanent environment effects can be seen as a special – and very extreme – case of the common environment problem. Consequently, we must always model this source of variance to protect against bias in V_A when we have repeated records. This is done by including an individual's identity twice in the animal model: first identity is associated with a pedigree structure to partition V_A; secondly, identity is fitted as a standard random effect to partition any additional non-genetic sources of fixed differences among individuals. This latter partition of 'permanent environment' variance V_{PE} will, despite its name, also include non-additive genetic effects (i.e. dominance variance) if present (see Table 1).

Using the worked example of lay date in gryphons, we can see that lessons from the first tutorial remain equally relevant. Thus, for instance, parameters can be influenced by inclusion of fixed effects. Here, we have simulated an age effect (a linear increase in lay date with age) and so fitting age will reduce V_R and increase h^2. However, similar changes to V_{PE} can also occur. For instance, if a trait is assayed in males and females in a sexually dimorphic organism then fitting sex will reduce the non-genetic between-individual variance (i.e. V_{PE}) rather than the within-individual environmental variance V_R. A final point to note is that to some extent inclusion of a permanent environment effect can be seen as a catch-all for unexplained environmental effects. We may well hypothesize that fixed differences among individuals arise from cohort, birth year or maternal effects and these explicit sources of common environment can be included in a repeated measures model. In the example provided birth year effects were simulated that have a permanent effect on an individual's phenotype

and when not explicitly modelled these effects contribute to V_{PE} for lay date. Inclusion of birth year results in an additional partition of V_{BY} as before with a corresponding decrease in the magnitude of V_{PE}. If inclusion of additional fixed and/or random effects explains environmental sources of among-individual variation, then V_{PE} may become small and could lack statistical significance. Under such circumstances it could be tempting to simplify the model by dropping the permanent environment effect but this should not be done. Unless V_{PE} truly is zero then its omission will upwardly bias V_A (see Kruuk & Hadfield (2007) for further discussion).

Fixed or random?

In the tutorials provided, we have obviously suggested which models to fit. However, when building your own models, it is necessary to decide not only what effects to include but also, in some cases, how they should be treated. One particular question that always arises for factors is whether they should be treated as fixed or random effects. Formally, the distinction can be made on the grounds that a factor will normally be treated as fixed if all levels are found in the data and the goal is to determine the effect on the mean of each factor level, whereas a factor will be treated as random if the levels represented in the data are a sample from a larger population about which the aim is to make some inference (Pinheiro & Bates 2000; Galwey 2006). Often the appropriate treatment of a particular effect is very obvious. For example, sex will (usually) have only two levels present and we are interested in knowing the effect of sex on the phenotypic mean. By contrast, the additive genetic effect will be treated as random since we do not have every genotype represented in the data and the goal is to use what we do have to make inferences about the level of genetic variance in the wider population. In addition to the additive effect we would typically model individual, mother and common environment effects in this way.

However, some effects are less clear cut. For example effects such as year of measurement or year of birth may plausibly be treated as either fixed or random. The former would generate an estimate of year specific effects on the mean (while soaking up one degree of freedom for each year represented in the data). Remaining variance, as partitioned across say additive and residual components, should then be interpreted as having been conditioned on the year effects. In contrast, treating year as a random effect will only use one degree of freedom and provide an estimate of how much variance is explained by year that could be extrapolated to a larger set (or population) of years than those actually present in the data. In making a decision about how to proceed it is useful to think about why the effect is being included and what information you actually want to extract from the model.

Hypothesis testing

In addition to parameter estimation, we will usually want to test the statistical significance of the one or more parameters against an appropriate null hypothesis. We may have deter-

mined that a trait's heritability is 0·2 but want to know whether this is significantly greater than zero. For a genetic correlation, we might sometimes wish to test against a null hypothesis of $r_G = 0$; other times, a more sensible null hypothesis might be that $r_G = 1$ (discussed further below). Appropriate tools of statistical hypothesis testing are not universally agreed upon for mixed models and the available methods may differ across software applications. The reader should therefore take the following as guidelines and hopefully useful suggestions, not as the definitive and unarguable truth! Although primarily focused on generalized mixed models, Bolker *et al*. (2009) provide a useful overview of some of the issues surrounding hypothesis testing for mixed models in general.

When fitting models using REML the standard errors associated with the estimated variance components should generally not be used for formal hypothesis testing, although we commonly do use them as a rough guide during preliminary model exploration. Instead a likelihood ratio test (LRT) can be constructed by comparing the log-likelihood of the model to a reduced model from which the effect of interest has been dropped. We normally define the test statistic as equal to twice the difference in log-likelihoods between the models, and assume that this is follows a chi-squared distribution with degrees of freedom equal to the number of additional parameters estimated in the more complex model (Pinheiro & Bates 2000). For the case of testing, the significance of a single variance component in a univariate model (e.g. V_A) this means that there is one degree of freedom. However, since this test is inherently two-tailed and variance components are (normally) expected to be constrained to positive parameter space, there is a good argument that this approach is overly conservative (see e.g. Gilmore *et al*. 2006). A widely used adjustment proposed by Stram & Lee (1994) amounts in practice to halving the *P*-value obtained from the conservative LRT. However, some authors have suggested this can sometimes result in an anticonservative test (Pinheiro & Bates 2000). Note also that this is not applicable when testing a covariance which is not bounded by zero. We do not take a strong position on this except to make the obvious point that, once made, a decision should probably be stuck with. It is certainly not acceptable to choose the form of test that gives you the most convenient answer.

Restricted Maximum Likelihood methodology is restricted in the sense that it only maximizes the likelihood that does not depend on the fixed effects. Therefore, likelihood comparisons are only valid under REML if models have identical fixed effect structures (Pinheiro & Bates 2000). Care must also be taken to ensure the same data set is being analysed since, for instance missing data for random effects (e.g. unknown mothers) could result in different patterns of data exclusion across models. Testing of fixed effects is also somewhat problematic as *F*-tests normally used in linear models require knowledge of the denominator degrees of freedom, which is hard to calculate for mixed models. Various options have been proposed for testing the significance of fixed effects in mixed models (Pinheiro & Bates 2000; Galwey

2006) and most software packages will provide some form of test of fixed effects. However, packages differ in the detail (e.g. method of determining the denominator degrees of freedom; Bolker *et al*. 2009) and users should therefore consult the relevant documentation on exactly what tests are being performed, how they are being constructed, and how they should be interpreted.

With Bayesian inference things are actually somewhat simpler. For example, using MCMC one can examine the posterior distribution for a parameter of interest (whether fixed or random) and see whether the 95% credible interval spans zero. One caveat to this is that prior specification may be such that some parameters must be positive and, as their 95% credible interval will never include zero, they will always be 'significant' based on this approach. For instance, this is the situation for variance components estimated using the R package 'MCMCglmm' and the use of an information theoretic approach based on the Deviance Information Criterion is consequently suggested for model selection (see 'MCMC methods for multi-response generalised linear mixed models: the MCMCglmm R package' available from the author, J.D. Hadfield, Institute of Evolutionary Biology, University of Edinburgh, UK).

What else is possible?

The examples presented in the tutorials above provide an introduction to the sorts of analyses that can be carried out using the animal model. Of course there are many variants on the basic model that can be specified depending on the questions being addressed. For example, by treating a phenotype measured in two environments as two distinct traits (rather than one trait with repeated measures) we can test for genotype-by-environment interaction (G × E). If the same genotype has different phenotypic expression in the two environments then we would expect $V_{A1} \neq V_{A2}$ while r_G across environments would be less than $+1$ (McAdam & Boutin 2003; Charmantier & Garant 2005). Similarly, we might divide a trait into numerous age-specific traits in order to explore the genetic processes underlying trade-offs between early and late fitness, or to test hypotheses stemming from evolutionary theory of ageing (e.g. Charmantier *et al*. 2006). A comparatively recent development within field studies of quantitative genetics is the use of so-called random regression animal models in which an individual's breeding value is modelled as a function of a covariate. The covariate may be an environmental variable in studies of plasticity and G × E, or age in studies of ontogeny, growth and senescence (Meyer & Kirkpatrick 2005; Nussey, Wilson & Brommer 2007; Wilson, Charmantier & Hadfield 2008). If a linear reaction norm is used this technique is equivalent to a 'random slope' model used to test for variation in plasticity (Nussey *et al*. 2007) and controls for pseudo-replication in behavioural studies (Schielzeth & Forstmeier 2009), but with separate functions included for both the individual level genetic (i.e. breeding value) and non-genetic (i.e. permanent environment) effects (Brommer, Rattiste & Wilson 2008).

An area of particular interest to behavioural ecologists is the analysis of sex-specific, or sex-limited traits, to explore questions relating to mate choice and sexual selection (Qvarnstrom, Brommer & Gustafsson 2006; Foerster *et al.* 2007). For instance, models of sexual selection on male secondary sexual traits typically rely on a genetic correlation between the male trait and female preference while inter-sex genetic correlations are also vital to our understanding of the evolution of sexual dimorphism (Fairbairn & Roff 2006). An interesting point is that genetic correlations can actually be estimated between sex-limited traits despite the fact that no individual ever expresses both phenotypes. This is because while a male will not express a female trait (and vice versa) he will have female relatives in the population who do. These relatives can then provide information as to the male's genetic merit for the unexpressed phenotype. Nevertheless, some caution is required when modelling sex-limited traits since while the genetic covariance can be estimated the same is not true for environmental covariance, and the total phenotypic correlation is obviously undefined.

Here, we have highlighted a few of the research topics in evolutionary ecology for which animal models offer great potential, but there are many others. For instance estimation of maternal genetic effects could give important insights into parent–offspring conflict (Wilson *et al.* 2005). Other types of indirect genetic effect, which are defined as occurring when the phenotype of a focal individual is influenced not only by its own genotype but by the genotype of others (Moore, Brodie & Wolf 1997), can also be estimated. Indirect genetic effects are expected to be important for behavioural traits like dominance and aggression (Moore *et al.* 1997; Wilson *et al.* 2009) as well as having implications for social evolution (Bijma & Wade 2008). Recently, Brommer *et al.* 2008 showed how this framework can also apply to reproductive traits with laying date in common gulls influenced by the male, as well as the female, genotype.

An exhaustive treatment of interesting scenarios that could be explored is beyond the scope of this paper but we refer the authors to Kruuk *et al.* 2008 for a broad review of applications to date. We have also produced a web-based resource at (http://www.wildanimalmodels.org) which provides tutorials including further code and example data sets to explore wider and more complex applications of the animal model.

On BLUP

Most biological questions suitable for quantitative genetic approaches can, and should, be translated into statistical hypotheses relating to components of (co)variances and parameters derived from these. However, in addition to these population-level parameters, the animal model can also be used to obtain estimates (or predictions) of the individual breeding values. These are usually obtained from REML-based analyses as best linear unbiased predictors (BLUP) of the true breeding values, which have long served as a useful tool in developing artificial selection schemes. More recently, evolutionary ecologists have used BLUP of individual breed-

ing values to explore some very interesting hypotheses. For example, by regressing fitness on the predicted breeding values studies have tested for selection acting on genotype and compared its strength to selection on the phenotype (e.g. Kruuk *et al.* 2001, 2002; Garant *et al.* 2004; Gienapp *et al.* 2006). Regressions of BLUP estimates of breeding values on time (or birth year) have been used to test for micro-evolutionary change in response to selection on traits (e.g. Merila, Kruuk & Sheldon 2001a; Coltman *et al.* 2003; Réale *et al.* 2003; Wilson *et al.* 2007). BLUPs have also been used to assess the genetic architecture and strength of selection on individual reaction norms (e.g. Brommer *et al.* 2005; Nussey *et al.* 2005).

Our omission of discussion of this topic until this point is deliberate. This is due to the fact that it is now clear that we cannot use BLUP to reliably test these hypotheses. Use of BLUP in the ways alluded to above can result in massive bias and extreme anti-conservatism. Problems arise not from an inherent problem with BLUP but from a failure to fully appreciate the statistical consequences of using *predicted*, rather that *true*, breeding values, coupled with the use of logically inconsistent models to generate the BLUP (Postma 2006; Hadfield *et al.* 2010). We refer the reader to Hadfield *et al.* (2010) for a full treatment of this. Here, we strongly echo the take home message of that paper: unless robust applications become apparent, the use of BLUP in evolutionary ecology should be discontinued. Fortunately, as discussed in Hadfield *et al.* (2010), there are also ways to explore the interesting hypotheses without resorting to analysing BLUP. For instance, the strength of selection on genotype can be determined from the genetic covariance between a trait and fitness and can therefore be directly estimated using a bivariate animal model. For tests of micro-evolutionary change, Bayesian approaches that properly account for measurement error in predicted breeding values can be used (Hadfield *et al.* 2010).

Other pitfalls and problems

Having highlighted the exciting possibilities opened up by application of the animal model to data from ecological studies, it is important to acknowledge that there are also methodological limitations and a number of pitfalls to avoid. Perhaps the most likely trap to fall into is that of failing to recognize or account for a likely source of bias. As we have seen, it is often possible to protect against common environment effects by inclusion of additional terms in the model, but this is not a bullet-proof strategy. The confounding variable will not always have been recorded, but this does not mean it isn't there. Even where an effect can be modelled, success will depend on data structure. For instance, the ability to statistically separate maternal from additive effects requires that mothers have multiple offspring and will be greatly improved by half-sib structure in the pedigree (Kruuk & Hadfield 2007). This is because paternal half sibs will share additive effects, while maternal half-sibs will share additive and maternal effects. Conversely, the separation will be made

problematic by large amounts of unknown paternity. Disentangling additive and common environment effects will also be improved by experimental manipulations where these are possible (e.g. cross-fostering in passerine studies) and so sophisticated analytical models should never be seen as a substitute for rigorous study design (Kruuk & Hadfield 2007).

A more general issue is that inclusion of additional effects in a model often leads to attendant declines in the precision of variance component estimation. In itself this should not necessarily be seen as a problem: large standard errors probably indicate that you are pushing the data too far so that estimated model parameters should be treated with appropriate caution. However, difficulties can arise if statistical significance is used as the basis of model selection. For example, imagine that a simple animal model yields a large and highly significant estimate for a trait's heritability. If we added a potentially confounding common environment effect (e.g. nest, territory or mother) but found that this second model was not a significantly better fit we might feel justified in presenting our heritability estimate from the first model. However, if the two effects are badly confounded then it could be that neither was statistically significant in the second model. How should one interpret such a situation in which the model selected could actually depend on the order of effects fitted? Faced with a choice between potentially biased but statistically significant h^2 estimate on the one hand and a more conservative but non-significant estimate on the other the optimist (e.g. researcher) may be tempted by the former while the cynic (e.g. reviewer) will almost certainly prefer the latter. Probably a reasonable middle ground in this particular case would be to present both models with appropriate caveats for the interpretation for each. An appropriate conclusion would be to say that there is a strong association between the pedigree and the phenotype, and while this is consistent with a high h^2 we cannot exclude the confounding common environment effect. Ultimately model selection will be determined not just by statistical significance, but also the biological question, the available data, and the potential sources of bias. Proper consideration of alternative models is therefore vital to properly interpreting the results of any study (Kruuk & Hadfield 2007). However, as noted at the outset, statistical power will almost always limit the models that can sensibly be tested and one could always posit a more complex model that could have been attempted if only the data were up to it. Highlighting power limitations and potentially confounding effects is therefore an important part of qualifying conclusions and should not necessarily be viewed as an admission of failure in study design.

Problems with biological interpretation can sometimes arise from a failure to interpret parameters in the context of the model in which they were estimated. We have already drawn attention to the way in which estimates of heritability can depend on the inclusion of other effects in the model (see above; Wilson 2008) and so must be interpreted with care. However, translating from quantitative genetic model to ecological interpretation offers the unwary reader plenty more opportunities for misunderstanding. For example, as we have illustrated in the tutorials, maternal effects are commonly modelled by inclusion of the maternal identity as a random effect. However, while this term is expected to account for fixed or permanent among-mother heterogeneity (i.e. differences between litters with different mothers), ecologists are frequently interested in maternal effects arising from maternal age, status or body condition that will vary between litters within a mother. Thus an absence of V_M does not necessarily mean that maternal effect in the wider sense do not occur, but rather that maternal identity does not contribute to among-litter differences (beyond the expected additive genetic effect). Other types of maternal effect could equally be looked at in an animal model framework, for example by fitting as fixed effects on offspring phenotype, or in the case of maternal condition by treating it as a second, potentially covarying, trait in a bivariate model.

Finally, having estimated quantitative genetic parameters it is perhaps not unreasonable that a researcher might wish to combine these with field-based estimates of natural selection in order to generate predictions of phenotypic change. Here, we would advocate caution. Although quantitative genetics is a predictive subject, it is fair to say that to date its predictions have had little success in natural systems (Merila, Sheldon & Kruuk 2001b; Kruuk 2004; Charmantier & Garant 2005; Kruuk *et al.* 2008). There are many reasons why the mismatch between prediction and observation occurs, all of which can be seen as violations of the simple models usually used to predict change (see Merila *et al.* 2001b). For instance the breeder's equation predicts that a per generation response to selection can be estimated as the product of a trait's heritability and the selection differential on it (the covariance between trait and fitness). However, this model is only intended as a complete description of phenotypic change in a population when a single trait is under selection, genetic drift is negligible, generations are discrete and non-overlapping, and the environment is constant. These conditions can certainly be approximated under experimental conditions but are very unlikely to hold true in any wild population, particularly when one recognizes that what is really meant by 'environment' in this context is everything other than the additive genetic effects on the focal trait (e.g. extrinsic environmental variables but also population size, demographic structure, resource abundance, predator and parasite levels). Consequently demonstrating that predictions from the breeder's equation do not match observations in natural populations is to some extent an attack on a straw man. The biological interest lies not in demonstrating that the model is too simplistic, but rather in exploring why.

Acknowledgements

Additional practical resources

PRACTICAL GUIDE TO QUANTITATIVE GENETICS FOR EVOLUTIONARY ECOLOGISTS

http://www.wildanimalmodels.org

Set of tutorials with sample data files highlighting evolutionary ecological applications of the animal model. Code is provided for use with ASReml, ASReml-R, WOMBAT and the R package 'MCMCglmm'.

ASREML COOKBOOK

http://uncronopio.org/ASReml/HomePage

Excellent site written and maintained by Luis Apiolaza providing information and code to help new ASReml users get started and progress to more complex models.

MIXED MODELS FOR GENETIC ANALYSIS

http://www.vsni.co.uk/products/asreml/user/geneticanaly-sis.pdf

Comprehensive notes written by Julius van der Werf describing the use of mixed models for genetic analysis. Emphasis is on applications for livestock breeding programs. Includes some code for ASReml.

References

Bijma, P. & Wade, M.J. (2008) The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *Journal of Evolutionary Biology*, **21**, 1175–1188.

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**, 127–135.

Brommer, J.E., Merila, J., Sheldon, B.C. & Gustafsson, L. (2005) Natural selection and genetic variation for reproductive reaction norms in a wild bird population. *Evolution*, **59**, 1362–1371.

Brommer, J.E., Rattiste, K. & Wilson, A.J. (2008) Exploring plasticity in the wild: laying date-temperature reaction norms in the common gull larus canus. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **275**, 687–693.

Carroll, S.P., Hendry, A.P., Reznick, D.N. & Fox, C.W. (2007) Evolution on ecological time-scales. *Functional Ecology*, **21**, 387–393.

Charmantier, A. & Garant, D. (2005) Environmental quality and evolutionary potential: lessons from wild populations. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **272**, 1415–1425.

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

Charmantier, A., Perrins, C., McCleery, R.H. & Sheldon, B.C. (2006) Age-dependent genetic variance in a life-history trait in the mute swan. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **273**, 225–232.

Coltman, D.W., O'Donoghue, P., Jorgenson, J.T., Hogg, J.T., Strobeck, C. & Festa-Bianchet, M. (2003) Undesirable evolutionary consequences of trophy hunting. *Nature*, **426**, 655–658.

Dohm, M.R. (2002) Repeatability estimates do not always set an upper limit to heritability. *Functional Ecology*, **16**, 273–280.

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

Fairbairn, D.J. & Roff, D.A. (2006) The quantitative genetics of sexual dimorphism: Assessing the importance of sex-linkage. *Heredity*, **97**, 319–328.

Falconer, D.S. & Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edition. Prentice Hall, London.

Flux, J.E.C. & Flux, M.M. (1982) Artificial selection and gene flow in wild starlings, *Sturnus vulgaris*. *Naturwissenschaften*, **69**, 96–97.

Foerster, K., Coulson, T., Sheldon, B.C., Pemberton, J.M., Clutton-Brock, T.H. & Kruuk, L.E.B. (2007) Sexually antagonistic genetic variation for fitness in red deer. *Nature*, **447**, 1107–1109.

Galwey, N.W. (2006) *Introduction to Mixed Modelling: Beyond Regression and Analysis Of Variance*. Wiley, Chichester.

Garant, D., Kruuk, L.E.B., McCleery, R.H. & Sheldon, B.C. (2004) Evolution in a changing environment: a case study with great tit fledging mass. *American Naturalist*, **164**, E115–E129.

Gienapp, P., Postma, E. & Visser, M.E. (2006) Why breeding time has not responded to selection for earlier breeding in a songbird population. *Evolution*, **60**, 2381–2388.

Gilmore, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2006) *Asreml User Guide Release 2.0*. VSN International Ltd., Hemel Hempsted.

Gustafsson, L. (1986) Lifetime reproductive success and heritability – empirical support for fisher's fundamental theorem. *American Naturalist*, **128**, 761–764.

Hadfield, J.D., Wilson, A.J., Garant, D., Sheldon, B.C. & Kruuk, L.E.B. (2010) The misuse of BLUP in ecology and evolution. *American Naturalist*, **175**, DOI:10.1086/648604.

Hairston, N.G., Ellner, S.P., Geber, M.A., Yoshida, T. & Fox, J.A. (2005) Rapid evolution and the convergence of ecological and evolutionary time. *Ecology Letters*, **8**, 1114–1127.

Henderson, C.R. (1953) Estimation of variance and covariance components. *Biometrics*, **9**, 226–252.

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

Henderson, C.R. (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph Press, Guelph.

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**, 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**, 1890–1903.

Kruuk, L.E.B., Clutton-Brock, T.H., Slate, J., Pemberton, J.M., Brotherstone, S. & Guinness, F.E. (2000) Heritability of fitness in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 698–703.

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

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

Kruuk, L.E.B., 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**, 525–548.

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

McAdam, A.G. & Boutin, S. (2003) Effects of food abundance on genetic and maternal variation in the growth rate of juvenile red squirrels. *Journal of Evolutionary Biology*, **16**, 1249–1256.

Merila, J. & Sheldon, B.C. (2001) Avian quantitative genetics. *Current Ornithology*, **16**, 179–255.

Merila, J., Kruuk, L.E.B. & Sheldon, B.C. (2001a) Cryptic evolution in a wild bird population. *Nature*, **412**, 76–79.

Merila, J., Sheldon, B.C. & Kruuk, L.E.B. (2001b) Explaining stasis: microevolutionary studies in natural populations. *Genetica*, **112-113**, 199–222.

Meyer, K. (1985) Maximum-likelihood estimation of variance-components for a multivariate mixed model with equal design matrices. *Biometrics*, **41**, 153–165.

Meyer, K. & Kirkpatrick, M. (2005) Up hill, down dale: quantitative genetics of curvaceous traits. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1443–1455.

Milner, J.M., Elston, D.A. & Albon, S.D. (1999) Estimating the contributions of population density and climatic fluctuations to interannual variation in survival of Soay sheep. *Journal of Animal Ecology*, **68**, 1235–1247.

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**, 804–813.

Moore, A.J., Brodie, E.D. & Wolf, J.B. (1997) Interacting phenotypes and the evolutionary process .1. Direct and indirect genetic effects of social interactions. *Evolution*, **51**, 1352–1362.

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**, 2309–2321.

van Noordwijk, A.J. (1984). Quantitative genetics in natural populations of birds illustrated with examples from the great tit, *parus major*. *Population Biology and Evolution* (ed. K.W.A.V. Loeschcke), pp. 67–79. Springer-Verlag, Berlin.

Nussey, D.H., Postma, E., Gienapp, P. & Visser, M.E. (2005) Selection on heritable phenotypic plasticity in a wild bird population. *Science*, **310**, 304–306.

Nussey, D.H., Wilson, A.J. & Brommer, J.E. (2007) The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, **20**, 831–844.

Ovaskainen, O., Cano, J.M. & Merila, J. (2008) A bayesian framework for comparative quantitative genetics. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **275**, 669–678.

Pelletier, F., Clutton-Brock, T., Pemberton, J., Tuljapurkar, S. & Coulson, T. (2007) The evolutionary demography of ecological change: Linking trait variation and population growth. *Science*, **315**, 1571–1574.

Pemberton, J.M. (2008) Wild pedigrees: the way forward. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **275**, 613–622.

Perrins, C.M. & Jones, P.J. (1974) Inheritance of clutch size in great tit (*Parus major L.*). *Condor*, **76**, 225–228.

Pinheiro, J.C. & Bates, D.M. (2000) *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York.

van de Pol, M.V. & Wright, J. (2009) A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour*, **77**, 753–758.

Postma, E. (2006) Implications of the difference between true and predicted breeding values for the study of natural selection and micro-evolution. *Journal of Evolutionary Biology*, **19**, 309–320.

Postma, E. & Charmantier, A. (2007) What 'animal models' can and cannot tell ornithologists about the genetics of wild populations. *Journal Of Ornithology*, **148**, S633–S642.

Qvarnstrom, A., Brommer, J.E. & Gustafsson, L. (2006) Testing the genetics underlying the co-evolution of mate choice and ornament in the wild. *Nature*, **441**, 84–86.

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

Réale, D., McAdam, A.G., Boutin, S. & Berteaux, D. (2003) Genetic and plastic responses of a northern mammal to climate change. *Proceedings of the Royal Society of London, Series B*, **270**, 591–596.

Saccheri, I. & Hanski, I. (2006) Natural selection and population dynamics. *Trends in Ecology & Evolution*, **21**, 341–347.

Schielzeth, H. & Forstmeier, W. (2009) Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, **20**, 416–420.

Shaw, R.G. (1987) Maximum-likelihood approaches applied to quantitative genetics of natural-populations. *Evolution*, **41**, 812–826.

Smith, H.G. & Wettermark, K.J. (1995) Heritability of nestling growth in cross-fostered European starlings *Sturnus vulgaris*. *Genetics*, **141** 657–665.

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

Stram, D.O. & Lee, J.W. (1994) Variance-components testing in the longitudinal mixed effects model. *Biometrics*, **50**, 1171–1177.

Thompson, R. (2008) Estimation of quantitative genetic parameters. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **275**, 679–686.

Van Noordwijk, A.J., van Balen, J.H. & Scharloo, W. (1981) Genetic and environmental variation in the clutch size of the great tit (*Parus major*). *Netherlands Journal of Zoology*, **31**, 342–372.

Wilson, A.J. (2008) Why h² does not always equal v_a/v_p? *Journal of Evolutionary Biology*, **21**, 647–650.

Wilson, A.J., Coltman, D.W., Pemberton, J.M., Overall, A.D.J., Byrne, K.A. & Kruuk, L.E.B. (2005) Maternal genetic effects set the potential for evolution in a free-living vertebrate population. *Journal of Evolutionary Biology*, **18**, 405–414.

Wilson, A.J., Pemberton, J.M., Pilkington, J.G., Clutton-Brock, T.H., Coltman, D.W. & Kruuk, L.E.B. (2007) Quantitative genetics of growth and cryptic evolution of body size in an island population. *Evolutionary Ecology Research*, **21**, 337–356.

Wilson, A.J., Charmantier, A. & Hadfield, J.D. (2008) Evolutionary genetics of ageing in the wild: empirical patterns and future perspectives. *Functional Ecology*, **22**, 431–442.

Wilson, A.J., Gelin, U., Perron, M.C. & Reale, D. (2009) Indirect genetic effects and the evolution of aggression in a vertebrate system. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **276**, 533–541.

Supporting Information

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

File S1. Notes on tutorials and software.

File S2. *Tutorial for ASReml* With required data files: 'gryphon.ped'; 'gryphon.dat'; 'gryphonRM.dat'.

File S3. *Tutorial for ASReml-R* With required data files: 'gryphon-ped.txt'; 'gryphon.txt'; 'gryphonRM.txt'.

File S4. *Tutorial for WOMBAT* With required data files: 'gryphon.ped'; 'gryphon_uni.dat'; 'gryphon_bi.dat'; 'gryphonRM.dat'; 'gryphonRM_pe.dat'.

File S5. *Tutorial for MCMCglmm* With required data files: 'gryphon-ped.txt'; 'gryphon.txt'; 'gryphonRM.txt'.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.