REVIEW

Why h^2 does not always equal V_A/V_P?

A. J. WILSON

Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

Abstract

Over the last decade, there has been a rapid growth in the application of quantitative genetic techniques to evolutionary studies of natural populations. Whereas this work yields enormous insight into evolutionary processes in the wild, the use of modelling techniques and strategies adopted from animal breeders means that estimates of trait heritabilities (h^2) are highly vulnerable to misinterpretation. Specifically, when estimated using animal models, h^2 will not generally be comparable across studies and must be interpreted as being conditioned on any fixed effects included in the model. Failure to realize the model dependency of published h^2 estimates will give a very misleading, and in most cases upwardly biased, impression of the potential for trait evolution.

Introduction

The heritability (h^2) of a phenotypic trait is defined as the proportion of phenotypic variance that is attributable to additive genetic effects, and is a key parameter in determining the evolutionary response to selection (Falconer & Mackay, 1996). As such, evolutionary biologists have long been interested in heritability, both for predicting selection responses, and as a tool to assess more general questions. For example, can laboratory-based estimates be usefully extrapolated to wild populations (Weigensberg & Roff, 1996)? Are some types of trait more heritable than others (Merilä & Sheldon, 1999)? Does heritability change consistently with the quality of the environment (Hoffman & Merila, 1999; Charmantier & Garant, 2005)? However, whereas it is well known that heritability is specific to a trait and a population, it is rather less appreciated that parameter estimates are also heavily determined by the structure of the model used for their estimation. This issue has enormous potential to confuse and mislead evolutionary ecologists, particularly researchers who are less familiar with quantitative genetic techniques.

Over recent years, there has been a surge of interest in the application of quantitative genetic models to data from natural populations (Kruuk, 2004; Postma & Char-

mantier, 2007). To a large extent, this endeavour has been facilitated by the adoption of the animal model, a form of mixed effects model that has long been used by animal breeders (Henderson, 1984). Mixed effects models contain both fixed and random effects. Fixed effects are used to model population-level average responses to explanatory variables, whereas random effects allow remaining variance to be partitioned into components attributable to any grouping factors present in the data (Galwey, 2006). By definition, an animal model includes an individual's additive genetic merit (or breeding value, Lynch & Walsh, 1998; as a random effect, and as genes are shared among relatives, this allows estimation of the additive genetic variance (V_A) for a trait of interest. By comparison with more traditional analytical techniques (e.g. parent–offspring regression), this method offers greater power and flexibility, particularly when dealing with complex pedigree structures typical of natural populations (Kruuk, 2004).

A further advantage of animal models is that fixed effects can readily be included, such that an individual's phenotype is 'corrected' for known sources of variation, such as age and sex, or environmental conditions, such as density or food abundance. For animal breeding applications, the inclusion of fixed effects is used to protect against downward bias in heritability estimates. For example, if one is interested in estimating the heritability of body weight, but individual animals have been measured at different stages of growth, then fitting age as a fixed effect in the model corrects for this allowing a

Correspondence: Alastair J. Wilson, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK.
Tel.: +44 131 6513608; fax: +44 131 6505455;
e-mail: alastair.wilson@ed.ac.uk

more meaningful comparison among individuals. This same argument also applies to studies of natural populations and a brief survey of the literature shows that studies using animal models to estimate h^2 in the wild have, almost without exception, included fixed effects in the models (for some recent examples, see Wilson *et al.*, 2005; Qvarnström *et al.*, 2006; Thériault *et al.*, 2007).

Whereas heritability is determined as the ratio of additive (V_A) to phenotypic (V_P) variance, V_P is most often defined as being the variance around the fixed effects mean. That is, to say, V_P is determined as the sum of the variance components associated with each random effect, and will therefore not include any variance explained by the fixed effects of a model. Consequently, heritability estimates will be critically determined by the fixed effects structure of the model used. The purpose of this note is to draw attention to the consequences of this practice, encourage discussion on the appropriate use of fixed effects, and demonstrate the vital importance of interpreting h^2 estimates in the context of the model used for estimation.

A simple example: h^2 of horn length in unicorns

This effect is most readily demonstrated by example. After defining arbitrary **G** (additive genetic) and **R** (residual) covariance matrices for a pair of traits, horn length and body mass in unicorns, phenotypes were simulated across a hypothetical pedigree structure (containing 1900 individuals drawn from three generations) using PEDANTICS (Morrissey *et al.*, 2007). Horn length and body mass were simulated such that both traits are heritable, and positively genetically correlated. Individual phenotypes were then completed by adding additional effects to horn length such that males have (on average) longer horns than females, horn length increases with age, and horns grow faster in males. Using ASReml, phenotypic variance was decomposed into additive genetic (V_A), and residual (environmental) variance (V_R), using a series of animal models differing only in their fixed effects structure (Table 1). Heritability (h^2) was then estimated as the ratio of V_A to phenotypic variance (V_P) under each model. Note that only a single

data set was analysed as the purpose is simply to illustrate the effect described above (i.e. not to perform a rigorous simulation exercise).

Following the common practice of defining V_P as the sum of variance components (i.e. $V_A + V_R$), it is clear that inclusion of fixed effects causes the expected increase in heritability estimates by reducing the magnitude of V_R and hence V_P (Table 1). Thus, the heritability of horn length increases from 0.116 (model 1a) to 0.554 when age and sex are included (model 1b) and to 0.585 when the interaction of age and sex is fit (model 1c).

Inclusion of these fixed effects necessarily produces a better fitting model as the age and sex effects are truly present in the simulated data. Here, statistical testing would therefore confirm the significance of the fixed effects and by conventional model selection procedures we would choose model 1c as the 'best' model. However, it does not necessarily follow that the heritability estimate is 'better' under model 1c. More generally, a common objective in biological modelling is to try and maximize the explanatory power of a model. Thus, a model is usually preferred if it explains more of the variance in the response (e.g. higher R^2 in a general linear model), and leaves less unexplained residual variance. With V_P calculated as the sum of variance components, a model selection strategy based on explaining as much variance as possible renders a naive interpretation of h^2 (as the proportion of variance explained by additive effects) useless. The more the knowledge about environmental effects on phenotype, the higher the heritability becomes.

Comparing the variance component estimates under models 1a–c, it is clear that fixed effects reduce V_R and hence increase h^2. However, the inclusion of fixed effects should not, in general, cause systematic changes in the estimated additive variance, and this is reflected by similar estimates of V_A across models 1a–c. Consequently, if V_A is scaled by the observed phenotypic variance in the data (as opposed to the sum of the variance components), then heritability is effectively constant across models (Table 1). Nevertheless, it is important to note that fixed and random effects are jointly estimated in an animal model and changing the fixed effects can therefore influence V_A estimates. This will especially be the case

Table 1 Variance components and heritability for horn length in unicorns estimated under models of differing fixed effects.

Model	Fixed effects	V_A	V_R	V_P	h^2	$V_{P(obs)}$	$V_A/V_{P(obs)}$
1a	Mean	0.361 (0.115)	3.117 (0.138)	3.478 (0.115)	0.116 (0.040)	3.466	0.104
1b	Mean + age + sex	0.362 (0.052)	0.653 (0.039)	1.014 (0.038)	0.554 (0.106)	3.466	0.104
1c	Mean + age + sex + age:sex	0.351 (0.049)	0.599 (0.036)	0.950 (0.035)	0.585 (0.110)	3.466	0.101
2a	Mean + weight	0.239 (0.097)	3.042 (0.129)	3.281 (0.1074)	0.078 (0.034)	3.466	0.069
2b	Mean + age + sex + weight	0.234 (0.039)	0.601 (0.033)	0.836 (0.030)	0.389 (0.080)	3.466	0.068
2c	Mean + age + sex + age:sex + weight	0.228 (0.036)	0.546 (0.029)	0.774 (0.028)	0.418 (0.083)	3.466	0.066

Phenotypic variance V_P is determined as the sum of V_A and V_R, and h^2 as the ratio of V_A to V_P. Also shown is the observed phenotypic variance estimated directly from the data [$V_{P(obs)}$] and the heritability calculated using $V_{P(obs)}$ [i.e. $V_A/V_{P(obs)}$].

if the incidence of fixed effects is nonrandom with respect to the pedigree structure. For example, if related individuals are more likely to be measured under similar environmental conditions, then failure to include environment effects could upwardly bias V_A (Kruuk & Hadfield, 2007).

Furthermore, a particular situation also exists where inclusion of a fixed effect can decrease the estimated genetic variance for a trait. In our example, this occurs when body weight is included as a fixed effect in the animal model of horn length (Table 1). Comparing models 2a–c with models 1a–c shows that including weight yields lower estimates of V_A, with concomitant declines in the heritability. This is because the structure of the **G** matrix used to simulate the phenotypes was such that a positive genetic correlation exists between weight and horn length. Thus, accounting for differences among individuals in the weight actually removes a portion of additive variance for the horn length. Thus, the estimates of V_A (and heritability) under models 2a–c are conditioned on the second trait of weight.

From a quantitative genetic perspective, a more sensible strategy here might be to model horn length and body weight as two traits in a bivariate animal model, thus estimating h^2 for each trait and the genetic correlation between them. More generally, a good rule of thumb might be not to include variables as fixed effects if they would reasonably be expected to have an additive genetic component of variance (and potentially additive covariance with the focal trait) themselves. However, a blanket application of this rule may perhaps be problematic where researchers need to use biological measures as surrogates for environmental conditions. For example, models of reproductive traits may attempt to control for among-individual differences in environments experienced by including weight or body condition as a proxy for food abundance. If carried out in a quantitative genetic model, it would at least seem prudent to estimate (co)variance components associated with such environmental proxies in order to correctly interpret results.

Implications of model dependence

The dependence of heritability on the fixed effects structure of a model has several major implications, and it is important that these should be kept in mind when interpreting the results of quantitative genetic studies in wild populations.

Heritability is not (necessarily) a standardized measure of V_A

Perhaps the most important point is that the h^2 values most commonly reported cannot be viewed as a properly standardized measure of genetic variance to be compared across traits or populations. For example, presented with the results in Table 1, an uninformed reader might

reasonably conclude that additive genetic effects on horn length contribute (significantly) more to phenotypic variation in some cases than others. The fallacy of this interpretation is self-evident, but might be considerably less apparent where results are derived from different studies.

It is also to be expected that routine inclusion of fixed effects will result in animal model estimates of heritability being higher than those derived from simpler techniques (e.g. parent–offspring regression), except where the latter are upwardly biased by common environment or maternal effects. It is therefore clear that current practices for generating h^2 estimates have enormous potential to mislead unwary readers, and will also render presented results largely unsuitable for meta-analytic studies. To a large extent, these problems could be avoided by using alternate standardizations of the genetic variance (e.g. scaling by the mean to give the coefficient of additive genetic variation; (Houle, 1992). However, given the enduring appeal of h^2 as a summary statistic, perhaps the most important point to note is that the values of h^2 need to be interpreted in the context not just of the biological system (trait and population) in which they were estimated, but also in the context of the model used to determine them.

Selection may be blind to fixed effects

The breeder's equation states that the per-generation change in a trait mean can be predicted as the product of its heritability and the selection differential, S (Falconer & Mackay, 1996). Thus, all else being equal, the rate of evolutionary change in a trait is directly proportional to h^2. However, in the above example, h^2 under models 1a–c showed a fivefold range (from 0.116 to 0.585) and it is therefore vital to use the appropriate estimate.

In this context, a question that arises is whether selection can 'see' fixed effects or is actually 'blind' to them. There is likely to be an important difference in this respect between the nature of artificial selection practised by animal breeders and that of natural selection occurring in a wild population. Thus, in optimizing a selective programme, an animal breeder may certainly wish to account for known effects on phenotype (e.g. if some animals have been raised on a different diet or measured at different stages of ontogeny) whereas natural selection will, in general, take no account of such effects. Thus, natural selection acts on phenotypes, not on residuals of phenotypes corrected for fixed effects. For example, a gape-limited predator (e.g. fish) will selectively remove animals below a certain size threshold, regardless of the reason that they are small.

In practice, the detailed mechanism of a selective event may be unknown in natural systems, but the strength of selection can still be estimated as the covariance between fitness (or some surrogate thereof) and the phenotypic trait. If the strength of selection is estimated 'blind', that

is to say using the raw phenotypic data without correction for fixed covariates, then the most appropriate estimate of heritability would be that derived from a model with no fixed effects. Use of alternate estimates will probably give an upwardly biased expectation of evolutionary change. Alternately, if the strength of selection is estimated in a way that is conditioned on covariates (e.g. age, sex and environmental conditions), then the appropriate estimate of h^2 would also be one conditioned on those effects.

Recommendations for evolutionary ecologists

Above, I have highlighted the pitfalls relating to interpreting published estimates of trait heritability from studies of wild populations. The question is what could or should be done? Fixed effects are included in animal models for good and sensible reasons and I certainly do not suggest a wholesale change to this practice. However, in the light of these issues, it is hoped that researchers performing quantitative genetic analyses will give greater consideration to the possibility of their results being misinterpreted. In general, the method of determining the denominator (i.e. V_P) has been stated explicitly in the methods sections of published work. However, it is fair to say that the full implications of this could, and should, be made more explicit. Furthermore, I suggest that researchers always provide a simple estimate of the phenotypic variance observed in the raw data, thereby allowing a standardized version of h^2 to be calculated by a reader if desired. More generally, it is vital that increased attention is given to the model dependency of heritability estimates. Thus, although heritability is certainly a population- and trait-specific parameter, it should also be viewed as model specific. This does not negate a straight-forward interpretation of h^2, but it is vital to understand that the phenotypic variance being partitioned is usually that which remains *after* conditioning on any fixed effects. Failure to realize this will lead, in some cases, to grotesquely biased expectations for evolutionary change.

Acknowledgments

References

Charmantier, A. & Garant, D. 2005. Environmental quality and evolutionary potential: lessons from wild populations. *Proc. R. Soc. Lond. B Biol. Sci.* **272**: 1415–1425.

Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics*, 4th edn. Longman, Essex.

Galwey, N.W. 2006. *Introduction to Mixed Modelling. Beyond Regression and Analysis of Variance*. Wiley, Chichester.

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

Hoffman, A.A. & Merila, J. 1999. Heritable variation and evolution under favourable and unfavourable conditions. *Trends Ecol. Evol.* **14**: 96–101.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* **130**: 195–204.

Kruuk, L.E.B. 2004. Estimating genetic parameters in natural populations using the 'animal model'. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**: 873–890.

Kruuk, L.E.B. & Hadfield, J. 2007. How to separate genetic and environmental causes of similarity between relatives. *J. Evol. Biol.* **20**: 1890–1903.

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

Merilä, J. & Sheldon, B.C. 1999. Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity* **83**: 103–109.

Morrissey, M.B., Wilson, A.J., Pemberton, J.M. & Ferguson, M.M. 2007. A framework for power and sensitivity analyses for studies of the quantitative genetics of natural populations, and a case study in Soay sheep (*Ovis aries*). *J. Evol. Biol.* **20**: 2309–2321.

Postma, E. & Charmantier, A. 2007. What 'animal models' can and cannot tell ornithologists about the genetics of wild populations. *J. Ornithol.* **148**: 633–642.

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

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*). *J. Evol. Biol.* **20**: 2266–2277.

Weigensberg, I. & Roff, D.A. 1996. Natural heritabilities: can they be reliably estimated in the laboratory? *Evolution* **50**: 2149–2157.

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. *J. Evol. Biol.* **18**: 405–414.