Quantitative genetic methods depending on the nature of the phenotypic trait

Pierre de Villemereuil

Quantitative genetic methods depending on the nature of the phenotypic trait

Pierre de Villemereuil[a]

[a] *School of Biological Sciences, University of Auckland, Auckland (New Zealand)*
bonamy@horus.ens.fr

Abstract

A consequence of the assumptions of the infinitesimal model, one of the most important theoretical foundations of quantitative genetics, is that phenotypic traits are predicted to be most often normally distributed (so-called Gaussian traits). But phenotypic traits, especially those interesting for evolutionary biology, might be shaped according to very diverse distributions. In this review, I show how quantitative genetics tools have been extended to account for a wider diversity of phenotypic traits using first the threshold model, then more recently using generalised linear mixed models. I explore the assumptions behind these models and how they can be used to study the genetics of non Gaussian complex traits. I also comment on three recent methodological advances in quantitative genetics that widen our ability to study new kinds of traits: the use of "modular" hierarchical modelling to e.g. study survival in the context of capture-recapture approaches for wild populations; the use of aster models to study a set of traits with conditional relationships (e.g. life-history traits); and finally, the study of high-dimensional traits such as gene expression.

Keywords: quantitative genetics, non Gaussian traits, heritability, threshold model, generalised linear mixed models

Corresponding author: Pierre de Villemereuil, E-mail: bonamy@horus.ens.fr

Short title: Quantitative genetics for non classical traits

Published in: *The Year in Evolutionary Biology 2017*, Annals of the New York Academy of Sciences

DOI: 10.1111/nyas.13571
Accepted version after peer-review.

Introduction: the infinitesimal model and its assumptions

Quantitative genetics is a conceptual and methodological framework allowing biologists to study the genetics of complex phenotypic traits,[1,2] i.e. traits influenced by a large number of genes. A central assumption of the quantitative genetics framework is that traits are normally distributed (hereafter called Gaussian traits), and when using several traits in a multivariate framework, that their relationships are simple and their number of dimensions is manageable. In this review, we will describe how new methodological development are overcoming these constraints, allowing us to study traits with non-normal error distribution (hereafter called non-Gaussian traits), complex conditional relationships between traits and high-dimensionality.

The foundations of quantitative genetics were established in 1918 by Fisher[3] using what is now known as the infinitesimal model. Apart from settling the theoretical disagreement between Mendelians and biometricians about the universality of Mendel's laws,[4] the infinitesimal model has been used for a century with much success, allowing the advent of efficient animal and plant breeding, while also deepening our understanding of evolutionary biology.[5] This success is explained not only by the simplicity of the model, but also by its robustness to various assumptions.[4] This model assumes that a complex phenotypic trait is influenced by an infinite number of loci, all of which have an infinitesimal effect. This generally results in the assumption of a Gaussian distribution of the additive genetic effects of individuals. The variance of these effects is the additive genetic variance V_A, which is a parameter of prime interest in quantitative genet-

ics. Indeed, the most common way to evaluate how a population can respond to selection is through the ratio between the additive genetic variance and the total phenotypic variance (V_P), namely the heritability (h^2):

$$h^2 = \frac{V_A}{V_P}. \quad (1)$$

To estimate V_A in modern analyses, the statistical assumptions of the infinitesimal model are best implemented using a linear mixed model (LMM), in which a random effect (with an assumed Gaussian distribution) reflecting the variance of additive genetic origin is included.[6,7] In its most simple form, this model can be written as:

$$\mathbf{z} = \mu + \mathbf{Za} + \mathbf{e}, \quad (2)$$

where $\mathbf{z}$ is the vector of observed phenotypic data, μ is the population mean phenotypic value, $\mathbf{Za}$ is a random effect of additive genetic origin (e.g. sire/dam effect or relatedness-based effect) and $\mathbf{e}$ is the vector of residuals. It is important to note that the assumption of the infinitesimal model implies that the additive genetic effects $\mathbf{a}$ (also called *breeding values*) are normally distributed following the matrix of relatedness between individuals in the population (here denoted $\mathbf{A}$):

$$\mathbf{a} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{A}V_A\right). \quad (3)$$

Since the residuals $\mathbf{e}$ are almost always assumed to be normally distributed as well (with a variance, called residual variance, noted V_R), it follows that the phenotypic trait $\mathbf{z}$ is also normally distributed.

Commonly used statistical tools in quantitative genetics are thus particularly well fitted for the study of Gaussian traits (i.e. normally distributed, at least conditionally on some fixed effects), but they are not a universal tools for any type of phenotypic trait. Curiously, despite the fact that the infinitesimal model is a century old, quantitative genetics remains very dependent on these assumptions of normality. However, in the last decade, much effort has been put into developing methods that are compatible with the core assumptions of the infinitesimal model, but that can also accommodate a broader range of characteristics for the phenotypic trait under study, especially about its distribution (but see e.g. the case of gene expression).

Expanding the model to more kinds of phenotypic traits

The threshold model

The first extension of the model towards non-Gaussian traits dates back to Wright,[8] working on the variable number (three or four, hence a binary trait) of digits in the guinea pigs. Wright refuted the idea that the complex results observed regarding the inheritance of this apparently simple phenotypic trait could be explained by an interaction between 3 different genes (the favoured hypothesis at the time) and proposed a model assuming instead a very large number of genes involved (i.e. the infinitesimal model). But, as explained above, the infinitesimal model naturally leads to a Gaussian trait and certainly not to a binary one. Wright thus assumed that, because of the mathematical consequences of the infinitesimal model, "something" (a latent trait, currently referred to as "liability", possibly through its use in quantitative genetics of disease literature) was normally distributed but that the resulting phenotype was separated into two categories: values of liability below a given threshold would produce, say, three digits while values above this threshold would produce four digits (see Fig. 1). This idea of applying a threshold to a Gaussian hypothetical trait can be traced back to even earlier work by Pearson.[9]

This model became known as the "threshold model"[10] and has been increasingly used, especially for the study of Human disease genetics,[11] because it allows to use the quantitative genetics tools on categorical traits. It was expanded to a "multiple threshold model" to account for traits with more than two categories.[12,13] The success of the threshold model has been such that it has been proposed as a catch-all model for any non-Gaussian trait, through transformation into a binary trait,[14] and its use has been promoted outside of the quantitative genetics field.[15] This success is also due to the fact that statistical implementations of the threshold model have been available for a long time.[10,13,16]

The fact that the observed phenotype is a threshold-based *transformation* of a Gaussian liability begs the question of "where" to calculate the additive genetic variance and heritability of the trait: at the level of the

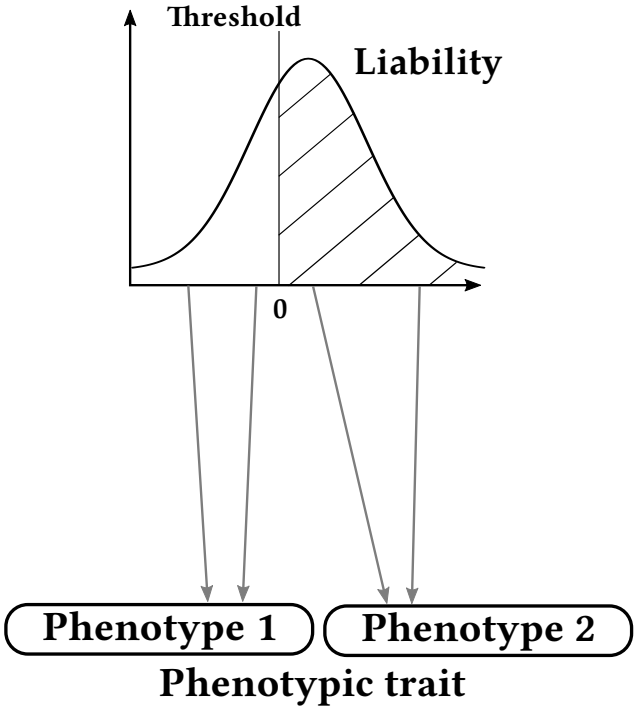


Figure 1: Graphic representation of the threshold model. Values below the threshold on the liability scale produce Phenotype 1 while values above produce Phenotype 2.

observed trait or at the level of the liability? Dempster & Lerner[10] (with help from Robertson) derived an equality to transform the heritability computed at the level of observed data h^2_{obs} to the level of the liability h^2_{liab}:

$$h^2_{\text{liab}} = \frac{p(1-p)}{t^2} h^2_{\text{obs}}, \tag{4}$$

where p is the proportion of one of the binary phenotypes in the population and t is the density of a standard normal distribution at the pth quantile. Later, van Vleck[17] found that analysing a binary trait *as if* they were Gaussian then using the transformation above, yielded well calibrated estimates of h^2_{liab} (in the case of parent-offspring regression, the two phenotypes are in comparable frequencies, i.e. p is between 0.25 and 0.75). This approach (to analyse binary data if they were Gaussian and use the transformation on the heritability estimate) has been notably promoted by Elston *et al.*[18] and Roff.[19] It has the advantage that no specific software implementation is needed, since it can basically use any tool designed for Gaussian traits, hence being extremely straightforward. To what extent this method is robust, e.g. to the presence of (fixed-effect) covariates on the model, is however unknown.

Recently, this approach has been successfully used within the animal model framework to study the heritability of binary traits such as fish life-history strategies (migrant/resident),[20] bird cooperative behaviour (helper/non-helper)[21] or dispersal status (allo-/philopatric).[22] It has also been shown to be unbiased when used with the animal model in a simulation study (see "REMLc" in Ref. 23), though, to the best of my knowledge, the sensitivity to extreme values of p has never been checked when using the animal model.

This historical evolution of the threshold model has three consequences. First, the threshold approach is straightforward to implement in practice because one can circumvent explicitly implementing the threshold model by a simple transformation of the heritability estimate. Second, and as a corollary, complex extensions of the model (e.g. multivariate models) are lacking, although they would be useful for the evolutionary biology community. This has led to the third consequence: the threshold model was "absorbed" into the generalised linear mixed models (GLMM) when they became more popular in evolutionary quantitative genetics over a decade ago. That is not to say that the threshold model has become useless. For example, Roff's[14] argument that traits with utterly *bizarre* distributions could be transformed into binary traits using a threshold and then analysed using a threshold model with reasonable success still holds, especially if these distributions do not fit within available distribution families in GLMM software.

The generalised linear mixed models

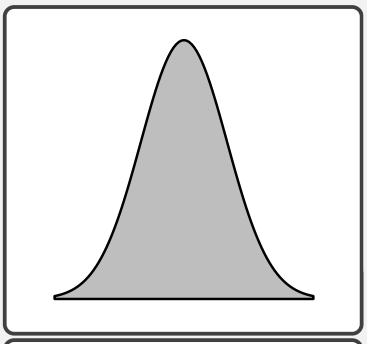
Structure of the model GLMMs are based on the same assumptions than LMM, but rather than being applied on the observed response variable, they are applied on a latent Gaussian variable (hereafter $\boldsymbol{\ell}$), much as in the threshold model. A link function g and an additional distribution $\mathcal{D}$ (with an optional parameter $\boldsymbol{\theta}$) are required to go from this latent variable to the response variable:

$$\boldsymbol{\ell} = \mu + \mathbf{Za} + \mathbf{e}, \tag{5a}$$

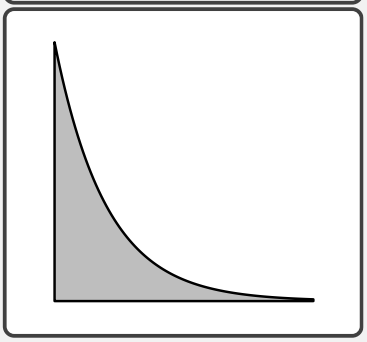
$$\boldsymbol{\eta} = g^{-1}(\boldsymbol{\ell}), \tag{5b}$$

$$\mathbf{z} \sim \mathcal{D}(\boldsymbol{\eta}, \boldsymbol{\theta}). \tag{5c}$$

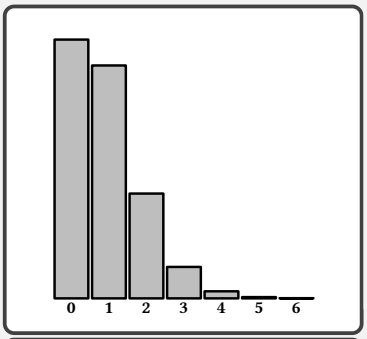
Box 1: Diversity of traits and related statistical distribution



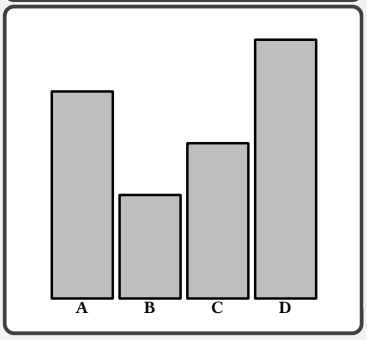
Gaussian continuous traits This type of trait might be the most common: it is definitely the most common in the published literature. Gaussian traits are readily analysed using LMMs. The residual error must be normally distributed, but not necessarily the trait itself. The influence of other variables might render the overall distribution non-Gaussian, but accounting for those variables in the model is sufficient to study the trait as a Gaussian trait. LMMs are also known to be robust to slightl deviations of the error distribution from a normal distribution.



Non-Gaussian continuous traits When the trait distribution is strongly non-Gaussian and no measured co-variate can account for this distortion from a normal distribution, relying on linear mixed models can become difficult. To analyse these traits, one option is to rely on transformations (using a function that renders the distribution of the trait closer to a normal distribution). This option raises the issue that the transformed, rather than the original trait, is analysed, hence the inferred parameters might lose their biological meaning (but they can be computed back on the original scale, see dispersal distance example in main text). If this is not possible, another option is to transform the trait into a binary trait and analyse it using a threshold model or a GLMM, as suggested by Ref. 14. A better option, if applicable, is to use a GLMM with an appropriate continuous distribution (e.g. the exponential distribution for the example on the left).



Count traits This refers to any trait measured as a positive integer. These traits are most often analysed using an appropriate GLMM (e.g. Poisson, binomial, negative-binomial), although using a LMM is possible when the distribution converges towards a Normal distribution (Poisson with high λ or binomial with high number of observations). Note that Poisson distributions are almost always used with a log-link function, meaning that the actual distribution of the trait is overdispersed compared to an actual Poisson distribution (i.e. the variance is expected to be greater than the mean).



Categorical traits This refers to traits measured as categories, such as presence/absence, morphotype or colour. The categories can be ordered (e.g. small/medium/large) in which case the traits are called ordinal. Ordinal traits are most often analysed using (multiple) threshold models (or corresponding GLMMs). Strict (non-ordered) categorical traits can be analysed using a multinomial distribution as implemented in some GLMM software implementations (see Table 1). Binary traits (with only two categories) can be seen as both categorical and ordinal, and are most often analysed using a threshold model (or corresponding GLMM).

Compared to the threshold model, GLMMs are broader, but also more complex. On the one hand, the genotype-to-phenotype model underlying the threshold model is fairly simple: some internal, invisible normally distributed variable is divided into two or several outputs using one or several thresholds. It is intuitive to imagine how such mechanisms could happen, as threshold effects are relatively common in biological systems.[24] GLMMs, on the other hand, assume that the latent variable (ℓ in Eq. 5a), which

stands for the combined additive genetic and a part of the environmental effects, is transformed into an expectation by the inverse of the link function ($\boldsymbol{\eta}$ in Eq. 5b). Observed values $\mathbf{z}$ (Eq. 5c) are then drawn around this expectation according to a certain distribution $\mathcal{D}$. Strictly speaking, this distribution should belong to the exponential family, like the binomial or Poisson distributions, but in practice, the scope of software implementing GLMMs has gone beyond this family, providing distributions such as negative binomial and zero-inflated Poisson. It is thus important to note that GLMMs are not LMMs "with a different error distribution", but are a much more complex type of model. The most commonly used distributions are the binomial and Poisson distributions to analyse count data such as binary traits, clutch size or fitness. However, the variety of distributions available in software implementing GLMMs is growing, with distributions such as the negative-binomial (for overdispersed Poisson-like data), zero-inflated Poisson (a mixture of a binomial and a Poisson distribution), multinomial or exponential (see Box 1 and Table 1). The availability of a wide variety of distributions is relevant for evolutionary biologists, as most evolutionary relevant traits have complex distributions. Such models have been used to study the genetics of various traits such as dispersal,[25] number of offspring,[26,27] disease,[28] resistance[29] and productivity.[30]

Estimating the parameters The likelihood of a GLMM is most often intractable, and specific approximate algorithms have been developed.[33] Each of these algorithms have strengths and weaknesses in terms of speed, accuracy and robustness to model complexity. For example, the penalised quasi-likelihood (PQL) is a fast algorithm, but is known to be biased in some cases.[34] In the context of quantitative genetics analysis, this has been shown to yield downward-biased estimates of additive genetic variance, and hence of heritabilities for binomial data.[23] On the contrary, Bayesian algorithms such as Markov Chain Monte Carlo (MCMC) usually yield more accurate variance estimates, though a sensitivity to the prior distribution is to be expected when the data sample size and the variance are low.[23] When the fitted models are highly complex (e.g. a zero-inflated Poisson with many random effects), likelihood-based methods might fail to converge at all, and only MCMC-based methods might yield an output. A more comprehensive review of algorithms to estimate GLMMs and their properties can be found in Box 2 of Ref. 35.

Computing quantitative genetic parameters from a GLMM

Choosing the scale From Eq. 5, we can note that GLMMs are composed of three layers, or "scales", instead of one for the LMM. This raises the following issue: on which scale should the heritability be computed? Parameters such as the additive genetic variance and heritability are easily computed for the latent variable ($\boldsymbol{\ell}$ in Eq. 5a, hereafter "latent scale"), which is by definition linear, Gaussian and behave according to the classical assumptions of the infinitesimal model, but obtaining those parameters on the scales of $\boldsymbol{\eta}$ in Eq. 5b and $\mathbf{z}$ in 5c (hereafter "data scale") is more complex.[36,37] However a case can be made that inferring quantitative parameters on the data scale should be the standard approach as the phenotype is exposed to natural selection at this scale. In any case, the relationships between these scales bear biological assumptions that researchers need to be aware of (see Box 2). Two main assumptions have notable biological consequences in terms of quantitative genetics. The first one is that the whole model is not linear (notably because of the impact of the non-linear link function). This, for example, generates non-additive genetic variance on the data scale even if only additive genetic variance was assumed on the latent scale. In other words, broad- and narrow-sense heritabilities are always differently computed on the data scale, even if no non-additive genetic component was assumed. The second assumption is that an always-existing intrinsic level of "noise" is assumed in almost all models. This level of noise is dependent on the mean and cannot be set to zero as in LMMs. For example, setting V_R do not result in a heritability of one on the data scale. In fact, for a given latent mean and total variance, there is a maximal value that the data scale heritability can reach that can be strictly lower than one.

Using a "link variance" To obtain parameters on the data scale, a parallel with the "liability scale" of the threshold model has been most commonly used. Indeed, it can be shown that a GLMM with a binomial distribution and a probit link function is equivalent to a threshold model,[31,38] because a probit function is the inverse function of a cumulative distribution function of a standard Normal distribution. In other

Table 1: List of R packages available to implement generalised linear mixed models (GLMMs) with some of their characteristics.

Software	Licensing	Algorithm	Available distributions	Accepts pedigree
lme4	GPL	Gauss-Hermite	binomial, gamma, Poisson, inverse-Gaussian	No
ASReml-R	Proprietary	PQL	binomial, multinomial, gamma, Poisson, negative-binomial	Yes
MCMCglmm	GPL	MCMC	binomial, multinomial[1], Poisson, exponential, geometric, zero-inflated Poisson and binomial	Yes
animalINLA	GPL	INLA	binomial, Poisson, zero-inflated Poisson	Yes

[1]: The models named "ordinal" and "categorical" can be broadly viewed as multinomial models with various properties of the link function and underlying latent scale. GPL: GNU Public Licence; PQL: Penalised Quasi-Likelihood; MCMC: Markov Chain Monte Carlo; INLA: Integrated Nested Laplace Approximation.

words, defining the probability $p = \text{probit}^{-1}(\ell)$ in a binomial GLMM is equivalent to adding a value drawn in a standard Normal distribution to ℓ and then applying a threshold. The conclusion is that the variance on a hypothetical liability of an equivalent threshold model is the variance of ℓ (e.g. $V_A + V_R$), plus the variance of a standard Normal distribution, which is equal to one and can be referred to as the "link variance". Note also that V_R is not identifiable in binomial models, and should thus be set to an arbitrary value (usually zero, more rarely one). The heritability, as if a threshold model was used to analyse the data and heritability was computed on the liability scale, is thus:[36]

$$h^2_{\text{liab}} = \frac{V_A}{V_A + V_R + 1}. \tag{6}$$

From this estimate, the heritability on the data scale can be easily obtained using Eq. 4. Recognising the logit function as the inverse of the cumulative distribution function of a logistic distribution leads to the same conclusion that a term needs to be added to obtain the variance on a liability scale (here a mixture of a Gaussian and logistic distributions). This term, corresponding to the variance of a standard logistic distribution, is $\pi^2/3$. This approach is very useful, as it allows heritability estimates to be compared on a common scale, whether they are estimated using a binomial GLMM or a threshold model. As such, it has become a standard to compute heritability estimates from binomial GLMMs.[37] Because of the simplicity of this approach, a similar reasoning was used as an approximation for Poisson GLMMs,[26,27,39] despite the logarithm link function not being related to a distribution function, contrary to the probit and logit functions. This usually leads to the following approximation:[39]

$$h^2_{\text{approxPoiss}} = \frac{V_A}{V_A + V_R + \lambda^{-1}}, \tag{7}$$

where λ is the population average of the phenotypic values. However, it is important to note that this calculation has no strong theoretical justification and the scale upon which it is performed is undefined. The actual heritability on the data scale is:[39]

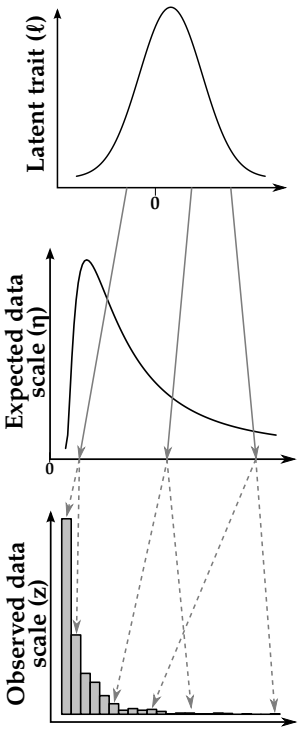
$$h^2_{\text{Poiss}} = \frac{V_A}{\exp(V_A + V_R) - 1 + \lambda^{-1}}, \tag{8}$$

which should generally be preferred. Note that Eqs. 7 and 8 converge as V_P tends toward zero. Note also that this is different from the estimate of the repeatability given by Ref. 36, because repeatability corresponds to an intra-class correlation coefficient, which in terms of heritability would be a broad-sense heritability[31,39] (see Box 2 as for why those two estimates would be different).

Using a general framework In general, this approach of adding a "link variance" is very specific to models related to the threshold model (e.g. binomial with a probit or logit link), but not applicable to all GLMMs. Ref 31 provides a more general solution which involves integral calculation. Such integrals allow to compute population mean ($\bar{\mathbf{z}}$) and variance ($V_{P,\mathbf{z}}$) on the data scale (i.e. related to $\mathbf{z}$), based on latent scale estimates (i.e. related to $\boldsymbol{\ell}$). Furthermore, the additive genetic variance on the data scale can be computed as the product of the latent additive genetic

Box 2: Biological assumptions behind GLMMs and their consequences

When using a statistical model in quantitative genetics, we are always making assumptions about the underlying genetics and the genotype-phenotype map of the phenotypic trait of interest. When using a LMM, these assumptions have a clear connection with the infinitesimal model and the biological assumptions are clear and mostly well-understood by biologists. However, as shown in Eq. 5, GLMMs are more complex models and it is important for biologists to be aware that this involves specific assumptions.



Distribution of the different scales under Poisson/log GLMM assumptions. Solid lines: deterministic relationships; dashed lines: stochastic relationships. Modified from Ref. 31.

As stated in the main text, the latent scale ℓ is the one on which assumptions are compatible with the infinitesimal model. This is illustrated in the figure on the left as the latent trait follows a Gaussian distribution. But looking at the expected (η) and observed data (z) scale, we can see they greatly differ from what would be expected under the sole infinitesimal model assumptions.

Assume that all of the variance in the latent scale is of additive genetic origin. The heritability on that scale is $h^2_{\text{lat}} = 1$. Because of the non-linearity of the inverse of the link function (i.e. the exponential function in this example), a part of this additive genetic variance is not additive any more on the expected data scale. Hence, we assume that the additive genetic components of the trait are interacting in a complex way (e.g. with epistatic interaction) to yield the genetic expectation of the trait. In other words, the variance of the expected data scale is still entirely of genetic origin (broad-sense heritability $H^2_{\text{exp}} = 1$), but now only a part of this variance is additive (narrow-sense heritability $h^2_{\text{exp}} < 1$). To go from the expected to the observed data scale, we sample phenotypic values around the expectation (here according to a Poisson distribution). This creates more variance of non-genetic origin. At this scale, neither the broad- nor the narrow sense heritabilities are equal to 1. This added "noise" around the expected values can be considered as either of biological origin or a measurement error.

This example shows two important consequences of the assumptions in GLMMs. First, the implicitly assumed genotype-phenotype map is complex and generates non-additive genetic variances even if none was included in the statistical model for the latent trait. Second, the models assumes that the trait is subject to an incompressible, intrinsic environmental noise that neither artificial, nor natural selection could remove. Whether this is a sensible assumption for the analysis at hand should be considered by biologists with caution.

Unfortunately, in-depth studies regarding the biological validity of these assumptions are currently lacking (but see Ref. 32 for more general considerations).

variance and a coefficient Ψ^2, which can be computed using integral calculation:[31,32]

$$V_{\text{A},\mathbf{z}} = \Psi^2 V_{\text{A}}. \tag{9}$$

Once these parameters are computed, the heritability can simply be estimated as:

$$h^2_{\text{obs}} = \frac{V_{\text{A},\mathbf{z}}}{V_{\text{P},\mathbf{z}}}. \tag{10}$$

Applying Eq. 10 to a Poisson GLMM exactly yields Eq. 8. That is not to say the estimate of h^2_{obs} in Eq. 10 is in every sense identical to an estimate obtained using a LMM. Indeed a strong issue when assuming the model underlying GLMMs (Eq. 5) is that it breaks the simple relationship between heritability, the strength of selection S and the response to selection R, as usually illustrated by the breeder's equation:[40] $R = h^2 S$. This equation holds only approximatively in the context of GLMMs and the heritability computed on the data scale.[31] A non-approximate strategy consists going back and forth between the observed data scale and the latent scale. First the fitness from observed data has to be computed, then the above equation (or

Price-Robertson identity[41,42]) has to be applied on the latent scale to predict the *latent* response to selection and finally go back to the data scale using the integral-based framework described above to obtain the corresponding phenotypic response. Yet another issue with GLMMs, again arising from the non-linearity of the link function is that fixed effects, when included in a GLMM impact the quantitative genetic estimates on the data scale. To account for this, it is necessary to integrate over the fixed effects. This whole new approach to compute quantitative genetic parameters and response to selection from GLMM-based analyses has been implemented in an R package named QGglmm.[43]

Multivariate analysis Multivariate models are important in quantitative genetics as genetic and environmental constraints on the covariances of traits have a strong impact on their ability to respond to selection.[44] An interesting feature of the above framework is that it can be used in a multivariate form. Multivariate GLMMs are handy in the sense that they allow to study the genetic or environmental covariation between traits with very different distributions (e.g. between a Gaussian, a binary and a Poisson trait). To do so, assuming T traits are studied, such models assume that the latent scale is of T latent traits, with e.g. a **G**-matrix of additive genetic variance-covariance and a **R**-matrix of residual variance-covariance between latent traits (similarly to multivariate LMMs). These latent traits are then independently transformed to their respective data scale according the specific process of their distribution (a log-link and a Poisson error for a Poisson trait, for example). This model yields a **G**-matrix on the latent scale, and given the possibly strong differences between the traits distribution, we might expect to run into trouble when trying to compute a **G**-matrix on the data scale. This is however not the case, because the transformation of each latent to its respective data scale is independent from each other. Thus a diagonal matrix $\mathbf{\Psi}$ containing an evaluation of each Ψ_t of the trait t in the diagonal can easily be computed. The data scale $\mathbf{G}_{\text{obs}}$ is then:

$$\mathbf{G}_{\text{obs}} = \mathbf{\Psi}\mathbf{G}\mathbf{\Psi}' \tag{11}$$

where $'$ is the transposition operator. $\mathbf{G}_{\text{obs}}$ can be directly interpreted as the additive genetic variance-covariance matrix on the data scale. Dividing the diagonal elements of this matrix by the diagonal elements of the data-scale phenotypic variance-covariance matrix ($\mathbf{P}_{\text{obs}}$, also obtainable from the estimations of the multivariate GLMMs) provides the data-scale heritabilities of each trait. Additionally, converting $\mathbf{G}_{\text{obs}}$ into a correlation matrix provides the data-scale genetic correlations between the traits. This approach has also been implemented in the QGglmm R package.[43]

An example study: the phoenix dataset

The dataset To illustrate some of the points stated in the previous section, this section will analyse a simplistic and simulated dataset (simulation code available online). This dataset concerns phenotypic traits measured on a mythical bird: the phoenix. Phoenixes are eagle-sized, non sexually dimorphic birds with a golden plumage and a capacity to revive from their ashes when they die. The dataset consists of a population of 1000 pedigreed individuals on which I have one measurement *per* trait. The traits were analysed using an animal model and the MCMCglmm R package.[45] The R code, in the form of a tutorial, as well as the data used here, are available online.

Tarsus length As is the case for other birds, tarsus length is a Gaussian (i.e. normally distributed) trait (Fig. 2A). The animal model output yields a relatively high heritability of 0.34 for this trait (Table 2). Because this is a Gaussian trait, there is no debate around the scale on which to compute the heritability, as there is only one way to produce an estimate.

Dispersal distance Phoenixes disperse from their site of birth to a site of breeding. The distribution of natal dispersal distances is quite typical,[46] with most birds dispersing only short distances, but occasionally over much longer distances (Fig. 2). Despite the continuous aspect of natal dispersal distance, this trait is thus heavily non-Gaussian. However, the log- transformed distance is normally distributed and can thus be analysed as a Gaussian trait:

$$\text{log-distance} = \log(\text{distance}) \tag{12}$$

The output of an analysis of the log-distance using an animal model is shown in Table 2. Using the output of this model directly, the heritability of the log-distance is equal 0.64 (Table 2). However, what I am interested

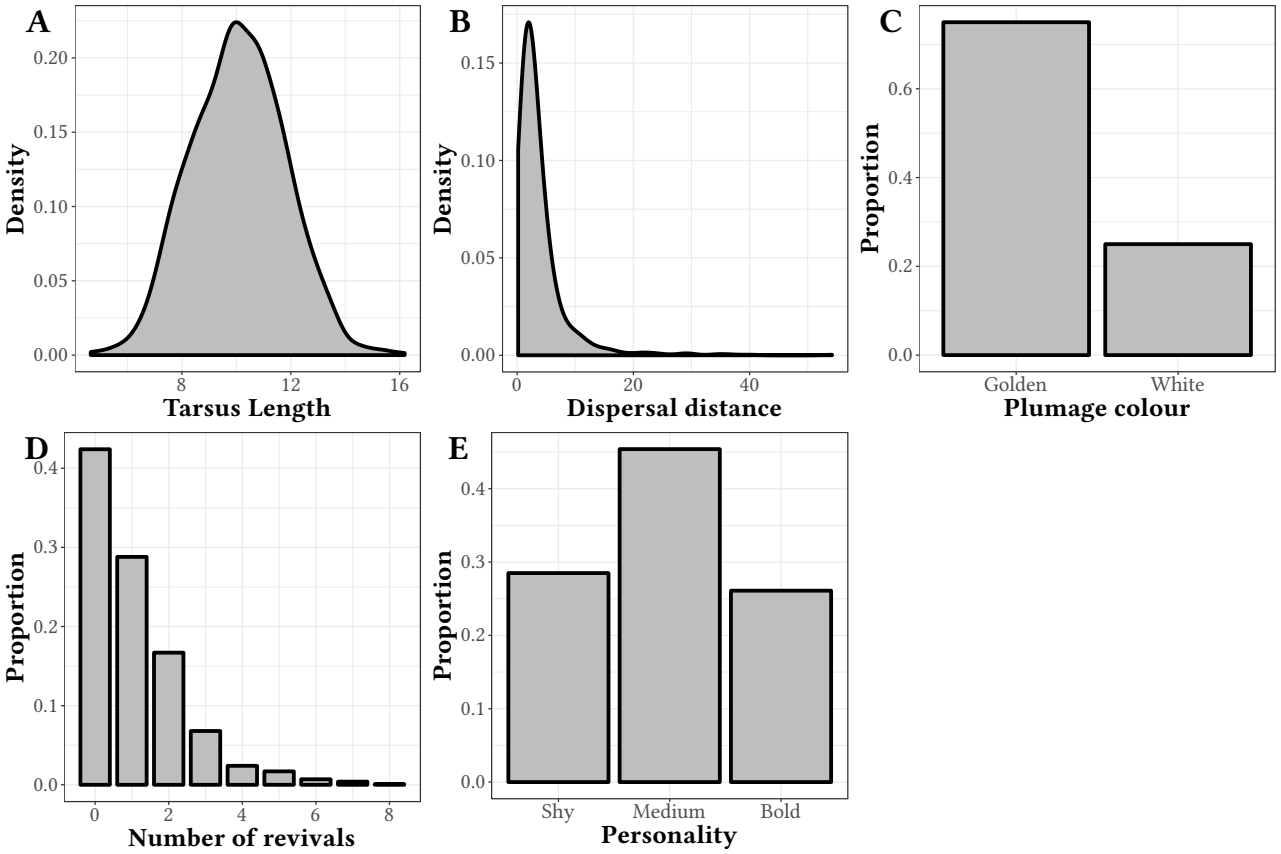


Figure 2: Trait distribution of the phoenix example. Traits for which a density is displayed (tarsus length and dispersal distance) are continuous, traits for which a histogram is shown (plumage colour, number of revivals and personality) are discrete, i.e. count or categorical traits.

Table 2: Point estimates of the (latent) intercept (μ), (latent) additive genetic (V_A) and (latent) residual variance (V_R) for four traits of the phoenix dataset. The latent heritability h^2_{lat} is computed as the ratio of V_A to the sum of V_A and V_R. The data scale heritability h^2_{obs} is computed using the QGglmm package, when relevant.

Trait	Distribution	μ	V_A	V_R	h^2_{lat}	h^2_{obs}
Tarsus Length	Gaussian	10.13	0.96	2.02	0.34	
Dispersal distance	log-Gaussian	0.84	0.45	0.25	0.64	0.44
Plumage colour	Binary	-1.1	1.86	1	0.69	0.27
Nb. of revivals	log-Poisson	-0.14	0.13	0.33	0.24	0.086

in is the heritability of the distance of natal dispersal, rather than the log-distance, as this is the quantity that might be targeted by selection. Although this type of data transformation is by no means a GLMM *per se*, it is possible to use the framework from Ref. 31 and the package QGglmm[43] to compute the heritability back on the original data scale. We can do this by considering the logarithm as a link function, but no distribution $\mathcal{D}$ (essentially $\mathbf{z} = \boldsymbol{\eta}$ in Eq. 5), see the script in Supplementary Information for details. The heritability on the original data scale is 0.44 (Table 2), which is much lower than the heritability of the log-distance. Both estimates are correct, they simply do not relate to the same quantity. For information, an (ill-fitted) model using the distance directly as a response variable (thus assuming it is a Gaussian trait) yields an intermediate heritability of 0.55.

Plumage colour Phoenixes usually have a golden plumage, but about 25% of them have a white plumage (Fig 2C). This might be of genetic origin, but no simple Mendelian model can account for the heredity of this binary trait. It is possible to analyse this trait using quantitative genetics by assuming a threshold model. To implement such a threshold model, I chose the approach of Refs 17–19 to analyse the binary trait as if it were normally distributed. This produced a heritability estimate of 0.30, which can be transformed on the liability scale using Dempster & Lerner's equation[10] (see Eq. 4). This yields a heritability estimate of 0.56. Again, both of those estimates are correct: one (0.30) is related to the actual data scale, while the other (0.56) is related to the hypothetical, continuous liability scale. I also used a more proper binomial GLMM with a probit link to analyse the trait. I obtained a very high heritability of 0.69 on the latent scale (see Table 2). However, this calculation does not account for the "link variance" and thus doesn't bear actual connection with the threshold model (see Eq. 6). When accounting for this, I obtained a lower heritability (0.51), very close to the one on the liability scale, as expected. When estimations from this GLMM are transformed through QGglmm to obtain estimates on the data scale, we obtain 0.27, again very close to the estimate on the data scale. Because GLMMs are more properly defined than the Gaussian-model approach, and since we can obtain comparable parameters (i.e. pertaining to the same scales) in both cases, I recommend using GLMMs whenever possible.

Number of revivals All phoenixes do not revive from their ashes, most of them do not. Those who do can only revive a given number of times before permanently dying, with only a few birds reaching as many as 8 revivals (Fig 2D). I analysed this trait using a Poisson GLMM with a logarithm link function. When using estimates to compute the heritability on the latent scale, I obtained an estimate of 0.24 (Table 2). Using QGglmm to compute the heritability on the data scale yields an estimate of 0.086 (Table 2), exactly the same as using Eq. 8. Using the simpler approximation of Eq. 7 yields a close estimate of 0.093. The approximation works fairly well in this case, as the estimated total latent variance (0.45) is low.

A multivariate model To illustrate the concepts related to multivariate GLMMs, I will analyse a joint model of the number of revivals (a Poisson trait) and plumage colour (a binary trait with, as above, the residual variance set to one). This model assumes two latent traits (possibly) with additive genetic and residual covariances. The model thus provides a two-elements vector for the intercepts and matrices of dimensions 2×2 for the additive genetic **G**-matrix and the residual **R**-matrix:

$$\begin{aligned} \mu &= \begin{pmatrix} -0.13 \\ -1.19 \end{pmatrix}, \\ \mathbf{G} &= \begin{pmatrix} 0.088 & 0.0042 \\ 0.0042 & 1.99 \end{pmatrix}, \\ \mathbf{R} &= \begin{pmatrix} 0.34 & 0.055 \\ 0.055 & 1 \end{pmatrix}, \\ \mathbf{P} &= \mathbf{G} + \mathbf{R}. \end{aligned} \tag{13}$$

The analysis of these estimates is already quite interesting: on the latent scale, the estimates of the additive genetic and residual covariances are quite low, suggesting that the two latent traits are in fact relatively independent. By dividing the diagonal elements of **G** by the diagonal elements of **P**, we obtain latent heritabilities (0.24 and 0.68) matching those of the univariate analysis in this case. Those latent estimates can be transformed to the data scale using QGglmm, yielding:

$$\begin{aligned} \text{mean}_{\text{obs}} &= \begin{pmatrix} 1.12 \\ 0.28 \end{pmatrix}, \\ \mathbf{G}_{\text{obs}} &= \begin{pmatrix} 0.11 & 0.00079 \\ 0.00079 & 0.056 \end{pmatrix}, \\ \mathbf{P}_{\text{obs}} &= \begin{pmatrix} 1.90 & 0.016 \\ 0.016 & 0.20 \end{pmatrix}. \end{aligned} \tag{14}$$

This confirms that on the data scale, the two traits seem to be quite uncorrelated. Dividing the diagonal elements of $\mathbf{G}_{\text{obs}}$ by the diagonal elements of $\mathbf{P}_{\text{obs}}$ again yields heritabilities comparable to the univariate cases (0.058 and 0.28).

The special case of non-binary categorical traits

Multiple threshold model and equivalents Just like the number of digits in Wright's study,[8] the analysis of many categorical traits (i.e. traits that are composed of categories such as blue/green/red) requires quantitative genetic tools, because their underlying genetics is complex. Such an analysis for non-binary categorical traits (i.e. with more than two categories)

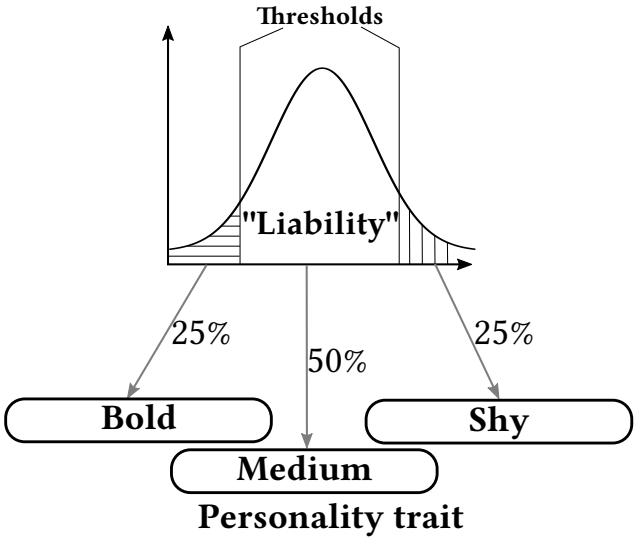


Figure 3: Graphical representation of the multiple threshold model used to simulate the personality trait of the phoenix dataset. More than one threshold is estimated, separating the underlying "liability" into the three phenotypes (Bold, Medium and Shy personality). The thresholds were set to be symmetric around the mean, yielding proportions of roughly 25%/50%/25%.

can be performed using the multiple threshold model introduced above or using a GLMM (e.g. using multinomial model, see Table 1). In most methods, the latent trait is divided into different groups according to various thresholds and then either directly transformed into categorical data (for the multiple threshold model, see Fig. 3) or, equivalently, after going through a link function and distribution (in the case of a GLMM). For example, Ref. 47 used this type of model on humans to study the genetic covariation between choroidal thickness and age-related macular degeneration (AMD), computed as different stages of disease seriousness (0 to 5).

Categorical or ordinal data It is important to distinguish between categorical traits in which categories cannot be ordered (such as colours) and ordinal traits in which categories can be naturally ordered (such as binned data). Strictly speaking, the multiple threshold model is designed for ordinal traits, because the categories need to be ordered (see Fig. 3). For purely (unordered) categorical data, models such as the "categorical" family in MCMCglmm[48] are available. Instead of assuming a one-dimensional latent scale as in the multiple threshold or its GLMM equivalent, these models assume a multi-dimensional latent scale: for a trait with K possible categories, this scale is composed of a $K-1$ dimensional Gaussian distribution. These models can be very useful as they allow to infer interesting information about categorical traits, such as the genetic correlation between the different categories.

Estimating the heritability Estimating the heritability on the observed data scale for these traits is relatively complicated. Despite the apparent simplicity of categorical and ordinal traits, they take the mathematical form of multivariate traits, with the number of dimensions equal to the number of traits. On the observed data scale (Eq. 5c), there is not one heritability, but a variance-covariance matrix providing the additive genetic variance for each category in the diagonal and the additive genetic covariances in the off-diagonal elements. I found no mention of this anywhere in the literature and the properties of this matrix (and of the corresponding observed data scale phenotypic variance covariance matrix) have not been properly described for two possible reasons. First, by far the most popular model to analyse categorical and ordinal traits is the multiple threshold model. Because of the deterministic relationship between the liability scale and the observed data scale (see Fig. 3), the heritability computed on the liability scale serves as a good summary statistic of the additive genetic properties of the trait (see example below). Second, a general framework to compute the quantitative genetic parameters of models other than the multiple threshold (e.g. multinomial model) has been only very recently available.[31]

Calculations for ordinal traits As an illustration, I applied this framework to the ordinal family available in the MCMCglmm R package (see Supplementary Information for the calculation details). The population average on the data scale is simply the proportion p_k of each category k (K categories):

$$\bar{\mathbf{z}} = \begin{pmatrix} p_1 \\ \vdots \\ p_K \end{pmatrix}. \tag{15}$$

The phenotypic variance-covariance matrix $\mathbf{P}_z$ is defined as:

$$P_z(k,l) = \begin{cases} p_k(1-p_k) & \text{if } k = l \\ -p_k p_l & \text{if } k \neq l \end{cases}. \tag{16}$$

Table 3: Point estimates of the (latent) intercept (μ), (latent) additive genetic (V_A) and (latent) residual variance (V_R) for the personality trait of the phoenix dataset. The latent heritability h^2_{lat} is computed as the ratio of V_A to the sum of V_A and V_R. The data scale heritability h^2_{obs} are computed from QGglmm, which provides an estimate for each category (Bold, Medium, Shy).

μ	V_A	V_R	h^2_{lat}	h^2_{obs} (Shy)	h^2_{obs} (Medium)	h^2_{obs} (Bold)
0.94	0.55	1	0.4	0.12	9.7×10^{-5}	0.12

More interestingly, one can define the vector $\mathbf{\Psi}$ (note that here $\mathbf{\Psi}$ is a vector, not a matrix as in the multivariate case above) as the difference of the liability density at consecutive thresholds, here noted γ (see Supplementary Information for the proof of this):

$$\Psi_k = f(\gamma_{k-1}) - f(\gamma_k), \tag{17}$$

where $\gamma_0 = -\infty$ and f is the density of a Normal distribution with mean μ and variance $V(\ell) + 1$ (i.e. the density of the liability). From there (see also the multivariate paragraph above), we can conclude that the additive genetic matrix on the data scale, $\mathbf{G}_z$ is:

$$\mathbf{G}_z = \mathbf{\Psi} V_A \mathbf{\Psi}^T. \tag{18}$$

By dividing the diagonal elements of $\mathbf{G}_z$ by the diagonal element of $\mathbf{P}_z$, we obtain a heritability estimate for each category. Beside the technicality of this calculation, we can learn the following properties. First, phenotypic covariances between categories are always negative. This comes from the fact that when a category increases in proportion in a population, the proportion of other categories necessarily decrease. Second, the sign of the additive genetic covariances between a category k and l will directly depend on the signs of Ψ_k and Ψ_l. Yet the signs of these quantities depend on whether we are "climbing" the Gaussian density (i.e. we are below its mean value) or "descending" it (i.e. were are above its mean value). As a result, $\mathbf{G}_z$ is composed of blocks: two groups of categories for which the threshold is strictly below (or above for the other group) the mean value are positively correlated among them, while the two groups are negatively correlated with each other. The biological meaning of this is that extreme categories (like the lowest or highest ranking categories) have phenotypic values far away from the population mean, hence on average breeding values far away from 0. As a consequence, they have a strong tendency to yield offspring belonging to "neighbouring" categories, hence the genetic correlation. Third, the $\mathbf{G}_z$ variance-covariance matrix is actually of rank 1 and all additive genetic correlations are either -1 or 1 (following the block structure described above for the signs). So even if the genetic structure takes the form of a matrix of dimension K, its actual dimensionality is the same as the liability scale it comes from (i.e. 1). The dimensionality of $\mathbf{P}_z$, however, is $K - 1$. This relates to the fact that a categorical trait with K categories has a dimensionality of $K - 1$ due to the constraint of the proportion of each category summing over to one. This also explains why binary traits can be summarised using a single heritability value: given that for $K = 2$, both $\mathbf{G}_z$ and $\mathbf{P}_z$ have a dimensionality of 1, all estimates (additive genetic and phenotypic variances and the resulting heritabilities) are equal between the two categories. In any case, this whole multivariate framework is actually not needed for $K = 2$, as shown above.

An example: personality of phoenixes When the phoenixes of our example are being caught, a personality trait is assessed based on their behaviour. Responsive and aggressive individuals are categorised as "bold", calmer individuals are categorised as "medium" and afraid or panicked individuals are categorised as "shy" (see Fig. 2E). Because of the clear and natural hierarchy of these three categories, this trait can be seen as an ordinal trait. Using the "ordinal" family (implementing a model equivalent to a multiple threshold model, see also the "threshold" family) of MCMCglmm,[45] we can infer latent parameters and hence a latent heritability of 0.40 (Table 3). As usual, we need to add the "link variance" to obtain an estimate actually comparable to the liability scale of a multiple threshold model. This yields an lower estimate of 0.23. Using QGglmm to obtain parameters on the data scale, we obtain a heritability parameter for each category: Bold and Shy phenotype have a 0.12 heritability, while the Medium category has an extremely low heritability (9.7×10^{-5}, see Table 3). These heritabilities are computed by dividing the diagonal elements of the $\mathbf{G}$ matrix by the diagonal elements of

the **P** matrix on the data scale. The distribution of these estimates are interesting and informative. Because of the 25%/50%/25% symmetry of the Bold/Medium/Shy categories (see Fig. 2E), individuals from the Medium category (i.e. close to the liability mean) exhibit little genetic variation compared to Bold and Shy individuals (all necessarily more distant to the liability mean). As always, all of those estimates are correct, but which one is the most sensible to comment on? Excluding the data scale for now, between the latent and liability scale, the latter should be preferred for two reasons: *(i)* this is the heritability that has historically been reported when using (multiple) threshold models, and *(ii)* this is the "closest" to the data, as the liability is linked to the data scale through a deterministic relationship. However, one could argue that the best estimates are data scale ones, as they are linked to the actual phenotype. This is a generally sensible reasoning. However, in the case of our personality trait (and in many other situations), the Bold/Medium/Shy division is a binning caused by our inability to finely measure continuous variation on a Bold-Shy continuum of personalities, thus that the actual phenotype is the (not so hypothetical, in this case) liability scale. Note that the same reasoning would apply if we had only a Bold/Shy categorisation (hence for a binary trait).

Cutting-edge and complex models for specific phenotypic traits

Heritability of survival: combining animal and capture-recapture models

Away from experimentally controlled settings, survival can be a challenging phenotypic trait on which to perform quantitative genetics. In many settings, and especially in wild populations, the survival of individuals is not directly observed. Rather, it is inferred from the prolonged absence of the individual from the records (which can also be due to emigration or coarse recapture efforts). As with any kind of inference, this is performed with uncertainty: from the onset of the absence of the individual from the records, one cannot tell when exactly the death happened, or if it happened at all. In practice, this uncertainty will depend on the depth of the population survey and the survey's ability to recapture existing individuals. Capture-recapture models[49–51] were designed to infer parameters such as survival while accounting for imperfect probability of recapture.

Estimating the heritability of survival while ignoring the uncertainty around its estimation might have two consequences. The first and immediate consequence is in confidence/credibility intervals around the heritability estimate that are narrower than they ought to be. The second consequence is a potentially biased estimation. Indeed many processes might lead to a better inference of survival in some individuals than in others (e.g. bold individuals are easier to catch than shy individuals). The use of statistical models (such as capture-recapture models above) and auxiliary variables should help to mitigate this bias.

To address this issue, Papaïx *et al.*[52] used the flexibility of Bayesian inference and MCMC algorithms[53] to estimate the heritability of survival, using an animal model, while "properly" estimating survival using a capture-recapture model (an approached named CRAM for Capture-Recapture-Animal-Model). The principle is as follows: a capture-recapture "module" estimates survival from one year to another using time-series of presence/absence and optional auxiliary variables whilst an animal model "module" makes use of these inferences to estimate the heritability of survival. Because both inferences are conducted within the same MCMC run, this approach naturally accounts for the uncertainty in estimating survival and should thus yield better credible intervals.

Unfortunately, this approach has not gained momentum and has, to the best of my knowledge, never been applied apart from the study of Papaïx *et al.*.[52] This lack of success might stem from two issues. A first issue could be that the number of available datasets for which this approach could be used on (i.e. a wild population survey for which a pedigree is available and with imperfect probability of recapture) is relatively small. A second issue could be the relative complexity of the approach. A WinBUGS/OpenBUGS code is available, but the method requires handling of large sparse matrices, for which these programs are not optimised. The use of JAGS and its `glm` module,[54] which should handle sparse matrices better, could help in that regard.

Despite the unfortunate (but still reversible) fate of CRAM, it led the way into an exciting approach using hierarchical modelling to perform quantitative ge-

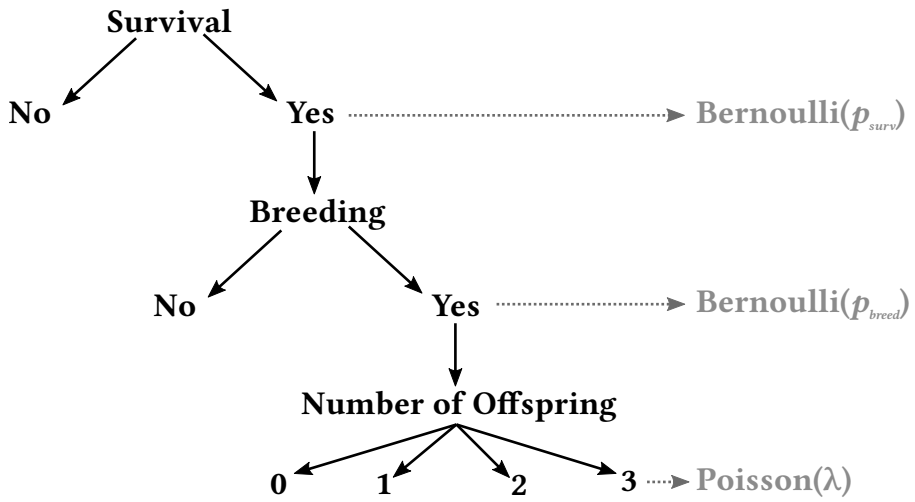


Figure 4: Example of a process that can be analysed using aster models. An individual survives according to a Bernoulli distribution with a probability p_{surv}. If an individual survives, it breeds with a probability p_{breed}. Finally, if the individual breeds, it yields a number of offspring following a Poisson distribution with a parameter λ.

netics on statistically inferred (rather than directly observed) phenotypic traits by combining "modules" in a hierarchical model, as suggested by Ref. 52 and 37. We can hope that beyond the special case of survival, this modular approach could be useful in other areas of quantitative genetics.

Compounded life-history traits: aster models

The relationship between components of fitness and life history traits is notoriously complex, with many traits influencing the total fitness of individuals at the same time. Often, there is a hierarchy in the influence of components of fitness. For example (Fig. 4), individuals have first to survive the non-breeding season, then engage in mating during the breeding season and finally have different numbers of offspring (a.k.a. breeding success). This kind of data is currently most often analysed as a zero-inflated Poisson distribution using directly the breeding success (and assigning a breeding success of zero to dead or non-breeding individuals). This is a correct approach, but a frustrating one as it does not allow all the steps to be studied separately. In our example, survival and not entering into breeding are merged when using a zero-inflated Poisson model, while being completely different phenomena. As these two phenomena could have very different genetic and environmental influences, much could be gained by modelling them separately rather than combining them in the same component of a bivariate analysis (as is the case in a zero-inflated Poisson).

In order to study this type of conditional relationship between e.g. life-history traits,[55] Geyer *et al.*[56] developed an approach named *aster models*, in which these hierarchical dependencies typical of life-history traits are accounted for. The approach was implemented in the `aster` R package. Since then, aster models have been used in various studies on adaptation, especially on plants.[57–61] Recently, aster models have been extended to include random effects[62] using a modified version of the PQL approach.[33] These effects can be used to estimate the additive genetic variance in some settings, e.g. the estimation of heritability of compounded life-history traits or fitness using sire/dam designs (see a tutorial in Ref. 63). Although no implementation is readily available, the aster models can be extended to use a pedigree. However the theoretical implications of aster models in terms of quantitative genetics (as was done for GLMMs in Ref. 31) have not yet been worked out. Hence some methodological challenges are still ahead, but aster models have a strong potential for the study of the quantitative genetics of life-history traits.

More generally, the core assumptions behind aster models (assumption of a causality pathway and conditional independence between components of this pathway) can already be used with much success in the complex analysis of fitness components. For example, Ref. 64 has shown that the "paradox of stasis"

in bird body size (where there is seemingly strong selection for higher body sizes but no change is observed over time) is likely to be explained by an antagonistic effect of chick body size on parental performance. To do so, they assumed a complex pathway to explain fitness, including among other variables, parental performance, body size, survival and fecundity. Assuming conditional independence between three models (for body size, juvenile survival and annual fecundity), they could fit these models using "plain" GLMMs. This approach is made possible by the computation of selection gradients that are complex functions of the models estimates and the use of a Bayesian approach allowing to compute posterior distributions (hence to obtain a measure of the statistical error) of such complex functions.[64] Although this approach is not an exact equivalent to aster models (which is using joint estimation and can come in a "conditional aster models"[55,56]), it could be a possible way to tackle issues such as the one in Fig. 4 (using subsequent Bayesian models including the parent node as a fixed effect and computing the complex functions for the resulting selection gradients).

High dimension data: the heritability of gene expression

Gene expression is a phenotypic trait of great interest because of its obvious causal proximity to the genes, allowing for detailed studies of the genotype-phenotype map.[65] Although a variety of distributions have been derived to model stochastic variations in protein or mRNA levels,[66] gene expression is most often analysed as a Gaussian trait after transformation.[67] As such, one could consider that it does not have peculiarities as a phenotypic trait. Yet studying gene expression using the tools of quantitative genetics might be quite challenging because it requires studying a very large number of traits (usually in the order of thousands) at once. This means that a high-dimension additive genetic variance-covariance matrix (or **G**-matrix) needs to be inferred. For a study on 1,000 genes for example, this would require the inference of 500,500 parameters for the **G**-matrix. The number of individuals necessary to infer such a matrix with satisfying precision (well above millions) is definitely out of our reach.

In order to work around this issue, Blows *et al.*[68] used an approach based on "block sampling" in which submatrices of low dimension (e.g. $k = 5$ or $k = 50$ in their study) are estimated and higher dimension matrices are reconstructed from them. More precisely, the complete **G**-matrix is separated into "blocks" $\mathbf{B}_{ij}$ of size $k = M/m$, where M is the number of dimensions of **G**:

$$\mathbf{G} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1m} \\ \mathbf{B}_{12} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{1m} & \mathbf{B}_{2m} & \cdots & \mathbf{B}_{mm} \end{pmatrix} \tag{19}$$

which is reconstructed from a simpler, block-diagonal matrix **K**:

$$\mathbf{K} = \begin{pmatrix} \mathbf{B}_{11} & 0 & \cdots & 0 \\ 0 & \mathbf{B}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{B}_{mm} \end{pmatrix} \tag{20}$$

with each block $\mathbf{B}_{ii}$ being independently inferred from the dataset (e.g. estimation of a small **G**-matrix from k random genes). The reconstruction of **G** from **K** is based on theoretical work in Ref. 69, 70. Of course, this approach is based on strong assumptions, namely that pleiotropy (and/or linkage) is widespread throughout the genome, allowing us to scavenge useful information on the structure of the **G**-matrix by using "smaller pieces" of it. Besides other methodological results, the authors confirmed, using *Drosophila serrata* gene expression data, the strong influence of pleiotropy/linkage in gene expression data.[68] Using gene ontology analysis,[71,72] they also found an enrichment in regulatory gene ontologies in the genes contributing the most to the major axis of their **G**-matrix (hence the most pleiotropic genes).[68]

Another promising approach to address this issue of quantitative genetics in high-dimensionality is the Bayesian sparse factor analysis.[73] This is again based on the assumption that pleiotropy is widespread and hence the **G**-matrix is of low rank. Following a statistical approach close to principal component analysis, the **G**-matrix is approximated using a number of independent *latent factors* (much smaller than the dimensionality of **G**), for which heritabilities can be estimated. In summary, the heritability of those latent factors is estimated in a diagonal matrix $\mathbf{\Sigma}_{h^2}$ and the total matrix **G** is reconstructed using the model:

$$\mathbf{G} = \mathbf{\Lambda}\mathbf{\Sigma}_{h^2}\mathbf{\Lambda}^T + \mathbf{\Psi} \tag{21}$$

where $\mathbf{\Lambda}$ is the "factor loadings" matrix relating the latent factors to the actual trait (as is the case in a

principal component analysis) and $\mathbf{\Psi}$ is a diagonal trait-specific additive genetic variance matrix. By relating these latent factors to the actual phenotypic traits, this model allows sets of traits linked through pleiotropy to be clustered. Using this model on a dataset on gene expression of 414 genes in *Drosophila melanogaster*, the authors identified 27 latent factors, most of which had moderate to high heritabilities and two of which were significantly genetically correlated with fitness.

Such approaches approximating large **G**-matrices by low-rank matrices are currently challenging to implement and their inferential value is based on strong assumptions (e.g. that pleiotropy is widespread). Hence they should be analysed with due caution. But despite these computational and statistical difficulties, they still deepen our understanding of genotype-phenotype maps and genetic correlations at the level of gene expression. Given the trending abundance of gene expression studies, we can hopefully expect such studies to be replicated in the near future and even connections to upper levels (morphological and life-history traits) to be made. Note that rank reduction through the simpler approach of factor analysis can also be useful for datasets of much smaller dimensionality, such as a dozen of traits.[74–76]

Future directions

This review of recent developments in the field of evolutionary quantitative genetics shows that, despite its age, the field is still active and undergoing strong developments. New and exciting theoretical and methodological work opens the way to more refined analyses of complex data, in turn enhancing our understanding of the sources of biological variation and the mechanism of its differential transmission to future generations. With these new tools, evolutionary biologists can understand the genetic bases of a broader variety of phenotypic traits, with an increasing ability to fit peculiarly distributed traits. However, despite the growing use of those tools, we are still lacking a good perspective on how these non-Gaussian traits are shaped by biological processes. If the (multiple) threshold model makes a convincing case that binary (or ordinal) traits can arise from a normally distributed variable, the cases of other models typical from GLMMs are far less supported. No theoretical work that I am aware of describes how the logarithm link of a Poisson model can be biologically justified (although a possible interpretation of the use of logarithm would be the assumption of a totally multiplicative model). The fact that it is not possible to select Poisson traits for a single specific value is also not properly described: as the variance is equal to the mean, even a latent trait without variation can still exhibit phenotypic variation on the data scale. Note that more flexible alternatives exists such as the double-Poisson[77] and the Conway-Maxwell Poisson[78] (COM-Poisson) distributions, but they are still making assumptions regarding a link between mean and variance, e.g. for the COM-Poisson, for which the mean reaches zero when the variance decreases to a minimal value. These assumptions are made by evolutionary biologists each time they use such kind of models, because they are statistically convenient, but almost no theoretical work has been done on whether they can be reasonably made on a biological system to begin with (although see Refs. 31, 32). Convincing genotype-phenotype maps accounting for the existence of quantitative, but non-Gaussian, traits are thus needed.

Another direction of methodological development would be to extend our ability to analyse biological systems in a holistic way. A first approach would be to use path diagrams, as was suggested by Wright[79] long ago and is currently becoming more promoted.[64,80,81] Aster models, discussed in this paper, uses such path diagrams in a convenient way for evolutionary biologists and could become a handy set of models, when they will be able to deal more efficiently with typical quantitative genetic data (e.g. pedigrees). A second way would be to enhance our ability to analyse datasets of high dimensions that are currently produced by gene expression studies, or to a lesser extent by large multivariate studies of multiple traits,[74–76] and might be produced in the near future by high-throughput phenotyping technologies.[82] I have discussed several recent attempts at doing this. Although they rely on strong statistical assumptions, they seem to be able to provide interesting information on, at least, the rough structure of the genetic covariation among traits and even on how groups of traits can be heritable and selected together. As we go down that road, however, we must keep in mind that these methods are using a scrap of information, coming from subset (or low-rank) analyses, to infer properties about the whole system, which will always be a limited process.

Conclusion

Despite being around a century old,[4] quantitative genetics is still a very dynamic field of research, possibly even gaining momentum over the last decades.[83] The field was first centred around the study of Gaussian traits because it is based on the infinitesimal model.[3] It was quickly applied to binary traits,[8] but only relatively recently was it extended to other trait distributions (see e.g. Ref. 84 for pioneer work on Poisson-distributed traits). As shown in this review, progress has been made to better understand the properties of the models used for non-Gaussian traits[23] and new ideas are being developed for particular types of traits.[52,63,68] However, our understanding of the genotype-to-phenotype map for non-Gaussian traits is still only superficial. For example, while GLMMs are convenient statistical models, whether their biological assumptions (see Box 2, notably that the model generates non-additive genetic variance and incompressible noise) are justified is most often unknown and left to the researcher's expert judgement. The growing interest in the study of the quantitative genetics of non-Gaussian traits will most certainly trigger new research in this area, which would in turn help design more appropriate quantitative genetics models. In the meantime, it should be acknowledged that these methods are indeed extremely useful and should not be discarded, but we should also acknowledge their underlying assumptions and be careful to interpret our parameters in the light of these assumptions.

Acknowledgements

I thank Olivier Gimenez, Ruth Shaw, Mark Blows, Anna Santure and Irène Till-Bottraud for their helpful comments on the manuscript, especially to ensure the description of their approaches was accurate. I also thank three reviewers, including Jarrod Hadfield and Loeske Kruuk, and the editor for their helpful criticism.

References

1. D. S. Falconer & T. F. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Harlow, Essex (UK): Benjamin Cummings.
2. M. Lynch & B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, Massachussets (US): Sinauer Associates.
3. R. A. Fisher. 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinburgh* **52:** 399–433.
4. D. A. Roff. 2007. A Centennial Celebration for Quantitative Genetics. *Evolution* **61**.5**:** 1017–1032.
5. W. G. Hill & M. Kirkpatrick. 2010. What Animal Breeding Has Taught Us about Evolution. *Annu. Rev. Ecol. Evol. Syst.* **41:** 1–19.
6. C. R. Henderson. 1950. Estimation of Genetic Parameters. *Ann. Math. Stat.* **21:** 309–310.
7. C. R. Henderson. 1976. A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics* **32**.1**:** 69–83.
8. S. Wright. 1934. An Analysis of Variability in Number of Digits in an Inbred Strain of Guinea Pigs. *Genetics* **19**.6**:** 506–536.
9. K. Pearson. 1900. Mathematical Contributions to the Theory of Evolution. Vii. on the Correlation of Characters Not Quantitatively Measurable. *Phil. Trans. R. Soc.* **195:** 1–405.
10. E. R. Dempster & I. M. Lerner. 1950. Heritability of Threshold Characters. *Genetics* **35**.2**:** 212–236.
11. P. M. Visscher & N. R. Wray. 2015. Concepts and Misconceptions about the Polygenic Additive Model Applied to Disease. *Hum. Hered.* **80**.4**:** 165–170.
12. T. Reich, J. W. James, & C. A. Morris. 1972. The Use of Multiple Thresholds in Determining the Mode of Transmission of Semi-Continuous Traits. *Ann. Hum. Genet.* **36**.2**:** 163–184.
13. D. Gianola & J. L. Foulley. 1983. Sire Evaluation for Ordered Categorical Data with a Threshold Model. *Genet. Sel. Evol.* **15**.2**:** 201.
14. D. A. Roff. 2001. The Threshold Model as a General Purpose Normalizing Transformation. *Heredity* **86**.4**:** 404–411.
15. J. Felsenstein. 2005. Using the Quantitative Genetic Threshold Model for Inferences between and within Species. *Philos. Trans. Biol. Sci.* **360**.1459**:** 1427–1434.
16. D. Gianola. 1982. Theory and Analysis of Threshold Characters. *J. Anim. Sci.* **54**.5**:** 1079–1096.

17. L. van Vleck. 1972. Estimation of Heritability of Threshold Characters. *J. Dairy Sci.* **55**.2**:** 218–225.

18. R. C. Elston, W. G. Hill, & C. Smith. 1977. Query: Estimating "Heritability" of a Dichotomous Trait. *Biometrics* **33**.1**:** 231–236.

19. D. A. Roff. 1997. *Evolutionary Quantitative Genetics*. New York (US): Chapman & Hall. 515 pp.

20. V. Thériault, D. Garant, L. Bernatchez, *et al.* 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**.6**:** 2266–2277.

21. A. Charmantier, A. J. Keyser, & D. E. Promislow. 2007. First Evidence for Heritable Variation in Cooperative Breeding Behaviour. *Proc. R. Soc. B Biol. Sci.* **274**.1619**:** 1757–1761.

22. A. Charmantier, M. Buoro, O. Gimenez, *et al.* 2011. Heritability of Short-Scale Natal Dispersal in a Large-Scale Foraging Bird, the Wandering Albatross. *J. Evol. Biol.* **24**.7**:** 1487–1496.

23. P. de Villemereuil, O. Gimenez, & B. Doligez. 2013. Comparing Parent-Offspring Regression with Frequentist and Bayesian Animal Models to Estimate Heritability in Wild Populations: A Simulation Study for Gaussian and Binary Traits. *Methods Ecol. Evol.* **4**.3**:** 260–275.

24. M. Freeman & J. B. Gurdon. 2002. Regulatory Principles of Developmental Signaling. *Annu. Rev. Cell Dev. Biol.* **18:** 515–539.

25. B. Doligez, G. Daniel, P. Warin, *et al.* 2011. Estimation and Comparison of Heritability and Parent–offspring Resemblance in Dispersal Probability from Capture–recapture Data Using Different Methods: The Collared Flycatcher as a Case Study. *J. Ornithol.* **152:** 539–554.

26. I. Olesen, M. Perez-Enciso, D. Gianola, *et al.* 1994. A Comparison of Normal and Nonnormal Mixed Models for Number of Lambs Born in Norwegian Sheep. *J. Anim. Sci.* **72**.5**:** 1166–1173.

27. C. A. Matos, D. L. Thomas, D. Gianola, *et al.* 1997. Genetic Analysis of Discrete Reproductive Traits in Sheep Using Linear and Nonlinear Models: I. Estimation of Genetic Parameters. *J. Anim. Sci.* **75**.1**:** 76–87.

28. P. B. Chi, A. E. Duncan, P. A. Kramer, *et al.* 2014. Heritability Estimation of Osteoarthritis in the Pig-Tailed Macaque (Macaca Nemestrina) with a Look toward Future Data Collection. *PeerJ* **2:** e373.

29. D. J. Lee, J. T. Brawner, & G. S. Pegg. 2014. Screening Eucalyptus Cloeziana and E. Argophloia Populations for Resistance to Puccinia Psidii. *Plant Dis.* **99**.1**:** 71–79.

30. G. Makouanzi, J.-M. Bouvet, M. Denis, *et al.* 2014. Assessing the Additive and Dominance Genetic Effects of Vegetative Propagation Ability in Eucalyptus—influence of Modeling on Genetic Gain. *Tree Genet. Genomes* **10**.5**:** 1243–1256.

31. P. de Villemereuil, H. Schielzeth, S. Nakagawa, *et al.* 2016. General Methods for Evolutionary Quantitative Genetic Inference from Generalised Mixed Models. *Genetics* **204**.3**:** 1281–1294.

32. M. B. Morrissey. 2015. Evolutionary Quantitative Genetics of Nonlinear Developmental Systems. *Evolution* **69**.8**:** 2050–2066.

33. N. E. Breslow & D. G. Clayton. 1993. Approximate Inference in Generalized Linear Mixed Models. *J. Am. Stat. Assoc.* **88**.421**:** 9–25.

34. N. E. Breslow & X. Lin. 1995. Bias Correction in Generalised Linear Mixed Models with a Single Component of Dispersion. *Biometrika* **82**.1**:** 81–91.

35. B. M. Bolker, M. E. Brooks, C. J. Clark, *et al.* 2009. Generalized Linear Mixed Models: A Practical Guide for Ecology and Evolution. *Trends Ecol. Evol.* **24**.3**:** 127–135.

36. S. Nakagawa & H. Schielzeth. 2010. Repeatability for Gaussian and Non Gaussian Data: A Practical Guide for Biologists. *Biol. Rev.* **85**.4**:** 935–956.

37. M. B. Morrissey, P. de Villemereuil, B. Doligez, *et al.* "Bayesian Approaches to the Quantitative Genetic Analysis of Natural Populations" In *Quantitative Genetics in the Wild*. Ed. by A. Charmantier, D. Garant, & L. E. Kruuk. Oxford (UK): Oxford University Press**:** pp. 228–253.

38. A. R. Gilmour, R. D. Anderson, & A. L. Rae. 1985. The Analysis of Binomial Data by a Generalized Linear Mixed Model. *Biometrika* **72**.3**:** 593–599.

39. J. L. Foulley & S. Im. 1993. A Marginal Quasi-Likelihood Approach to the Analysis of Poisson Variables with Generalized Linear Mixed Models. *Genet. Sel. Evol.* **25**.1**:** 101.

40. J. L. Lush. 1937. *Animal Breeding Plans*. Ames, Iowa (US): Iowa State College Press.

41. A. Robertson. 1966. A Mathematical Model of the Culling Process in Dairy Cattle. *Anim. Sci.* **8:** 95–108.

42. G. R. Price. 1970. Selection and Covariance. *Nature* **227:** 520–521.

43. P. de Villemereuil. 2016. *QGglmm: Estimate Quantitative Genetics Parameters from Generalised Linear Mixed Models*. URL: https://cran.r-project.org/web/packages/QGglmm/index.html.

44. B. Walsh & M. W. Blows. 2009. Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annu. Rev. Ecol. Evol. Syst.* **40**.1**:** 41–59.

45. J. D. Hadfield. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *J. Stat. Softw.* **33**.2**:** 1–22.

46. E. Paradis, S. R. Baillie, W. J. Sutherland, *et al.* 1998. Patterns of Natal and Breeding Dispersal in Birds. *J. Anim. Ecol.* **67**.4**:** 518–536.

47. R. J. Sardell, M. G. Nittala, L. D. Adams, *et al.* 2016. Heritability of Choroidal Thickness in the Amish. *Ophthalmology* **123**.12**:** 2537–2544.

48. J. D. Hadfield. 2016. *MCMCglmm Course Notes*. URL: http://cran.r-project.org/web/packages/MCMCglmm/index.html.

49. J.-D. Lebreton, K. P. Burnham, J. Clobert, *et al.* 1992. Modeling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach with Case Studies. *Ecol. Monogr.* **62**.1**:** 67–118.

50. J. A. Royle. 2008. Modeling Individual Effects in the Cormack–Jolly–Seber Model: A State–space Formulation. *Biometrics* **64**.2**:** 364–370.

51. O. Gimenez & R. Choquet. 2010. Individual Heterogeneity in Studies on Marked Animals Using Numerical Integration: Capture–recapture Mixed Models. *Ecology* **91**.4**:** 951–957.

52. J. Papaïx, S. Cubaynes, M. Buoro, *et al.* 2010. Combining Capture-Recapture Data and Pedigree Information to Assess Heritability of Demographic Parameters in the Wild. *J. Evol. Biol.* **23**.10**:** 2176–2184.

53. R. B. O'Hara, J. M. Cano, O. Ovaskainen, *et al.* 2008. Bayesian Approaches in Evolutionary Quantitative Genetics. *J. Evol. Biol.* **21**.4**:** 949–957.

54. M. Plummer. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling" In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, March***:** pp. 20–22.

55. R. G. Shaw, C. J. Geyer, S. Wagenius, *et al.* 2008. Unifying Life-history Analyses for Inference of Fitness and Population Growth. *Am. Nat.* **172**.1**:** E35–E47.

56. C. J. Geyer, S. Wagenius, & R. G. Shaw. 2007. Aster Models for Life History Analysis. *Biometrika* **94**.2**:** 415–426.

57. S. Wagenius, H. H. Hangelbroek, C. E. Ridley, *et al.* 2010. Biparental Inbreeding and Interremnant Mating in a Perennial Prairie Plant: Fitness Consequences for Progeny in Their First Eight Years. *Evolution* **64**.3**:** 761–771.

58. P. H. Leinonen, D. L. Remington, & O. Savolainen. 2011. Local Adaptation, Phenotypic Differentiation, and Hybrid Fitness in Diverged Natural Populations of *Arabidopsis Lyrata*. *Evolution* **65**.1**:** 90–107.

59. R. I. Colautti & S. C. H. Barrett. 2013. Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. *Science* **342**.6156**:** 364–366.

60. E. J. Kleynhans, S. P. Otto, P. B. Reich, *et al.* 2016. Adaptation to Elevated CO2 in Different Biodiversity Contexts. *Nat. Commun.* **7:** 12358.

61. K. E. Samis, A. López-Villalobos, & C. G. Eckert. 2016. Strong Genetic Differentiation but Not Local Adaptation toward the Range Limit of a Coastal Dune Plant. *Evolution* **70**.11**:** 2520–2536.

62. C. J. Geyer, C. E. Ridley, R. G. Latta, *et al.* 2013. Local Adaptation and Genetic Effects on Fitness: Calculations for Exponential Family Models with Random Effects. *Ann. Appl. Stat.* **7**.3**:** 1778–1795.

63. C. J. Geyer & R. G. Shaw. 2013. *Aster Models with Random Effects and Additive Genetic Variance for Fitness*. URL: http://conservancy.umn.edu/handle/11299/152355.

64. C. E. Thomson, F. Bayer, N. Crouch, *et al.* 2017. Selection on Parental Performance Opposes Selection for Larger Body Mass in a Wild Population of Blue Tits. *Evolution* **71**.3**:** 716–732.

65. M. V. Rockman. 2008. Reverse Engineering the Genotype–phenotype Map with Natural Genetic Variation. *Nature* **456**.7223**:** 738–744.

66. V. Shahrezaei & P. S. Swain. 2008. Analytical Distributions for Stochastic Gene Expression. *Proceedings of the National Academy of Sciences* **105**.45**:** 17256–17261.

67. R. S. Parrish, H. J. Spencer III, & P. Xu. 2009. Distribution Modeling and Simulation of Gene Expression Data. *Comput. Stat. Data Anal.* Statistical Genetics & Statistical Genomics: Where Biology, Epistemology, Statistics, and Computation Collide **53**.5**:** 1650–1660.

68. M. W. Blows, S. L. Allen, J. M. Collet, *et al.* 2015. The Phenome-Wide Distribution of Genetic Variance. *Am. Nat.* **186**.1**:** 15–30.

69. J.-C. Bourin & E.-Y. Lee. 2012. Unitary Orbits of Hermitian Operators with Convex or Concave Functions. *Bull. Lond. Math. Soc.* **44**.6**:** 1085–1102.

70. J.-C. Bourin & E.-Y. Lee. 2013. Decomposition and Partial Trace of Positive Matrices with Hermitian Blocks. *Int. J. Math.* **24:** 1350010.

71. M. Ashburner, C. A. Ball, J. A. Blake, *et al.* 2000. Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* **25**.1**:** 25–29.

72. H. Mi, A. Muruganujan, & P. D. Thomas. 2013. PANTHER in 2013: Modeling the Evolution of Gene Function, and Other Gene Attributes, in the Context of Phylogenetic Trees. *Nucleic Acids Res.* **41:** D377–386.

73. D. E. Runcie & S. Mukherjee. 2013. Dissecting High-Dimensional Phenotypes with Bayesian Sparse Factor Analysis of Genetic Covariance Matrices. *Genetics* **194**.3**:** 753–767.

74. K. McGuigan & M. W. Blows. 2007. The Phenotypic and Genetic Covariance Structure of Drosphilid Wings. *Evolution* **61**.4**:** 902–911.

75. E. Schroderus, I. Jokinen, M. Koivula, *et al.* 2010. Intra- and Intersexual Trade-offs between Testosterone and Immune System: Implications for Sexual and Sexually Antagonistic Selection. *The American Naturalist* **176**.4**:** E90–E97.

76. C. A. Walling, M. B. Morrissey, K. Foerster, *et al.* 2014. A Multivariate Analysis of Genetic Constraints to Life History Evolution in a Wild Population of Red Deer. *Genetics* **198**.4**:** 1735–1749.

77. B. Efron. 1986. Double Exponential Families and Their Use in Generalized Linear Regression. *J. Am. Stat. Assoc.* **81**.395**:** 709–721.

78. G. Shmueli, T. P. Minka, J. B. Kadane, *et al.* 2005. A Useful Distribution for Fitting Discrete Data: Revival of the Conway–Maxwell–Poisson Distribution. *J. R. Stat. Soc. Ser. C Appl. Stat.* **54**.1**:** 127–142.

79. S. Wright. 1921. Correlation and Causation. *J. Agric. Res.* **20:** 557–585.

80. M. B. Morrissey. 2014. Selection and Evolution of Causally Covarying Traits. *Evolution* **68**.6**:** 1748–1761.

81. C. E. Thomson & J. D. Hadfield. 2017. Measuring Selection When Parents and Offspring Interact. *Methods Ecol. Evol.* **8**.6**:** 678–687.

82. J. L. Araus & J. E. Cairns. 2014. Field High-Throughput Phenotyping: The New Crop Breeding Frontier. *Trends Plant Sci.* **19**.1**:** 52–61.

83. A. Charmantier, D. Garant, & L. E. B. Kruuk, eds. *Quantitative Genetics in the Wild*. Oxford (UK): Oxford University Press. 293 pp.

84. J. L. Foulley, D. Gianola, & S. Im. 1987. Genetic Evaluation of Traits Distributed as Poisson-Binomial with Reference to Reproductive Characters. *Theoret. Appl. Genetics* **73**.6**:** 870–877.