Estimating Heritability and Genetic Correlation of Exercise-Related Traits in Diversity Outbred Mice

Kathie Sun[1,2], Greg Keele[1,2], Daniel Pomp[1], and William Valdar[1,3]

1. Department of Genetics, University of North Carolina at Chapel Hill, NC, USA
2. Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, NC, USA
3. Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, NC, USA

Summary

Motivation: Narrow sense heritability is a fundamental concept in quantitative genetics that describes the proportion of phenotypic variation for a given trait that is due to additive genetic variation within the population.

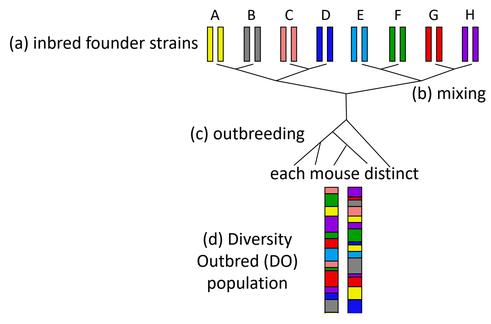
In addition, many common diseases are highly polygenic and thus the result of variation in many common alleles acting in concert [1, 2]. Shared causal loci (pleiotropy) results in correlations between traits. Genetic correlation is an estimate of the portion due to additive genetic effects common to pairs of traits.

These two quantities are often estimated using likelihood-based methods, which do not easily produce interval estimates in variance components. In addition, default implicitly-uniform priors on variance components in maximum likelihood estimation may be informative and influence estimation.

Experiment: We used phenotype and genotype data from experiments in two structured mouse populations to derive and implement a Bayesian procedure for estimating heritability and genetic correlation. Subsequently, we used these procedures to obtain point and interval estimates from exercise- and metabolism-related traits.

Results: Heritability and genetic correlation estimates have been published in previous mouse studies for several of the traits explored in this project. Overall, our results for traits related to running and body size broadly agree with the literature, particularly estimates from the mouse population fed with a consistent diet. [3, 4].

Diversity Outbred Models



Estimates of heritability and genetic correlation can be derived from studies in structured populations with known characteristics. Diversity Outbred (DO) mice are a large, heterogeneous population derived from eight inbred founder strains bred randomly over many generations [5]. The resulting mice contain extensive genetic diversity from random matings.

DO1 A DO population of 315 male and female mice were tested with sex as a covariate. h^2 was calculated on 10 phenotypes.

DO2 A separate DO population of 287 female mice was fed either a chrolic acid or high protein diet, which was included in the models for 8 phenotypes.

Statement of Model

Narrow-sense heritability, h^2, is calculated as $\frac{Var(Additive)}{Var(Phenotype)}$. Under such a definition,

$$h^2 = \frac{\tau^2}{\tau^2 + \sigma^2}$$

where τ^2 is a parameter of variance due to inherited genetic factors and σ^2 is a parameter of residual variance that is due to noise.

Bayesian Model for Heritability

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \mathbf{I}\sigma^2)$$

$$\boldsymbol{\beta} \sim \mathrm{N}(0, \mathbf{I}\sigma^2_\beta)$$

where **y** is an *n*-vector containing a phenotype for *n* individuals, **R** is an $n \times n$ similarity kernel for those individuals, and **X** is an $n \times p$ design matrix of *p* covariates. **u** represents a vector of additive genetic effects that impact expression of the phenotype and is influenced by an individual's genetic similarity to the rest of the cohort.

$$\mathbf{u} \sim \mathrm{N}(0, \mathbf{R}\tau^2)$$

Sampling of h^2 was done through Markov Chain Monte Carlo procedures via Gibbs sampling to find

$$p(h^2|\mathbf{y})$$

The h^2 mean estimate is calculated with 350 samples from the Gibbs sampler from an original 5000 iterations (1500 burn-in and thinning by 10).

Prior distributions

Instead of explicitly putting a prior on h^2, we can choose additive genetic and residual components of variance that follow independent inverse gamma distributions.

$$\tau^{-2} \sim \mathrm{Gamma}(a_\tau, b)$$

$$\sigma^{-2} \sim \mathrm{Gamma}(a_\sigma, b)$$

Using an auxillary random variable, U, which can be defined as $U = \tau^{-2} + \sigma^{-2}$, we can make the one-to-one transformation

$$h^2 = f_1(\tau, \sigma) = \frac{\tau^2}{\tau^2 + \sigma^2} = \frac{\sigma^{-2}}{\tau^{-2} + \sigma^{-2}},\ 0 < h^2 < 1,$$

$$U = f_2(\tau, \sigma) = \tau^{-2} + \sigma^{-2},\ U > 0$$

In the case that τ^{-2} and σ^{-2} share the same shape parameter, b, the prior distribution of h^2 reduces to the density of a beta distribution with parameters a_σ and a_τ. h^2 ranges from 0-1 so we can choose an implicit prior distribution that is truly uniform over 0 to 1 [6].

$$h^2 \sim \mathrm{Beta}(a_\tau = 1, a_\sigma = 1)$$

Model for Genetic Correlation

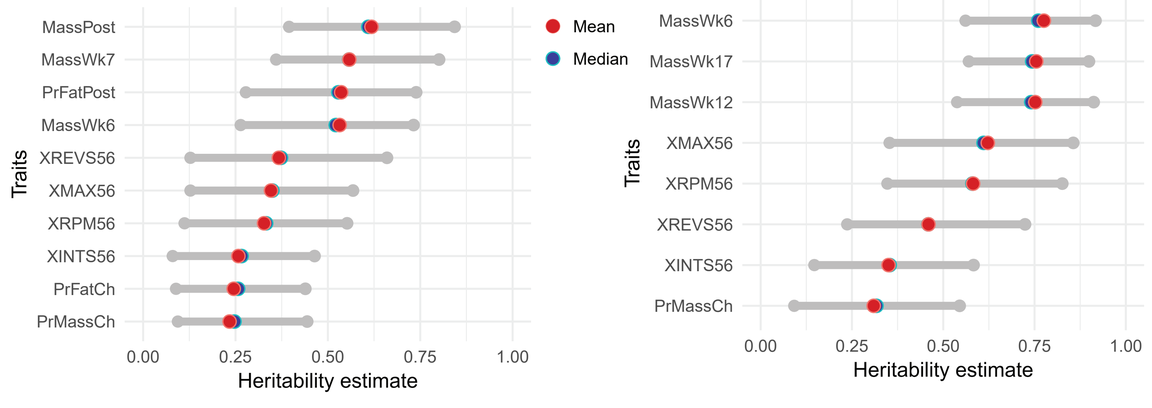
$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 + \mathbf{u_1} \\ \boldsymbol{\mu}_2 + \mathbf{u_2} \end{bmatrix}, \begin{bmatrix} \mathbf{I}\sigma_1^2 & \rho_\epsilon \mathbf{I}\sigma_1\sigma_2 \\ \rho_\epsilon \mathbf{I}\sigma_1\sigma_2 & \mathbf{I}\sigma_2^2 \end{bmatrix} \right)$$

where $\mathbf{y_1}$ and $\mathbf{y_2}$ contain two phenotypes measured on the same set of *n* individuals, and **R** is an $n \times n$ genetic relatedness matrix for those individuals.

$$\begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{R}\tau_1^2 & \rho_R \mathbf{R}\tau_1\tau_2 \\ \rho_R \mathbf{R}\tau_1\tau_2 & \mathbf{R}\tau_2^2 \end{bmatrix} \right)$$

The genetic correlation is defined as the correlation ρ_R in the bivariate mixed model.
We calculated genetic correlation between traits using the $\mathbf{u}$ vectors from Gibbs sampling for two different phenotypes according to the model and Bayesian methods as written for heritability.

Heritability Estimates



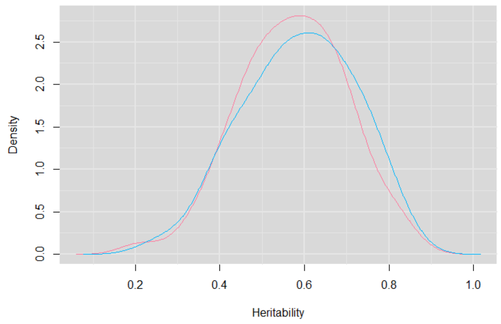
Heritability estimates and highest 95% posterior density (HPD) region for 10 phenotypes in DO1 (**left**) and 8 phenotypes in DO2 (**right**)

References and Funding

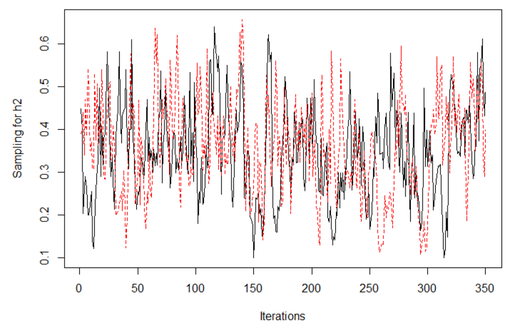
[1] Bulik-Sullivan B, et al. (2015). *Nat Genet.* 47(11):1236-41.
[2] Lees SH, et al. (2012). *Bioinformatics.* 28(19):2540-2.
[3] Leamy LJ, et al. (2005).*Genetics Selection Evolution.* 37:151.
[4] Swallow JG, et al. (1998). *Behavior Genetics.* 28:3.
[5] Churchhill GA, et al. (2012). *Mann Genome.* 23:713-8.
[6] Sorenson D and Gianola D. (2002). ISBN: 0-387-954406.

Funding UNC-CH Biological and Biomedical Sciences Program, Caroline H. and Thomas S. Royster Fellowship, R01-GM104125 (Valdar)

Performance of Gibbs Sampler

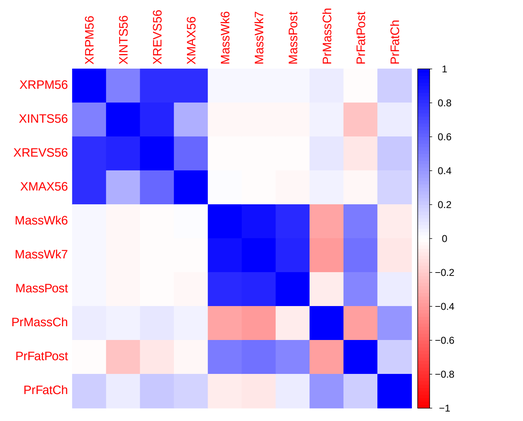


Posterior h^2 distribution for XRPM56 in two runs



Trace plot of h^2 sampling for XINTS56 in two runs

Genetic Correlation Estimates



Pearson correlation coefficients for 10 phenotypes in DO1 (**above**) and 8 phenotypes in DO2 (**below**) using **u** data averaged across samples