Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Opportunities for improvement of phenotypic variability: influence of direct vs epistatic effects

Xia Shen[1], William G. Hill[2]

[1]Swedish University of Agricultural Sciences, Uppsala, Sweden, [2] University of Edinburgh, Edinburgh, United Kingdom.

ABSTRACT Genetic regulation of phenotypic variability has recently become a hot topic in quantitative genetics. Empirical evidence has shown that individual loci controlling phenotypic variability or "vQTL" can be mapped in population-based studies, and also that there are polygenic effects on the variability. In the breeding context, scientists have been investigating whether it is possible to select for phenotypic variability or uniformity, in order to better control the production quality. However, since apparent vQTL effects can be caused by various types of interactions, such as epistasis which influence the genetic variance, possibility of selection for phenotypic variability may not be guaranteed. Here, based on epistatic models, we theoretically analyze and discuss the selection for phenotypic variability. Our results indicate that such selection would be effective in changing the phenotypic variance via interaction effects only for a few special situations.

Keywords: variation in phenotypic variability, one vs many loci, epistasis

Introduction

Recent evidence indicates that the uniformity or variability, rather than the magnitude, of complex traits is also under genetic control. Evidence comes both at the multi-locus level (*e.g.* Hill et al., 2010; Mackay et al., 2005) and can even be mapped to the individual locus level (*e.g.* Geiler-Samerotte et al., 2013; Jimenez-Gomez et al., 2011; Shen et al., 2012; Wolc et al., 2012; Yang et al., 2012). Such individual loci have been named "vQTL" (Rönnegård et al., 2010a). vQTL can arise not just from the effects of individual loci on controlling variation of the phenotype directly through their influence on environmental variance, but also from epistasis on mean phenotype, as variability among genotypic values at one locus depends on the allele at the other. In human genetics, for example, genome-wide association analysis of variability-controlling sites is a powerful way to detect gene-gene and gene-environment interactions because fitting variance differences at one site may give much higher power than fitting interaction effects at two or more (*e.g.* Paré et al., 2010; Struchalin et al., 2010). It is therefore potentially very useful in re-construction of interaction pathways for prediction purposes for both disease and other complex traits. Genome-wide data also indicate that there is a substantial amount of inherited variation in such variance-controlling loci (Shen et al., 2012). Hence, new directions should be developed for selection using such knowledge, incorporating genomic data, and can be important in better understanding and partitioning the variation in quantitative traits.

The idea or concept of variance heterogeneity, that genes and genotypes may influence the variability as well as the mean of complex traits, is of long standing, but only relatively recently has evidence become available, *e.g.* (Hill et al., 2010) and subsequent to that review, at the level of individual loci. In agriculture, genes regulating robustness could be of benefit in producing more uniform products, saving in post-processing costs, and/or animals less sensitive to management or environment. There are a limited number of laboratory experiments in which selection to change phenotypic variance has been undertaken, with some success (see the review by Geiler-Samerotte et al., 2013). Breeding projects, such as RobustMilk (http://www.robustmilk.eu), have been initiated to test the effectiveness of selection for phenotypic uniformity.

As the cause of apparent vQTL could be epistasis, however, it is important to know whether selection is changing effects on phenotypic variance exhibited at individual loci or genetic variance due to epistasis among loci. Since selection for phenotypic variability or uniformity is based on a polygenic scenario rather than individual loci, the feasibility of such selection is unclear (*e.g.* counterexamples in Hill et al., 2010, with rather low estimated heritability of genetically regulated variance heterogeneity). In this paper, we investigate models of epistatic interaction, and then discuss how these influence the possibility of successful selection for phenotypic variability. We first consider pairs of interacting loci, and then generalize to more. There are assumed to be no direct effects on the phenotypic variability (*i.e.* variance of phenotype given genotype), but potential differences in genetic variance at individual or multiple sites as a consequence of epistasis.

Theoretical Analysis & Results

Epistasis between two loci and its effect on variance heterogeneity Consider a two-locus model with additive effect at each locus and additive x additive interaction in an outbred population of a diploid organism, assuming Hardy-Weinberg equilibrium (HWE) at all loci and no linkage disequilibrium between them,

$$y = \mu + g_1\beta_1 + g_2\beta_2 + g_1g_2\beta_{12} + e$$

where the individual phenotypic value y is determined by an intercept μ, genotypes at loci 1 and 2 (g_1 and g_2, coded as the number of copies of the major allele) with their main effects β_1 and β_2, their interaction (epistatic) effect β_{12} and a residual $e \sim N(0,\sigma^2)$. Given the minor allele frequency (MAF) q at locus 2, we have $g_2 \sim B(2,q)$, $E[g_2] = 2q$ and

$V(g_2) = 2q(1-q)$. In order to evaluate the phenotypic variability (variance heterogeneity) across locus 1 genotypes caused by its interaction with locus 2, we have

$$\begin{aligned} V(y|g_1=0) &= V(E[y|g_1=0,g_2]) + E[V(y|g_1=0,g_2)] \\ &= V(\mu + g_2\beta_2) + E[\sigma^2] \\ &= 2\beta_2^2 q(1-q) + \sigma^2 \end{aligned}$$

Similarly,

$$\begin{aligned} V(y|g_1=1) &= 2(\beta_2+\beta_{12})^2 q(1-q) + \sigma^2 \\ V(y|g_1=2) &= 2(\beta_2+2\beta_{12})^2 q(1-q) + \sigma^2 \end{aligned}$$

Combining terms,

$$V(y|g_1) = 2q(1-q)(\beta_2+\beta_{12}g_1)^2 + \sigma^2 \qquad (1)$$

Clearly, $V(y|g_1)$ is a constant if and only if $\beta_{12}=0$, which implies that, if there is such an epistatic effect, there will be phenotypic variance heterogeneity regulated by locus 1.

Polygenic epistasis with one apparent vQTL as "controller" Let us extend the model to a polygenic scenario where locus 1 (vQTL) influences the action of multiple (k) loci and so has epistatic effects with each, and k is large enough to approximate statistical asymptotic properties.

$$y = \mu + g_1\beta_1 + \sum_{j=2}^{k+1}(g_j\beta_j + g_1 g_j \beta_{1j}) + e$$

Similarly, $g_j \sim B(2,q_j)$,where q_j is the MAF at locus j. $E[g_j] = 2q_j$ and $V(g_j) = 2q_j(1-q_j)$ $(2 \leqslant j \leqslant k+1)$. Hence, extending eq.(1)

$$V(y|g_1) = 2\sum_{j=2}^{k+1} q_j(1-q_j)(\beta_j + g_1\beta_{1j})^2 + \sigma^2$$

Similarly, $V(y|g_1)$ depends on g_1 if there is at least one j for which $\beta_{1j} \neq 0$, which means that if there are epistatic effects between locus 1 and multiple loci interacting with locus 1, the phenotypic variance will depend on the genotype of locus 1. If for all j, $\beta_j\beta_{1j} > 0$, *i.e.* the epistatic effect (if non-zero) and the main effect of the corresponding QTL have the same direction, then

$$V(y|g_1=2) > V(y|g_1=1) > V(y|g_1=0)$$

and the variance difference will be largest, otherwise the phenotypic variance does not necessarily vary across locus 1 genotypes. Nevertheless, assuming for example k is large and for each j, $\beta_j \sim N(0,\sigma_j^2)$, $\beta_{1j} \sim N(0,\sigma_{1J}^2)$, and $Cov(\beta_j,\beta_{1j}) = 0$, the phenotypic variance given the genotype of locus 1 is approximately

$$\begin{aligned} E_{\beta_j,\beta_{1j}}[V(y|g_1)] &= 2\sum_{j=2}^{k+1} q_j(1-q_j)E[(\beta_j+g_1\beta_{1j})^2] + \sigma^2 \\ &= 2(\sigma_J^2 + g_1^2\sigma_{1J}^2)\sum_{j=2}^{k+1} q_j(1-q_j) + \sigma^2 \end{aligned}$$

and therefore we could still expect $V(y|g_1=2) > V(y|g_1=1) > V(y|g_1=0)$ in practice where the effects are polygenic.

Polygenic epistasis with multiple apparent vQTL If we further extend the model to

$$y = \mu + \sum_{i=1}^{m} g_{1i}\beta_{1i} + \sum_{i=1}^{m}\sum_{j=2}^{k+1}(g_j\beta_j + g_{1i}g_j\beta_{1ij}) + e$$

with m vQTL, rather than only locus 1, interacting with the other k loci, then

$$V(y|g_{1i}) = 2\sum_{i=1}^{m}\sum_{j=2}^{k+1} q_j(1-q_j)(\beta_j + g_{1i}\beta_{1ij})^2 + \sigma^2$$

where $V(y|g_{1i})$ does not depend only on the genotypes of a single locus. There are in total 2^m vQTL haplotypes, and the phenotypic variance for each particular haplotype approximates

$$\begin{aligned} & E_\beta[V(y|g_{11},g_{12},\cdots,g_{1m})] \\ = \; & E_\beta\left[2\sum_{i=1}^{m}\sum_{j=2}^{k+1} q_j(1-q_j)(\beta_j+g_{1i}\beta_{1ij})^2 + \sigma^2\right] \\ = \; & 2\sum_{i=1}^{m}(\sigma_J^2 + g_{1i}^2\sigma_{1iJ}^2)\sum_{j=2}^{k+1} q_j(1-q_j) + \sigma^2 \end{aligned}$$

given that for each j, $\beta_j \sim N(0,\sigma_j^2)$ and $\beta_{1ij} \sim N(0,\sigma_{1iJ}^2)$, where β_j and β_{1ij} are independent, and β stands for the set of $\{\beta_j,\beta_{1ij}\}$. Although obviously, $E_\beta[V(y|g_{11}\neq 0, g_{12}\neq 0,\cdots,g_{1m}\neq 0)] > E_\beta[V(y|g_{11}=g_{12}=\cdots=g_{1m}=0)]$, these two extreme haplotypes are likely to be rare or absent in a real population, as are many other haplotypes if m is large. Furthermore, if $g_{1i} \sim B(2,f_{1i})$, for each individual that has randomly sampled genotypes for the m vQTL, the variance quantity above is approximately

$$\begin{aligned} & E_{g_{1i}}[E_\beta[V(y|g_{11},g_{12},\cdots,g_{1m})]] \\ = \; & 2\sum_{i=1}^{m}(\sigma_J^2 + E[g_{1i}^2]\sigma_{1iJ}^2)\sum_{j=2}^{k+1} q_j(1-q_j) + \sigma^2 \\ = \; & 2\sum_{i=1}^{m}(\sigma_J^2 + 2q_{1i}(1-q_{1i})\sigma_{1iJ}^2)\sum_{j=2}^{k+1} q_j(1-q_j) + \sigma^2 \end{aligned}$$

Thus, for such a polygenic scenario, because each locus is involved with interactions among very many loci and signs of these are likely to vary, any genetically regulated variance heterogeneity among individuals is likely to be small and undetectable. Indeed, this variance will merely comprise part of the epistatic variance, if indeed that could be estimated.

Simulation Based on the *Arabidopsis thaliana* genome data (250K SNP array) of 100 randomly selected lines from the RegMap population (Hancock et al., 2011), we simulated phenotypes from polygenic epistatic models and estimated the heritability of the residual variance. The phenotype was simulated as

$$\mathbf{y} = \mu + \mathbf{g}_{add} + \mathbf{g}_{epi} + \mathbf{e}$$

where $\mathbf{g}_{add} \sim N(\mathbf{0}, \mathbf{G}\sigma^2_{add})$, $\mathbf{g}_{epi} \sim N(0, \mathbf{G} \circ \mathbf{G}\sigma^2_{epi})$, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$, and $\circ$ represents the Hadamard (element-wise) product between matrices. $\mathbf{G}$ is the genomic kinship matrix constructed using all the available SNPs and weighted by the allele frequencies. In the simulation, σ^2_{add}, σ^2_{epi} and σ^2 were set to 1. Using the **hglm** package (Rönnegård et al., 2010b), a linear mixed model including the additive genetic effects

$$\mathbf{y} = \mu + \mathbf{g}_{add} + \mathbf{e}$$

was fitted, and the estimated deviances (squared residuals) were $\hat{\mathbf{d}}$. A second linear mixed model

$$\hat{\mathbf{d}} = \nu + \mathbf{g}_{var} + \mathbf{e}^*$$

where ν is an intercept, $\mathbf{g}_{var} \sim N(\mathbf{0}, \mathbf{G}\sigma^2_{var})$ and $\mathbf{e}^* \sim N(\mathbf{0}, \mathbf{I}\sigma^{*2})$, was fitted to estimate the heritability of the residual variance (variance heterogeneity), defined as

$$h^2_{var} = \frac{\hat{\sigma}^2_{var}}{\hat{\sigma}^2_{var} + \hat{\sigma}^{*2}}$$

Such a simulation procedure was repeated 30 times. The mean heritability for the variance was 0.0044 with standard error 0.0031 and $P = 0.17$ from a student t-test against 0. This agreed with our theoretical analysis that such epistatic effects could not explain variation in phenotypic variability, and that selection for variance heterogeneity caused by polygenic epistasis is unlikely to be effective, even though a large number of interacting "vQTL" do exist.

Conclusion

Assuming underlying epistatic models, we showed that selection for phenotypic variability or uniformity may be successful only when a small number of vQTL are involved. Polygenic scenarios likely even out the genetic regulation of phenotypic variability due to polygenic epistasis. There is no similar constraint on the potential effectiveness on selection to utilize individual or multi-locus differences in the environmental and thus phenotypic variance due to loci which affect the variance directly. Such selection can act both at the individual or multi-locus level just as selection on trait mean (Hill et al., 2004; Mulder et al., 2007). Our results motivate future discoveries of the genetic basis in populations with relatively high heritability of residual variance (*e.g.* Rönnegård et al., 2013), and potentially also explain why such selection may not succeed in practice.

Literature Cited

Geiler-Samerotte, K. A., Bauer, C. R., Li, S., et al. (2013). In: *Current Opinion in Biotechnology*, pp. 1–8.

Hancock, A. M., Brachi, B., Faure, N., et al. (2011). In: *Science* 334, pp. 83–86.

Hill, W. G. and Zhang, X.-S. (2004). In: *Genetics Research, Cambridge* 83, pp. 121–132.

Hill, W. G. and Mulder, H. A. (2010). In: *Genetics Research, Cambridge* 92, pp. 381–395.

Jimenez-Gomez, J. M., Corwin, J. A., Joseph, B., et al. (2011). In: *PLoS Genetics* 7, e1002295.

Mackay, T. F. C. and Lyman, R. F. (2005). In: *Philosophical Transactions of the Royal Society, Series B* 360, pp. 1513–1527.

Mulder, H. A., Bijma, P., and Hill, W. G. (2007). In: *Genetics* 175, pp. 1895–1910.

Paré, G., Cook, N. R., Ridker, P. M., et al. (2010). In: *PLoS Genetics* 6, e1000981.

Rönnegård, L. and Valdar, W. (2010a). In: *Genetics* 188, pp. 435–447.

Rönnegård, L., Shen, X., and Alam, M. (2010b). In: *The R Journal* 2, pp. 20–28.

Rönnegård, L., Felleki, M., Fikse, F., et al. (2013). In: *Journal of Dairy Science* 96, pp. 2627–2636.

Shen, X., Pettersson, M., Rönnegård, L., et al. (2012). In: *PLoS Genetics* 8, e1002839.

Struchalin, M. V., Dehghan, A., Witteman, J. C. M., et al. (2010). In: *BMC Genetics* 11, p. 92.

Wolc, A., Arango, J., Settar, P., et al. (2012). In: *Animal Genetics* 43 Suppl 1, pp. 87–96.

Yang, J., Loos, R. J. F., Powell, J. E., et al. (2012). In: *Nature* 490, pp. 267–272.