

Do Molecular Markers Inform About Pleiotropy?

Daniel Gianola,^{*,†,‡,1} Gustavo de los Campos,[§] Miguel A. Toro,^{**} Hugo Naya,^{††}

Chris-Carolin Schön,^{†,‡} and Daniel Sorensen^{**}

^{*}Departments of Animal Sciences, Dairy Science, and Biostatistics and Medical Informatics, University of Wisconsin–Madison, Wisconsin 53706, [†]Department of Plant Sciences Technical University of Munich, Center for Life and Food Sciences, D-85354 Freising-Weihenstephan, Germany, [‡]Institute of Advanced Study, Technical University of Munich, D-85748 Garching, Germany, [§]Department of Biostatistics, Michigan State University, East Lansing, Michigan 48824, ^{**}Escuela Técnica Superior de Ingenieros Agrónomos, Universidad Politécnica de Madrid, 20840 Madrid, Spain, ^{††}Institut Pasteur de Montevideo, Mataojo 2020, Montevideo 11400, Uruguay, and ^{††}Department of Molecular Biology and Genetics, Aarhus University, DK-8000 Aarhus C, Denmark

ABSTRACT The availability of dense panels of common single-nucleotide polymorphisms and sequence variants has facilitated the study of statistical features of the genetic architecture of complex traits and diseases via whole-genome regressions (WGRs). At the onset, traits were analyzed trait by trait, but recently, WGRs have been extended for analysis of several traits jointly. The expectation is that such an approach would offer insight into mechanisms that cause trait associations, such as pleiotropy. We demonstrate that correlation parameters inferred using markers can give a distorted picture of the genetic correlation between traits. In the absence of knowledge of linkage disequilibrium relationships between quantitative or disease trait loci and markers, speculating about genetic correlation and its causes (e.g., pleiotropy) using genomic data is conjectural.

KEYWORDS genetic correlation; genomic correlation; genomic heritability; linkage disequilibrium; pleiotropy

THE interindividual differences for a trait or disease risk that can be explained by genetic factors, such as trait heritability (h^2), the genetic correlation (r_G), and the coheritability between two traits ($r_{Gh_1h_2}$), are very important parameters in quantitative genetic studies of animals, humans, and plants. These quantities play a role in the study of evolution due to artificial and natural selection, and knowledge thereof is required for statistical prediction of outcomes in animal and plant breeding as well as medicine. Traditionally, these parameters have been estimated using phenotypes and pedigrees, e.g., family and twin data in human genetics. The availability of dense panels of common single-nucleotide polymorphisms (SNPs) and of sequence data more recently has made it possible to assess kinship among distantly related individuals (Morton *et al.* 1971; Thompson 1975; Ritland

1996; Lynch and Ritland 1999). This development has opened new opportunities for study of the genetic architecture of complex traits and diseases. For instance, Yang *et al.* (2010) suggested using whole-genome regressions (WGRs) (Meuwissen *et al.* 2001) to assess the proportion of variance of a trait or disease risk that can be explained by a regression of phenotypes on common SNPs or genomic heritability and a related parameter, the “missing heritability.” More recently, WGR models have been extended for the analysis of systems of multiple traits, so the concept of *genomic correlation* also has entered into the picture (Jia and Jannink 2012; Lee *et al.* 2012). For instance, Maier *et al.* (2015) used multivariate WGR models and reported estimates of genetic correlations between psychiatric disorders, and Furlotte and Eskin (2015) presented a methodology that incorporates genetic marker information for the analysis of multiple traits that, according to the authors, “provide fundamental insights into the nature of co-expressed genes.” In a similar spirit, Korte *et al.* (2012) argued that multitrait-marker-enabled regressions can be useful for understanding pleiotropy. More recently, Bulik-Sullivan *et al.* (2015) proposed a methodology for “estimating genetic correlation” using statistics derived from single-marker genome-wide

Copyright © 2015 by the Genetics Society of America
doi: 10.1534/genetics.115.179978

Manuscript received June 26, 2015; accepted for publication July 21, 2015; published Early Online July 23, 2015.

¹Corresponding author: Department of Animal Sciences, University of Wisconsin–Madison, 1675 Observatory Drive, Madison, WI 53706.
E-mail: gianola@ansci.wisc.edu

association studies (GWAS) and reported estimates of such correlations among 25 human traits.

de los Campos *et al.* (2015) discussed potential problems that emerge when trying to infer genetic parameters using molecular markers that are imperfectly associated with the genotypes at the causal loci. In this paper, the framework described in de los Campos *et al.* (2015) is extended for the analysis of systems of traits, and it is demonstrated that correlation parameters inferred using markers can give a distorted picture of the genetic correlation between traits. For instance, it is shown that an analysis based on markers may suggest a genetic correlation when none exists or may fail to detect a genetic correlation when one does exist. It is concluded that in the absence of knowledge about linkage disequilibrium (LD) relationships between quantitative trait loci (QTL) and markers, speculating about genetic correlations, and even more about their causes (e.g., pleiotropy), using genomic data is conjectural.

Theory

To set the stage, consider a single-locus model. In an additive-inheritance framework, a phenotype (y) is regressed on a QTL genotype code Q (0, 1, and 2 for genotypes aa , Aa , and AA , respectively) according to the linear model

$$y = \alpha_0 + \alpha_1 Q + E \quad (1)$$

where α_0 and α_1 are fixed parameters, and Q and E are independent random variables, the latter representing a model residual. The proportion of phenotypic variance explained by the linear regression on Q , or narrow-sense heritability, is

$$h^2 = \frac{\alpha_1^2 \Sigma_Q}{\alpha_1^2 \Sigma_Q + \sigma_E^2}$$

where $\Sigma_Q = \text{var}(Q)$ is the variance in allelic content, and $\sigma_E^2 = \text{var}(E)$ is the residual variance. If Q is standardized to a unit variance,

$$\Sigma_Q = 1 \quad \text{and} \quad h^2 = \frac{\alpha_1^2}{\alpha_1^2 + \sigma_E^2}$$

In quantitative genomic analysis, marker genotypes (X) are used in lieu of the QTL genotypes Q because the latter are unknown or unobserved. The marker-based or instrumental model, assuming a single marker, is a linear regression on marker genotype X with form

$$y = \alpha_0 + \beta_1 X + E' \quad (2)$$

where E' is a regression residual. Assuming without loss of generality that both X and Q are in standard deviation units, the marker effect can be shown to be $\beta_1 = \rho_{QX} \alpha_1$, where ρ_{QX} is the correlation between the marker and the QTL genotypes, which depends on their LD. In this setting, the proportion of variance of phenotypes explained by the linear regression on the marker, or genomic heritability, is $h_{\text{marked}}^2 = \rho_{QX}^2 h^2$, and

missing heritability is $h_{\text{missing}}^2 = (1 - \rho_{QX}^2) h^2$. Hence, missing heritability is a function of the LD between the marker and the QTL. Genomic heritability has h^2 as an upper bound (de los Campos *et al.* 2015).

The regression model just described can be extended to the analysis of multiple traits affected by multiple QTL. For simplicity, we consider only two markers (X_1 and X_2) and two QTL (Q_1 and Q_2). A multivariate representation of the model with an arbitrary number of QTL and markers is provided in the Appendix. Figure 1 depicts a system with two traits, two QTL, and two markers. The left panel represents the regression of the phenotypes on the two QTL, with blue arrows denoting effects from QTL on traits and green arcs denoting LD between QTL. In the QTL model of Figure 1, the genetic correlation is (see Appendix)

$$r_G = \frac{\alpha_1' \Sigma_Q \alpha_2}{\sqrt{(\alpha_1' \Sigma_Q \alpha_1)(\alpha_2' \Sigma_Q \alpha_2)}} \quad (3)$$

where $\alpha_1' = (\alpha_{11} \quad \alpha_{12})$ contains the effects of QTL 1 and 2 on trait 1, and $\alpha_2' = (\alpha_{21} \quad \alpha_{22})$ contains the effects of QTL 1 and 2 on trait 2. The variance-covariance matrix between QTL genotypes is given by Σ_Q . If genotypes are standardized,

$$\Sigma_Q = \begin{bmatrix} 1 & \rho_{Q_{12}} \\ \rho_{Q_{21}} & 1 \end{bmatrix}$$

with $\rho_{Q_{12}}$ being the correlation between genotypes at QTL 1 and 2. In the QTL model of Figure 1, there are two sources of genetic correlation: pleiotropy (*i.e.*, the same QTL affects more than one trait) and LD between QTL, in this case represented by $\rho_{Q_{12}} \neq 0$. This is well known in quantitative genetics (Falconer and Mackay 1996; Knott and Haley 2000).

We now bring the two markers into the picture, as shown in the right panel of Figure 1; here gray arrows are regressions on markers (these are distinct from regressions on QTL genotypes), and arcs denote correlations between genotypes due to LD. In the Appendix, we show that the genomic correlation is

$$r_{G,\text{marked}} = \frac{\alpha_1' \Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ} \alpha_2}{\sqrt{(\alpha_1' \Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ} \alpha_1)(\alpha_2' \Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ} \alpha_2)}} \quad (4)$$

In this expression, Σ_{QX} is the covariance matrix between QTL and marker genotypes (reflecting marker-QTL LD), and Σ_X is the covariance matrix between marker genotypes, reflecting mutual LD relationships among markers. If markers and genotypes are in standard deviation units,

$$\Sigma_X = \begin{bmatrix} 1 & \rho_{X_{12}} \\ \rho_{X_{21}} & 1 \end{bmatrix} \quad \text{and} \quad \Sigma_{QX} = \begin{bmatrix} \rho_{Q_1 X_1} & \rho_{Q_1 X_2} \\ \rho_{Q_2 X_1} & \rho_{Q_2 X_2} \end{bmatrix}$$

Comparison of the genomic correlation (4) with the genetic correlation (3) indicates that in $r_{G,\text{marked}}$, $\Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ}$ replaces Σ_Q . Inspection of (4) reveals that the sources of the genomic correlation are (1) pleiotropic QTL effects via α_1 and α_2 , (2)

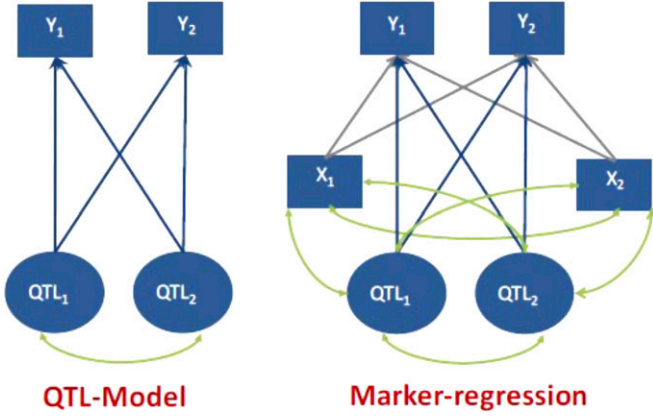


Figure 1 Two-trait system. A system of two traits (Y) involving two QTL and two markers (X). Single-pointed blue arrows denote causal effects, green double-pointed arrows denote LD, and single-pointed gray arrows represent regression coefficients.

marker-QTL LD patterns conveyed by Σ_{QX} , and (3) among-marker LD relationships, as conveyed by Σ_X . Notably, one of the sources of genetic correlation, *i.e.*, LD between QTL, as conveyed by Σ_Q , has no effect on $r_{G,\text{marked}}$. Conversely, there are sources that contribute to the genomic correlation, *i.e.*, marker-marker and marker-QTL LD, that do not enter into r_G .

Because the sources affecting genetic and genomic correlations are distinct, the two parameters can differ greatly. This point is strengthened by considering four stylized cases represented in Figure 2. All the demonstrations supporting the discussion that follows can be found in the Appendix.

Application to Four Situations

Case 1: Independent marker-QTL pairs and absence of pleiotropy (Figure 2, upper-left panel)

This is the simplest case: it consists of two marker-QTL pairs with linkage equilibrium (LE) between pairs but LD within pairs. Each trait is affected by only one QTL; QTL 1 affects trait 1, and QTL 2 affects trait 2. Several simplifications take place here. For instance, because of LE between pairs, $\rho_{Q_1X_2} = 0$, so Σ_{QX} becomes an identity matrix. Therefore, the genetic covariance in the numerator of (3) reduces to $\alpha'_1\alpha_2$. In the absence of pleiotropy, $\alpha'_1 = (\alpha_{11} \ 0)$ and $\alpha'_2 = (0 \ \alpha_{22})$ are orthogonal; *i.e.*, $\alpha'_1\alpha_2 = 0$. Therefore, the genetic correlation is null. Furthermore, with LE between pairs, $\rho_{Q_1X_2} = \rho_{Q_2X_1} = \rho_{X_1X_2} = 0$, leading also to an absence of genomic correlation. Thus, in case 1 there is complete agreement between the genomic and genetic correlations: both are null.

Case 2: Phantom correlation (Figure 2, upper-right panel)

The setting is obtained by adding LD between the two markers to case 1. There is no pleiotropy, and the two QTL are in LE, so the genetic correlation is zero (genetically, the system is equivalent to case 1). However, because of the LD between markers, $\Sigma_{QX}\Sigma_X^{-1}\Sigma_{XQ}$ in (4) is no longer diagonal. Consequently, there will be nonzero genomic correlation even in absence of genetic

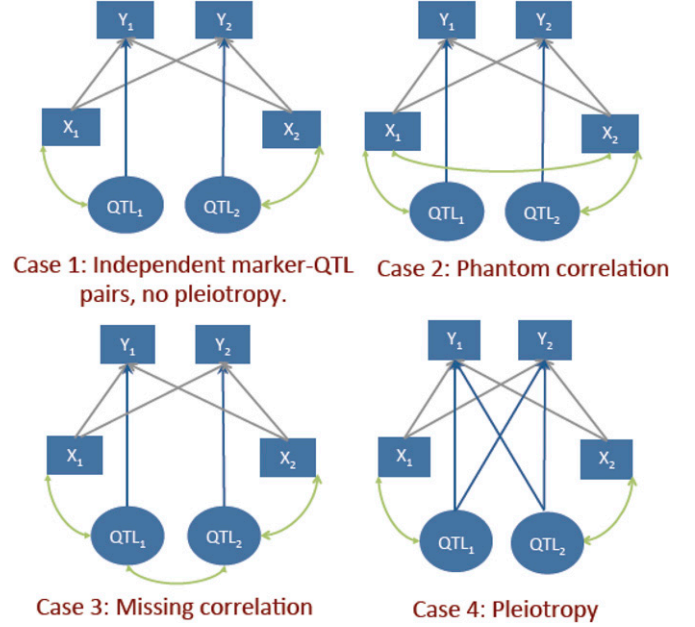


Figure 2 Two-trait system. Four possible cases of interplay between QTL, markers, and phenotypes. The arrows have the same interpretation as in Figure 1.

correlation: markers can induce genomic correlation when traits are genetically uncorrelated—a crucial issue.

Case 3: Missing correlation (Figure 2, lower-left panel)

This scenario illustrates a situation in which the genetic correlation is undetected by the markers and is obtained from case 1 by adding LD between QTL, which, in the absence of pleiotropy, is the only source of genetic correlation between traits. However, $\Sigma_{QX}\Sigma_X^{-1}\Sigma_{XQ}$ remains diagonal as in case 1. Furthermore, in the absence of pleiotropy, $\alpha'_1\alpha_2 = 0$ (orthogonality); consequently, $r_{G,\text{marked}}$ is null. This example shows how one source of genetic correlation, namely, LD among QTL, may be completely lost in a genomic analysis.

Case 4: Pleiotropy (Figure 2, lower-right panel)

Here we allow each of the two QTL to affect both traits; otherwise, the setting is as in case 1. Pleiotropy now induces a genetic and a genomic correlation. However, r_G and $r_{G,\text{marked}}$ differ in magnitude depending on the patterns of LD and on the magnitude of the pleiotropic effects. To illustrate, we set $\Sigma_X = \Sigma_Q = \mathbf{I}_2$, an identity matrix of order 2; this implies LE between pairs of QTL and pairs of markers. Further, we take

$$\Sigma_{QX} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \quad \text{or} \quad \Sigma_{QX} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.8 \end{bmatrix}$$

i.e., homogeneity or heterogeneity of marker-QTL LD, respectively. Finally, QTL effects are set to $\alpha'_1 = (1 \ \alpha_{12})$ and $\alpha'_2 = (\alpha_{21} \ 1)$, with $\alpha_{12} = \alpha_{21}$; this pleiotropic effect was varied over the set of values $[-0.9, -0.8, \dots, 0.8, 0.9]$. Figure 3 displays the resulting values of the genomic (vertical axis) vs. genetic (horizontal axis) correlations computed using (3) and (4); the blue curve represents the case where

marker-QTL LD was the same for both pairs, and the red curve represents the case where LD differed between pairs 1 and 2. The figure shows how different patterns of LD induce different magnitudes of genomic and genetic correlations that, however, do not differ in sign in this example.

The genomic covariance does not always preserve the sign of the genetic covariance. Suppose that the two QTL are not pleiotropic but are in LD, with effects $\alpha'_1 = (\alpha \ 0)$ and $\alpha'_2 = (0 \ -\alpha)$ and with

$$\Sigma_Q = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

Using the expression in the numerator of (3), the genetic covariance is

$$(\alpha \ 0) \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} (0 \ -\alpha)' = -\frac{\alpha^2}{2}$$

which is negative at any nonnull value of α . Now let the LD relationships between markers and between QTL and markers be such that

$$\Sigma_X = \begin{bmatrix} 1 & \frac{4}{5} \\ \frac{4}{5} & 1 \end{bmatrix} \quad \text{and} \quad \Sigma_{QX} = \begin{bmatrix} \frac{4}{5} & 0 \\ 0 & \frac{4}{5} \end{bmatrix}$$

The genetic system is such that QTL 1 (QTL 2) is in LD with marker 1 (marker 2), but there is LE between QTL 1 and marker 2 and QTL 2 and marker 1. In the numerator of expression (4),

$$\Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ} = \begin{bmatrix} \frac{16}{9} & -\frac{64}{45} \\ -\frac{64}{45} & \frac{16}{9} \end{bmatrix}$$

and the genomic covariance is $(64/45)\alpha^2$, always positive. In this example, the genomic correlation is $4/5$, and the genetic correlation is $-1/2$.

Discussion

In the analysis of systems of complex traits, none of the cases just discussed are likely to “hold” exactly as described, and there is an enormous range of possibilities in terms of within and between marker-QTL genotypes as well as allelic effects sizes and signs. However, the underlying mechanisms that our examples describe are an integral part of the multivariate system involving QTL and markers and are key to an understanding of why genomic and genetic correlations are distinct parameters. Importantly, there is an ambiguous link between the two parameters. For instance, all or a fraction of the

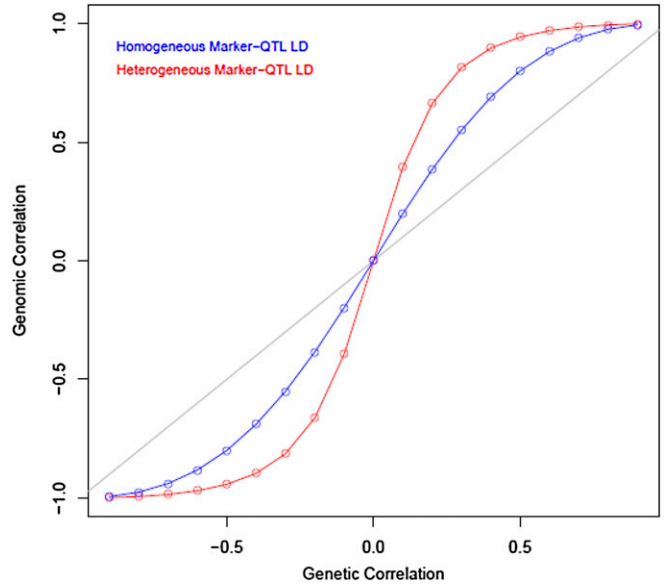


Figure 3 Genomic vs. genetic correlation in the system described by case 4.

component of r_G that is due to LD among QTL is likely to be missed by an analysis based on markers that are in imperfect LD with QTL. Also, a fraction of the genetic correlation due to pleiotropy is likely to be missed as a result of imperfect LD between marker and QTL. Finally, LD between markers can create illusory genetic correlations.

What happens if all QTL genotypes are included in the panel of markers, as may be expected if full DNA sequence information is available? Here the sequence can be partitioned into neutral markers (\mathbf{x}) and QTL (\mathbf{q}) such that for a given individual the genomic data presents as $\mathbf{x}_{\text{seq}} = (\mathbf{x}', \mathbf{q}')$. Thus, the *sequence* covariance matrix is

$$\text{var}(\mathbf{x}_{\text{seq}}) = \Sigma_{\mathbf{x}_{\text{seq}}} = \begin{bmatrix} \Sigma_X & \Sigma_{XQ} \\ \Sigma_{QX} & \Sigma_Q \end{bmatrix} \quad (5)$$

The marked genotype for trait i using the DNA sequence is

$$\hat{G}_i = \alpha'_i \Sigma_{Q\mathbf{x}_{\text{seq}}} \Sigma_{\mathbf{x}_{\text{seq}}}^{-1} \mathbf{x}_{\text{seq}}, \quad i = 1, 2 \quad (6)$$

and the genomic or marked covariance is

$$\begin{aligned} \text{cov}(\hat{G}_1, \hat{G}_2) &= \alpha'_1 \Sigma_{Q\mathbf{x}_{\text{seq}}} \Sigma_{\mathbf{x}_{\text{seq}}}^{-1} \Sigma_{\mathbf{x}_{\text{seq}}} \Sigma_{\mathbf{x}_{\text{seq}}}^{-1} \Sigma_{\mathbf{x}_{\text{seq}}Q} \alpha_2 \\ &= \alpha'_1 \Sigma_{Q\mathbf{x}_{\text{seq}}} \Sigma_{\mathbf{x}_{\text{seq}}}^{-1} \Sigma_{\mathbf{x}_{\text{seq}}Q} \alpha_2 \end{aligned} \quad (7)$$

Using partitioned matrix techniques for obtaining the inverse of $\Sigma_{\mathbf{x}_{\text{seq}}}$ de los Campos *et al.* (2015) showed that

$$\Sigma_{Q\mathbf{x}_{\text{seq}}} \Sigma_{\mathbf{x}_{\text{seq}}}^{-1} \Sigma_{\mathbf{x}_{\text{seq}}Q} = [\Sigma_{QX} \quad \Sigma_Q] \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} = \Sigma_Q \quad (8)$$

Hence, $\text{cov}(\hat{G}_1, \hat{G}_2) = \alpha'_1 \Sigma_Q \alpha_2$, the genetic covariance defined in equation (A2) in the Appendix. This shows that if the sequence information contains the variants at the causal loci, the marked covariance is equal to the genetic covariance

between traits, and therefore, the genomic correlation is identical to the genetic correlation in that case. However, the genetic correlation depends on allelic frequencies and allele effect sizes at the QTL as well as on LD relationships between the QTL; these parameters, as well as the trait-specific QTL, will still need to be learned properly. Apart from finite-sample-size statistical problems, technical issues such as a large percentage of singleton reads and incomplete gene coverage will complicate matters (Kerr Wall 2009). Hence, when sequence data become available for quantitative genetic studies, unraveling the structure of the genetic correlation will not be an easy task, even under the simplifying assumptions of an additive model of inheritance.

In conclusion, multivariate quantitative genetic analysis based on markers can be used to obtain more accurate predictions of complex traits and to estimate genomic correlations. However, these parameters cannot always be viewed as genetic correlations because the sources of genetic and genomic correlations are distinct. Imperfect LD between markers and QTL produces missing heritability in single-trait analysis; in multivariate models, the problem becomes one of missing, excessive, or spurious (MES) correlation. Care must be exercised when interpreting estimates of genomic correlations between complex traits when these traits are assessed by molecular markers as opposed to QTL and even more so when interpreted from a causality perspective. Unfortunately, considerably more information is needed than what is now available for a meaningful interpretation of estimates of genomic correlations between pairs of traits when gene action involves many additive QTL. Speculating on the multivariate statistical genetic architecture of complex traits using imperfect instruments such as markers seems risky at this time.

Acknowledgments

This work was supported in part by the Wisconsin Agriculture Experiment Station and by a U.S. Department of Agriculture Hatch Grant (142-PRJ63CV) to D.G. C.C.S. and D.G. acknowledge support of the Technische Universität München Institute for Advanced Study, funded by the German Excellence Initiative. G.D.L.C. received support from National Institutes of Health grants R01GM099992 and R01GM101219. M.A.T. wishes to acknowledge funding

from the European Union's Seventh Framework Programme (KBBE.2013.1.2-10) under grant agreement 61361.

Literature Cited

- Bulik-Sullivan, B., H. K. Finucane, V. Anttila, A. Gusev, F. R. Day *et al.* 2015 An atlas of genetic correlations across human diseases and traits. *bioRxiv* doi: <http://dx.doi.org/10.1101/014498>.
- de los Campos, G., D. Sorensen, and D. Gianola, 2015 Genomic heritability: what is it? *PLoS Genet.* 11: e1005048.
- Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Longmans Green, Harlow, UK.
- Furlotte, N. A., and E. Eskin, 2015 Efficient multiple-trait association and estimation of genetic correlation using the matrix-variate linear mixed model. *Genetics* 200: 59–68.
- Knott, S. A., and C. S. Haley, 2000 Multitrait least squares for quantitative trait loci detection. *Genetics* 156: 899–911.
- Jia, Y., and J.-L. Jannink, 2012 Multiple trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192: 1513–1522.
- Korte, A., B. J. Vilhjálmsson, V. Segura, A. Platt, Q. Long, and M. Nordborg, 2012 A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* 44: 1066–1071.
- Lee, S. H., J. Yang, M. E. Goddard, P. M. Visscher, and N. R. Wray, 2012 Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 28: 2540–2542.
- Lynch, M., and K. Ritland, 1999 Estimation of pairwise relatedness with molecular markers. *Genetics* 152: 1753–1766.
- Maier, R. *et al.*, 2015 Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am. J. Hum. Genet.* 96: 283–294.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Morton, N. E., S. Yee, D. E. Harris, and R. Lew, 1971 Bioassay of kinship. *Theor. Popul. Biol.* 2: 507–524.
- Ritland, K., 1996 A marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* 50: 1062–1073.
- Thompson, E. A., 1975 The estimation of pairwise relationships. *Ann. Hum. Genet.* 39: 173–188.
- Kerr Wall, P., 2009 Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics* 10: 347.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of heritability for human height. *Nat. Genet.* 42: 565–569.

Communicating editor: G. A. Churchill

Appendix

Genetic Correlation

Let $G_1 = \alpha_1' \mathbf{q}$ and $G_2 = \alpha_2' \mathbf{q}$ be additive genetic values for a pair of traits, where α_1 and α_2 are vectors of fixed allelic substitution effects affecting traits 1 and 2, respectively, and \mathbf{q} is a random vector indicating the incidence of genotypes at the corresponding QTL. Following de los Campos *et al.* (2015), the additive genetic variance of trait i is

$$\text{var}(G_i) = \alpha_i' \Sigma_Q \alpha_i, \quad i = 1, 2 \quad (\text{A1})$$

The additive genetic covariance between traits 1 and 2 is then

$$\text{cov}(G_1, G_2) = \alpha_1' \Sigma_Q \alpha_2 \quad (\text{A2})$$

where Σ_Q is a covariance matrix between allelic contents at loci affecting the traits. For example, with two QTL (assuming Hardy-Weinberg equilibrium at each of the two QTL),

$$\Sigma_Q = \begin{bmatrix} 2p_1(1-p_1) & 2D_{12} \\ 2D_{12} & 2p_2(1-p_2) \end{bmatrix} \quad (\text{A3})$$

where p_j is the frequency of the reference allele at locus j ($j = 1, 2$), and D_{12} is the LD statistic between alleles at the two loci. In scalar notation, (A2) takes the more explicit form

$$\text{cov}(G_1, G_2) = 2[p_1(1-p_1)\alpha_{11}\alpha_{21} + p_2(1-p_2)\alpha_{12}\alpha_{22}] + 2D_{12}(\alpha_{12}\alpha_{21} + \alpha_{11}\alpha_{22}) \quad (\text{A4})$$

The genetic covariance has a pleiotropy component (the first part of the expression) plus a LD component that vanishes if the QTL are in pairwise equilibrium, *i.e.*, $D_{12} = 0$. The genetic correlation (Falconer and Mackay 1996) is

$$r_G = \frac{\alpha_1' \Sigma_Q \alpha_2}{\sqrt{(\alpha_1' \Sigma_Q \alpha_1)(\alpha_2' \Sigma_Q \alpha_2)}} \quad (\text{A5})$$

Genomic Correlation

Let \mathbf{x} be a vector of genotypes at p marker loci. The multiple linear regressions of G_1 and G_2 on \mathbf{x} produce as fitted values $\hat{G}_i = \alpha_i' \Sigma_{QX} \Sigma_X^{-1} \mathbf{x}$. The genomic covariance (or marked genetic covariance) is defined as

$$C_{\text{marked}} = \text{cov}(\hat{G}_1, \hat{G}_2) = \alpha_1' \Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ} \alpha_2 \quad (\text{A6})$$

The *genomic correlation* is

$$r_{G,\text{marked}} = \frac{C_{\text{marked}}}{\sqrt{(\alpha_1' \Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ} \alpha_1)(\alpha_2' \Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ} \alpha_2)}} \quad (\text{A7})$$

Interpreting this parameter meaningfully requires knowledge of (1) bivariate QTL effects at all loci, (2) LD relationships between QTL affecting the two traits and the markers via the Σ_{QX} matrices, and (3) LD relationships among markers. Unfortunately, only phenotypes, marker genotypes, and LD relationships between markers are observable. Most of the required ingredients in the formula are yet unknown. Importantly, note that Σ_Q , conveying LD between QTL, does not enter into the genomic correlation.

Independent QTL-Marker Blocks (Case 1 in Figure 2)

Each of two independently segregating QTL is in LD with a marker, with the two markers being in mutual LE, and there is no pleiotropy. Here

$$\Sigma_{QX} = \Sigma_{XQ} = \Sigma_X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (\text{A8})$$

so the genetic and genomic correlations both become

$$r_G = r_{G,\text{marked}} = \frac{\alpha_1' \alpha_2}{\sqrt{(\alpha_1' \alpha_1)(\alpha_2' \alpha_2)}} \quad (\text{A9})$$

Because there is no pleiotropy, $\alpha_1' \alpha_2 = (\alpha_{11} \ 0)'(0 \ \alpha_{22}) = 0$, and both correlations are null.

Phantom Correlation (Case 2 in Figure 2)

Consider $\Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ}$ in (A6), where (given standardized genotypes)

$$\begin{aligned} \Sigma_{QX} &= \begin{bmatrix} \rho_{Q_1 X_1} & \rho_{Q_1 X_2} \\ \rho_{Q_2 X_1} & \rho_{Q_2 X_2} \end{bmatrix} = \begin{bmatrix} \rho_{Q_1 X_1} & 0 \\ 0 & \rho_{Q_2 X_2} \end{bmatrix} \\ \Sigma_X^{-1} &= \begin{bmatrix} 1 & \rho_{X_1 X_2} \\ \rho_{X_1 X_2} & 1 \end{bmatrix}^{-1} = \frac{1}{1 - \rho_{X_1 X_2}^2} \begin{bmatrix} 1 & -\rho_{X_1 X_2} \\ -\rho_{X_1 X_2} & 1 \end{bmatrix} \\ \Sigma_{XQ} &= \begin{bmatrix} \rho_{Q_1 X_1} & 0 \\ 0 & \rho_{Q_2 X_2} \end{bmatrix} \end{aligned} \quad (\text{A10})$$

Then

$$\Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ} = \frac{1}{1 - \rho_{X_1 X_2}^2} \begin{bmatrix} \rho_{Q_1 X_1}^2 & -\rho_{Q_1 X_1} \rho_{Q_2 X_2} \rho_{X_1 X_2} \\ -\rho_{Q_1 X_1} \rho_{Q_2 X_2} \rho_{X_1 X_2} & \rho_{Q_2 X_2}^2 \end{bmatrix} \quad (\text{A11})$$

The off-diagonals of this matrix are nonnull, so a genomic correlation will arise when there is no genetic correlation.

Missing Correlation (Case 3 in Figure 2)

Because the markers are in LE, Σ_X^{-1} is an identity matrix, so

$$\Sigma_{QX} \Sigma_X^{-1} \Sigma_{XQ} = \Sigma_{QX} \Sigma_{XQ} = \begin{bmatrix} \rho_{QX}^2 & 0 \\ 0 & \rho_{QX}^2 \end{bmatrix}$$

Therefore, in the absence of pleiotropy, C_{marked} in (A6) and, thus, $r_{G,\text{marked}}$ will be null no matter what the value of the genetic correlation.

Pleiotropy (Case 4 in Figure 2)

The results in Figure 3 were obtained using expressions (3) and (4) with the parameter values described in the main body of the paper.