Short Note: More Generalised Estimation of Between-trait Genetic Correlations Using Data from Collateral Relatives

By R. D. BURDON[1]) and L. A. APIOLAZA[2])

(Received 7th January 1998)

Summary

Sib-family or clonal data are commonly used to estimate genetic correlations between traits. Usually both traits of a pair are measured on the entire genetic sample, but quite often the respective traits are measured on independent samples from within the genetic groups. Using a method based on analysis of variance, the estimation procedure is generalised to include cases where the two traits are measured on partially overlapping subsamples, provided group size and degree of overlap are constant. This has a clear potential for improving cost-efficiency, especially for pairs of traits that differ widely in heritability and/or costs of measurement.

Key words: Genetic correlation, estimation, sampling, cost-efficiency.

FDC: 165.3.

Introduction

For estimating genetic correlations between traits half-sib family data are widely used, but other categories of collateral relatives can also be used for that purpose. Among the other categories, full-sib families and clones are the most likely to be used, since they are often available and have high coefficients of relationship which are conducive to reasonably precise estimates. The choice of what category to use will depend on the availability of material, the likely precision of estimates obtainable from any category, and whether concerns exist over non-additive gene effects being possibly involved in materially different genetic correlations from the additive effects. Choice of category, however, is secondary to the issue that we will address.

Estimation of genetic correlations between a pair of traits from groups of collateral relatives is usually done in one of two types of situation (BURDON, 1977): 'Type A', where both traits are measured on the same sample of individuals from the genetic groups involved, which is the classical application, or 'Type B', where the two traits are measured on independent samples from within groups.

These cases are actually the two extremes of a continuum. Intermediate cases can include: where one trait is evaluated only on a subset of the individuals that are evaluated for the other trait, or where the individuals that are evaluated for both traits are only a partially overlapping subset of the total of those that are measured for respective traits. This note generalises the estimation procedure to all such cases, subject to the restrictions that the numbers of individuals represented for each trait and jointly for both traits are constant among genetic groups. This solution is 'ANOVA-based', i.e. based on classical analysis of variance using sums of squares and corresponding sums of cross-products. As such, it is a very straightforward solution which will be satisfactory if the sample is adequate and the imbalance in the classification meets the stated restrictions.

A still more general, but much more technical demanding solution is addressed in a separate paper (APIOLAZA *et al.*, in prep.) based on Restricted Maximum Likelihood (REML) methods.

The Theory

A generalised ANOVA-based procedure for estimating genetic correlations requires a general formulation of the expected covariance of group means between two traits, x and y. Assume that x and y are evaluated on random samples of n and m individuals per group respectively, and both traits are evaluated on a subset of q individuals, with the conditions $m \leq n$, $q \leq m$, with the individuals sampled per group totalling $m + n - q$ (= N). For the classical Type A correlation estimates $m = n = q = N$, and for the Type B case $q = 0$.

There can be practical reasons for intermediate sampling schemes, provided they still afford satisfactory genetic correlation estimates. If a high-heritability trait is very costly to measure, or must be measured destructively, it is intuitively attractive to measure it on a smaller sample ($m < n$). If one can evaluate for either trait non-destructively, but not for both, then a small or zero value of q may be indicated. A likely situation is where one trait is measured just on a subset of the individuals that are assessed for the other trait ($q = m < n$), but partially overlapping subsets ($q < m$) might still be used.

Taking traits individually, and assuming a balanced classification and complete individual randomisation of individuals with k random groups (e.g. half-sib families, full-sib families or clones), we have for trait x the expectations for mean squares (*Table 1*).

The expected variance of group means ($\sigma^2_{\bar{g}}$) for trait x is given by:

$$\sigma^2_{\bar{g}} = \sigma^2_g + \sigma^2_w / n \qquad (1)$$

Considering trait y, m is substituted for n.

Assuming an additivity plus dominance genetic model, with no maternal effects in the case of seedlings or 'c-effects' in the case of clones, the more detailed expectations for $\sigma^2_{\bar{g}}$ for trait x are given by:

$$\sigma^2_{\bar{g}} = \tfrac{1}{4}\sigma^2_A + (\tfrac{3}{4}\sigma^2_A + \sigma^2_D + \sigma^2_e) / n, \text{ for half-sibs} \qquad (2)$$

$$\sigma^2_{\bar{g}} = (\tfrac{1}{2}\sigma^2_A + \tfrac{1}{4}\sigma^2_D) + (\tfrac{1}{2}\sigma^2_A + \tfrac{3}{4}\sigma^2_D + \sigma^2_e) / n, \text{ for full-sibs} \qquad (3)$$

$$\sigma^2_{\bar{g}} = (\sigma^2_A + \sigma^2_D) + \sigma^2_e / n, \text{ for clones} \qquad (4)$$

where σ^2_A and σ^2_D are additive and dominance genetic variances respectively, and σ^2_e is environmental variance. Again, m is substituted for n when considering trait y.

[1]) New Zealand Forest Research Institute, Private Bag 3020, Rotorua, New Zealand

[2]) Department of Animal Science, Massey University, Palmerston North, New Zealand

Table 1. – Form of analysis of variance, for trait x.

Item	df	Expected mean square
Among groups	k - 1	$\sigma_w^2 + n\sigma_g^2$
Among individuals within groups	k(n - 1)	σ_w^2

σ_g^2 and σ_w^2 are among-group and within-group variances respectively.

For the limiting case of m or n = 1, we just have phenotypic variance (σ_P^2) among the individuals measured, of the expectation

$$\sigma_P^2 = \sigma_A^2 + \sigma_D^2 + \sigma_e^2 \qquad (5)$$

in which the three variance components are fully confounded. If the heritability (h^2) approaches unity ($\sigma_A^2 \rightarrow \sigma_P^2$), which may occur with certain expensive-to-determine traits, this confounding could be unimportant.

Considering a pair of traits, in a Type A situation (m = n = q = N), the mean cross-products have expected compositions analogous to those in *table 1*, with an expected covariance of group means ($cov_{\bar{g}}$) between the traits (cf Equation 1) given by

$$cov_{\bar{g}} = cov_g + cov_w / n \qquad (6)$$

where cov_g and cov_w are respectively the among- and within-group covariances between the traits. The more detailed expectations, in terms of genotypic and environmental covariance components, are analogous to the expected composition of group-mean variances (Equations 2, 3 and 4).

However, the generalised covariance of family means, which is of prime interest, is between the means that are based on the n individuals for trait x and the m individuals for trait y ($cov_{\bar{g}mn}$). It can be shown that

$$cov_{\bar{g}mn} = cov_g + \frac{q}{nm} cov_w \qquad (7)$$

For Type A situations (q = m = n = N) this duly simplifies to equation 6, and for Type B situations (q = 0) cov_w duly disappears from the expectation. If k (q – 1) is large enough to provide a good estimate of cov_w then it should be possible to obtain a good estimate of cov_g, which may differ appreciably from $cov_{\bar{g}}$ if nm is not large and a strong non-genetic covariance exists.

The estimated between-trait genetic correlation (r_g) is of course given by

$$r_g = \hat{cov}_g / (\hat{\sigma}_{gx} . \hat{\sigma}_{gy}) \qquad (8)$$

where σ_{gx}^2 and σ_{gy}^2 are the between-group variances for x and y respectively.

Estimation of σ_g^2 is straightforward (*Table 1*), although m should be large enough to give good estimates. Estimation of cov_g can be done from equation 7, using observed $cov_{\bar{g}mn}$, $\hat{cov}_w$, and q, m and n.

Discussion

Two issues arising are: (1) how to cope with additional types of imbalance in the classification, viz where one or more of n, m, q and N vary among groups, and (2) characterising the sampling distribution of r_g.

An ANOVA-based solution for more general imbalance has not been developed, and if such imbalance must be addressed a REML-based solution seems to be indicated. In practice, however, such imbalance may often not be an issue. It would be a greater problem if m varies among groups than if n varies, but if cost rather than available individuals dominates choice of m, one should standardise m for all groups.

Characterising estimation errors, and thence confidence limits, is a besetting problem with genetic correlations (LIU *et al.*, 1997). The distributions are often highly non-normal with finite sample populations. Hence the formulae for standard errors can become very crude approximations, especially when point estimates of the correlations and even the heritabilities approach their theoretical bounds. Bootstrap estimates of error distributions therefore have attractions even for Type A and Type B cases, and should be very attractive for the intermediate cases.

While bootstrap estimates of precision would seem appropriate for actual data there is also the question of expected efficiency in terms of precision, for choosing sampling schemes. Monte Carlo simulation would seem appropriate (cf BROWN, 1969; LIU *et al.*, 1997; APIOLAZA *et al.*, in prep.). Parameters to be varied would include: heritabilities of the respective traits, genetic correlation between traits, sample size (m, n) and overlap configuration (q) for the traits concerned, and costs of evaluation for the respective traits to help predict cost-efficiency.

As it is, we have extended a rigorous and very straightforward method of estimating genetic correlations to cover some situations that readily arise in practice (e.g. BURDON and LOW, 1992; DVORAK and WRIGHT, 1994). In those situations, there is a potential for large gains in cost-efficiency.

References

APIOLAZA, L. A., BURDON, R. D. and GARRICK, D. J.: Effect of univariate subsampling on the efficiency of bivariate parameter estimations and selection from half-sib progeny tests. (in preparation). — BROWN, G. H.: An empirical study of the distribution of the genetic correlation coefficient. Biometrics **25**: 63–72 (1969). — BURDON, R. D.: Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. Silvae Genet. **26**: 168–175 (1977). — BURDON, R. D. and LOW, C. B.: Genetic survey of *Pinus radiata*. 6: Wood properties: variation, heritabilities and interrelationships with other traits. NZ J. For. Sci. **22**: 228–245 (1992). — DVORAK, W. and WRIGHT, J. A.: Offspring-parent correlations for wood density of *Pinus tecunumannii* between nature and exotic environments. Can. J. For. Res. **24**: 1593–1596 (1994). — LIU, B.-H., KNAPP, S. J. and BIRKES, D.: Sampling distributions, biases, variances and confidence intervals for genetic correlations. Theor. Appl. Genet. **94**: 8–19 (1997).

Buchbesprechungen

Teil I: Nachhaltige Entwicklung von Kiefernwäldern. Von J. H. KUPER. **Teil II: Wald und Forstwirtschaft Niedersachsens im Kohlenstoffhaushalt.** Von K. BÖSWALD und R. WIERLING. Schriftenreihe „Aus dem Walde", Mitteilungen aus der Niedersächsischen Landesforstverwaltung, Heft 50. Herausgegeben vom Niedersächsischen Ministerium für Ernährung, Landwirtschaft und Forsten. 1997. Bezug: Niedersächsisches Forstplanungsamt, Forstweg 1A, D-38302 Wolfenbüttel. 333 Seiten. DM 20,–.

In dem vorliegenden Heft der Schriftenreihe „Aus dem Walde" wird im ersten Teil (233 Seiten), die aus dem Englischen übersetzte Promotionsschrift des Leiters der Königlich-Niederländischen Forsten Het Loo, Dr. JAAP KUPER, ungekürzt wiedergegeben. In ihr werden Ansätze einer naturgemäßen Waldwirtschaft mit bewußt niedrig gehaltenem Aufwand auf ärmsten Sandstandorten beschrieben, die einen Vergleich mit Nordwestdeutschland reizvoll machen. Phänomene, wie Eichensaaten durch Eichelhäher, Sukzession von Birke, Kiefer und anderen Baumarten sowie deren waldbauliche Behandlung, werden in dieser Schrift systematisch ausgewertet. Ungewöhnliche Versuche zur Anreicherung von nicht ausreichenden spontanen Verjüngungen wurden angestellt, z. B. Ergänzungspflanzungen mit Kiefern unter Kiefernaltholzschirm. Für deutsche Forstleute ungewohnt ist der betriebswirtschaftliche Teil der Arbeit, dem Ansätze der angelsächsischen forstökonomischen Theorie zugrunde liegen.

Im zweiten Teil des Heftes (86 Seiten) ist eine Arbeit abgedruckt, die der Frage nachgeht, welche Leistungen Wald, Forst- und Holzwirtschaft in Niedersachsen für den Kohlenstoffhaushalt gegenwärtig erbringen und mit welcher zukünftigen Entwicklung zu rechnen ist. Die mit Sorgfalt und Vorsicht angestellten Berechnungen weisen am Beispiel des verhältnismäßig kleinen Landes Niedersachsen nach, welch großen Stellenwert der Wald und die Forstwirtschaft zur Abwendung drohender Umweltrisiken besitzt; ein Sachverhalt, dem mehr politische Aufmerksamkeit gewidmet werden sollte.

Dem Heft bleibt zu wünschen, daß es über Niedersachsen hinaus Beachtung findet.

M. LIESEBACH (Grosshansdorf)

Plant Molecular Biology. A Laboratory Manual. Edited by M. S. CLARK. 1997. Springer Verlag, Berlin. 529 pages with 37 figures. Price DM 120,–.

This manual consists of three parts, which deal with basic molecular techniques, characterization of plant DNA, and genetic engineering methodology and analysis. Part I is further divided into four chapters: 1) Genomic DNA isolation, Southern blotting and hybridization; 2) Cloning from genomic DNA and production of libraries; 3) Extraction of RNA, cloning and subtractive hybridization; and 4) Characterization of clones. Each chapter in this Part I has further been divided into sub-chapters. Sub-chapters in chapter 1 include: 1.1) isolation of total genomic DNA (WILKE); 1.2) Southern blotting (CLARK); 1.3) hybridization with radioactive probes (CLARK); 1.4) non-radioactive methods of detection on Southern blots (LEROY et al.). There are 5 sub-chapters in Chapter 2: 2.1) use of PCR in plant molecular biology (GUEVARA-GARCIA et al.); 2.1) plasmid libraries (CLARK); 2.3) Lambda genomic cloning (ELGAR); 2.4) Cosmic libraries (LIU); and 2.5) yeast artificial chromosome (YAC) libraries (MATALLANA et al.). Chapter 3 consists of the following sub-chapters: 3.1) isolation and analysis of messenger RNA from plant cells: cloning of cDNAs (LESSARD et al.); and 3.2) subtractive hybridization of different mRNA populations (FOOTE et al.). Chapter 4 of Part I consists of two sub-chapters: 4.1) DNA sequencing (CLARK); and 4.2) expression of cloned genes (PEHU). Part II consists of 3 chapters: 5) organelle DNA isolation (LANDGREN and GLIMELIUS); 6) RAPD analysis: use of genomic characterization, tagging and mapping (WAUGH); and 7) RFLP mapping of plant nuclear genomes: planning of experiments, linkage map construction, and QTL mapping (VAN DEN BERG et al.). Part III consists of 5 chapters: 8) plant gene transfer (POTTER and JONES); 9) molecular characterization of transformed plants (TOPPING and LINDSEY); 10) molecular characterization of somatic hybrids (XU and PEHU); 11) Cytological characterization of transformed plants: mapping of low-copy and repetitive DNA sequences by fluorescent in situ hybridization (FISH) (LEITCH et al.); and 12) cytological characterization of somatic hybrids: detection of genome origin by genomic in situ hybridization (GISH) (KENTON et al.). In addition to these 12 broad chapters, there is an addendum to chapters 11 and 12: A1) in situ hybridization in plant species with small chromosomes (LAPITAN). A one-page appendix on nuclear genome sizes of important plant species (C.-N. LIU) completes the manual.

This laboratory manual covers a broad range of latest techniques in the area of plant molecular biology. Each chapters contains basic information for the molecular technique, which is followed by the standard requirements for the protocol, that is, reagents, equipment, procedures, trouble-shooting notes, and references. The 12 chapters (and their sub-chapters wherever applicable; an addendum and appendix are contributed by 42 contributors. Although basic techniques are part of all the chapters, the arbitrary classification into Part I, II and III and sub-chapters in only part I is somewhat puzzling. In my opinion, the chapters could have been numbered from one to 21 or 22. But that is not a weak point of the book. On the whole, the manual is a good compilation of the latest techniques in molecular biology of plants that would be useful not only for the students but also researchers interested in this area of rapid advancements.

M. R. AHUJA (Grosshansdorf)

Herausgeberin: Bundesforschungsanstalt für Forst- und Holzwirtschaft: Schriftleitung: Institut für Forstgenetik und Forstpflanzenzüchtung, Siekerlandstrasse 2, D-22927 Grosshansdorf — Verlag: J. D. Sauerländer's Verlag, Finkenhofstrasse 21, D-60322 Frankfurt a. M. — Anzeigenverwaltung: J. D. Sauerländer's Verlag, Frankfurt am Main. — Satz und Druck: Graphische Kunstanstalt Wilhelm Herr, D-35390 Giessen
Printed in Germany.

ISSN 0037-5349