Proc. Assoc. Advmt. Breed. Genet. 21: 437-440

MISSING HERITABILITY OF ADAPTATION PHENOTYPES IN TROPICAL CATTLE

Laercio R. Porto-Neto, Nicholas J. Hudson, William Barendse, Sigrid A. Lehnert and Antonio Reverter

CSIRO Agriculture Flagship, 306 Carmody Road, St Lucia, QLD 4067, Australia

SUMMARY

The 'missing heritability problem' is the inability to tag all the genetic variance of a trait using genome wide single nucleotide polymorphism. Here, we compute missing heritability for two populations of cattle phenotyped for ten tropical adaptation traits that exhibit variable genetic architectures. We derived genomic relationship matrices (GRM) using both low and high density SNP panels, and computed the missing heritability through comparison to pedigree (NRM). Overall, the low density indicine panel performs very well in characterising the Brahman population. We found that estimation of missing heritability was broadly similar for both panels across the ten phenotypes. This implies similar amounts of genetic variation relevant to those phenotypes have been captured. The phenotypes with the lowest missing heritability (coat type and sheath score in Tropical Composites) possess an architecture that can be characterised simply. That is, they are dominated by genes of large effect.

INTRODUCTION

The total genetic variance of a trait is usually estimated using pedigree information. Then the total variance (phenotypic variance) is partitioned into genetic and environmental variance. The heritability or ratio of genetic variance to total phenotypic variance can be calculated. In genome-wide association studies (GWAS), the shrunk variance associated with each significant single nucleotide polymorphism (SNP) can be estimated. If all SNP are considered simultaneously, a large proportion of the variance that would have been missed due to small individual effects can be captured (Yang *et al*. 2010). However, it has been shown that in all studies to date this sum of those variances is usually far less than the total genetic variance. Since the SNP in the GWAS cover 90% or more of the whole genome, the inability to account for the total genetic variance is called the missing heritability problem.

The degree of missing heritability varies for different traits in the same dataset. Here we estimate the different degrees of missing heritability for 10 traits of tropical cattle for two breeds to determine whether there are large differences either between traits or between breeds in the amount of the genetic variance that is not tagged by SNP.

To estimate the missing heritability we use the approach of simultaneously analysing the genetic variance using both pedigree and SNP data, in which the genetic variance not accounted for by the SNPs will be assigned to the pedigree component. This allows us to maximize the amount of genetic variance assigned to SNP, avoids the issues of significance thresholds, and biased estimation of the SNP effects due to small sample sizes (in the thousands instead of in the hundreds of thousands).

MATERIAL AND METHODS

Animals, genotypes and phenotypes. We used the genetic and phenotypic resources outlined in Porto-Neto *et al*. 2014. In brief, 2112 Brahman and 2550 Tropical Composite animals were genotyped with either the Bovine HD BeadChip (Illumina Inc., San Diego, CA) that includes more than 770,962 SNP or the GGP Indicine chip that includes 71,726 SNP.

Calculation of genetic relationship matrix (GRM) and numerator relationship matrix (NRM). The GRM was computed based on the methodology developed by Van Raden (2008).

$$GRM = \frac{\mathbf{ZZ}^T}{2\sum p_i(1-p_i)}$$

where **Z** is a matrix that relates SNP alleles to individuals and p_i is the frequency of the second allele for the *i*-th SNP. $\mathbf{ZZ}^T$ represents the number of shared SNP alleles among two individuals and the division of $\mathbf{ZZ}^T$ by $2\sum p_i(1-p_i)$ aims at scaling the GRM to make it analogous to the NRM obtained based on the pedigree information.

Variance components. A single model with two random components was run in Qxpak (Perez-Enciso and Misztal 2011). We estimated missing heritability using the approach of Roman-Ponce *et al*. 2014.

$$C_{miss} = 1 - \frac{\sigma_u^2}{\sigma_a^2 + \sigma_u^2}$$

where σ_u^2 is the variance due to the genotype data (ie GRM) and σ_a^2 is the additive genetic variance due to the pedigree (ie. the NRM in our context).

RESULTS AND DISCUSSION

To provide an estimate of how well the two SNP panels characterised the genetics of the two populations, we plotted the population level frequency of the reference allele of each loci on a genome-wide basis (Figure 1).

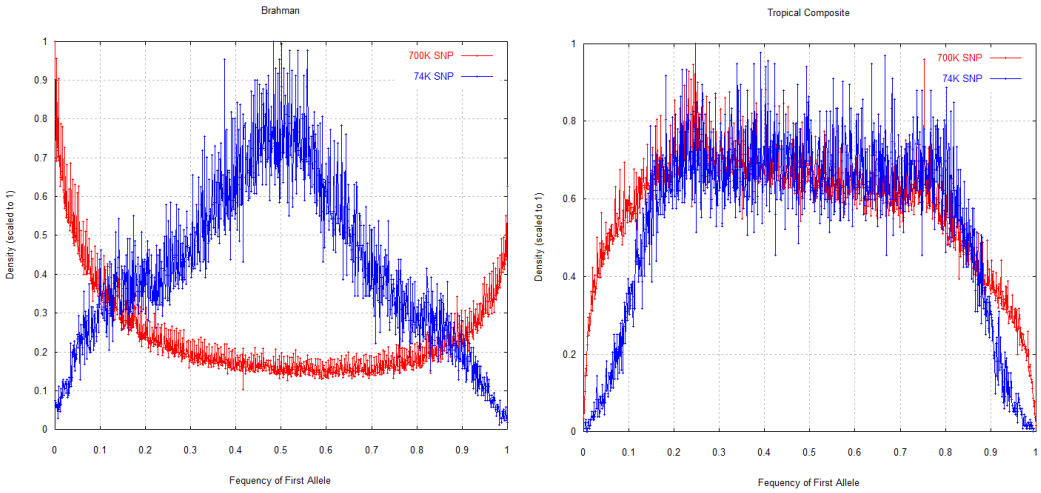

Figure 1. Population level allelic frequency for each SNP panel, 74K (blue lines) and 700K (red lines) on the Brahman (left hand plot) and Tropical Composite (right hand plot) populations.

The left hand plot shows that the Brahman sample has a relative deficiency of low frequency alleles in the low density Indicine SNP panel (74K), compared to the HD SNP (700K) panel, and the frequency of the reference allele is biased to p = 0.5. On the other hand, there is a bias on the HD SNP panel towards low frequent alleles. The right hand plot shows that the Tropical Composite animals have an allele frequency distribution that is similar for both the low density

Indicine SNP panel as well as the HD SNP panel, although there does seem to be a relative lack of very low frequency SNPs in the low density panel for this sample. It is worth mentioning that the 74K selection is part of the full HD panel, so the impact of two independent SNP selections was not tested here. We could speculate that a set of SNP that poorly characterise the structure of a population would also poorly perform in capturing the traits' variance. Put another way, a random selection of 74K SNP would perform worst than a set of SNP with on average high minor allelic frequencies.

We next computed missing heritability scores for the Brahman (Table 1) and Tropical Composite (Table 2) using both SNP panels.

Table 1. Genetic parameters and missing heritability for Brahman cattle based on low and high density SNP panels.

TRAIT	74K			700K		
	NRM	GRM	Missing	NRM	GRM	Missing
Coat type	0.232	0.264	0.468	0.228	0.269	0.460
Coat colour	0.269	0.324	0.453	0.186	0.362	0.339
Condition score	0.223	0.375	0.372	0.139	0.409	0.253
Worm eggs (n/gr)	0.351	0.363	0.491	0.351	0.365	0.491
Fly lesions	0.231	0.282	0.450	0.261	0.287	0.476
Flight time	0.286	0.233	0.552	0.287	0.233	0.552
Sheath score	0.284	0.335	0.459	0.281	0.360	0.438
Temperature	0.230	0.195	0.541	0.233	0.187	0.555
Tick score	0.413	0.384	0.518	0.413	0.383	0.519
Yearling Weight	0.210	0.316	0.399	0.208	0.306	0.405

Table 2. Genetic parameters and missing heritability for Tropical Composite cattle based on low and high density SNP panels.

TRAIT	74K			700K		
	NRM	GRM	Missing	NRM	GRM	Missing
Coat type	0.169	0.502	0.252	0.267	0.430	0.383
Coat Colour	0.284	0.389	0.422	0.279	0.400	0.411
Condition score	0.175	0.321	0.354	0.175	0.306	0.365
Worm eggs (n/gr)	0.369	0.359	0.507	0.365	0.365	0.500
Fly lesions	0.359	0.364	0.497	0.358	0.365	0.495
Flight time	0.251	0.311	0.447	0.241	0.327	0.424
Sheath score	0.235	0.529	0.308	0.235	0.531	0.307
Temperature	0.218	0.157	0.581	0.215	0.156	0.581
Tick score	0.382	0.370	0.508	0.387	0.375	0.507
Yearling Weight	0.293	0.363	0.446	0.285	0.385	0.425

These missing heritability estimates are somewhat higher than the 36.6% reported by Roman-Ponce *et al* (2014) across 11 cattle traits and the 38.5% reported by Haile-Mariam *et al* (2013) across 29 cattle traits. In the Brahman sample, the extra genetic information provided by the high density SNP panel only led to a systematic reduction in missing heritabilities in coat colour and condition score. In the Tropical Composite sample, the larger SNP panel did not lead to a clear reduction in missing heritability, and interestingly, in the trait coat type the missing heritability was lower when the lower density panel was used. This particular phenotype is controlled by a small number of genes and so the addition of non-informative SNP from across the genome may have reduced the accuracy of prediction.

We noted that the two phenotypes with the lowest missing heritability in the Tropical Composite populations (Coat type and Sheath score) are phenotypes with a relatively simple genetic architecture characterised by relatively few gene regions of large effect (Porto Neto *et al.* 2014). It is not surprising that this kind of simple genetic architecture would be more amenable to accurate modelling of genotype to phenotype relationships.

REFERENCES

Haile-Mariam M., Nieuwhof G.J., Beard K.T., Konstatinov K.V., Hayes B.J. (2013) *J. Anim. Breed. Genet.* **130**: 20.

Perez-Enciso M. and Misztal I. (2011) *BMC Bioinf.* **12**: 202.

Porto-Neto L.R., Reverter A., Prayaga K.C., Chan E.K.F., Johnston D.J., Hawken R.J., Fordyce G., Garcia J.F., Sonstegard T.S., Bolormaa S., Goddard M.E., Burrow H.M., Henshall J.M., Lehnert S.A. and Barendse W. (2014) *PLoS ONE* **9**: e113284.

Roman-Ponce S.I., Samore A.B., Dolezal M.A., Bagnato A., Meuwissen T.H.E. (2014) *Genet. Selec. Evol.* **46**: 36.

Van Raden P.M. (2008) *J. Dairy Sci* **91**: 4414.

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E. & Visscher P.M. (2010) *Nature Genetics* **42**: 565.