



Using genetic markers in unpedigreed populations to detect a heritable trait

DODDS Ken G.^{†1}, AMER Peter R.², AUVRAY Benoît¹

(¹Applied Biotechnologies Group, AgResearch Limited, Invermay Research Centre, Private Bag 50034, Mosgiel 9053, New Zealand)

(²AbacusBio Limited, PO Box 5585, Dunedin 9016, New Zealand)

[†]E-mail: ken.dodds@agresearch.co.nz

Received Sept. 20, 2007; revision accepted Sept. 28, 2007

Abstract: Before a breeder invests selection pressure on a trait of interest, it needs to be established whether that trait is actually heritable. Some traits may not have been measured widely in pedigreed populations, for example, a disease or deformity may become more prevalent than previously, but is still relatively rare. One approach to detect inheritance would be to screen a commercial population to obtain a sample of “affecteds” (the test group) and to also obtain a random control group. These individuals are then genotyped with a set of genetic markers and the relationships between individuals within each group estimated. If the relatedness is higher in the test group than in the control group, this provides initial evidence for the trait being heritable. A power simulation shows that this approach is feasible with moderate resources.

Key words: Relatedness, Genetic markers, Heritable trait

doi:10.1631/jzus.2007.B0782

Document code: A

CLC number: Q78; TP31

INTRODUCTION

When breeding and farming animals, there sometimes appears a group of individuals with an unusual phenotype. This phenotype may be the result of the environment (for example unusual weather conditions or presence of disease-causing agents) or it may be due to a recent genetic change in the population (for example, a recent increase in frequency of a genetic defect). Often the phenotype is seen only in commercial, unpedigreed populations. For example, if the phenotype is sufficiently rare, it may be only the large commercial populations where it is noticed in sufficient numbers to be reported. In addition, breeders may not wish to disclose the presence of an undesirable trait, and would cull any animal showing the trait.

Before undertaking a large-scale experiment to investigate the possibility of genetic inheritance, it would be useful to have some indication of whether the trait is inherited. One possibility is to use genetic markers to elucidate the relationships between indi-

viduals showing the phenotype and to compare these to the relationships between individuals not showing the phenotype. If members of the former group are more closely related than those of the latter group, this is evidence for the trait having a heritable component. A number of studies have developed methods for estimating relatedness using genetic markers and many of these are reviewed and compared by Blouin (2003), Milligan (2003), Oliehoek *et al.* (2006) and Weir *et al.* (2006). Relatedness (r) is defined as twice the probability that random alleles from each of the individuals are identical by descent (IBD). A more precise description can be made by considering the probability of sharing 0, 1 or 2 alleles that are IBD (kinship coefficients k_0 , k_1 and k_2), in which case $r=k_1/2+k_2$. Table 1 shows these values for some common relationships.

Relatedness estimation methods have a high sampling variance when trying to estimate the relationship between a pair of individuals (van de Castele *et al.*, 2001). It has been suggested that a reasonable estimation of relatedness (i.e. standard

Table 1 Values of relatedness (r) and kinship coefficients (k_0 , k_1 and k_2) for some common relationships, assuming non-inbred individuals

Relationship	r	k_0	k_1	k_2
Self	1.0000	0	0	1.00
Parent-offspring	0.5000	0	1.000	0
Full-sibs	0.5000	0.250	0.500	0.25
Half-sibs	0.2500	0.500	0.500	0
Cousin	0.1250	0.750	0.250	0
Half-cousin	0.0625	0.875	0.125	0

deviation less than 0.1) requires 30~40 microsatellite markers (Blouin, 2003) or 100~200 single nucleotide polymorphism (SNP) markers (Weir *et al.*, 2006). Using this number of markers may be beyond the budget of the type of study being investigated here, particularly if marker development is required. However the prospect for estimating the average relatedness within a group using a smaller number of markers is better (Queller and Goodnight, 1989).

We investigate the utility of using genetic markers to detect a difference in within-group relatedness as a method for inferring that a trait is heritable. This is done using simulation methods to estimate the power for population structures and marker sets that reflect practical situations. An aquaculture and a livestock situation are modelled as examples of possible population structures.

METHODS

Simulated populations

Many aquaculture species have the ability to produce cohorts of extremely large (in the thousands) full-sib families. For the aquaculture simulation, full-sib families were created by one to one mating (i.e., each parent has only one mate), but these families may be related to each other. The simulated population consisted of three generations, with five unrelated individuals of each of the four grandparental types. From their simulated progeny, 20 male and 20 female individuals were randomly chosen as the parents of the final (progeny) generation. In each generation there were either 200 or 400 progeny per family (to see if there were important effects due to the choice of this parameter). The phenotype of in-

terest (test phenotype) was modelled as a dominant phenotype, and one of the paternal grandsires was chosen to be heterozygous for the genotype while all other grandparents were non-carriers. If a simulation replicate resulted in no test phenotypes, that replicate was discarded and not counted. With this population structure, pairs of test phenotype individuals are either cousins or full-sibs. A set of either 20 or 40 test phenotype individuals and a set of the same number of normal (control) individuals were randomly sampled from the population for 'genotyping'.

Many livestock populations contain large (maybe 100 or more) half-sib families and there may be small full-sib families within these (depending on the species). For the livestock simulation, each female was mated to a male that was randomly chosen from those available. The simulated population contained 5 paternal grandsires, 50 paternal granddams, 20 maternal grandsires and 2000 maternal granddams. From their progeny, 20 male and 2000 female individuals were randomly chosen as the parents of the progeny generation. In each generation each female had two progeny. The test phenotype was modelled in the same way as for the aquaculture simulation. With this population structure, most pairs of test phenotype individuals are either half-cousins or half-sibs.

Marker sets

We used a 9-marker set for the aquaculture simulation, and a 6-marker set and a 12-marker set for the livestock simulation. Allele frequencies are shown in Table 2. The (9-marker) aquaculture set had a probability of identity (Taberlet and Luikart, 1999; Ayres and Overall, 2004) (to an unrelated individual) of 3×10^{-14} , so each marker was approximately equivalent to a marker with seven equally frequent alleles. The 6- and 12-marker livestock sets had probabilities of identity of 2×10^{-7} and 4×10^{-13} , respectively, approximately equivalent to the same number of markers with four to five equally frequent alleles. In the simulation genotypes were determined without error and without missing values.

Test of relatedness difference

For all pairs of test phenotype individuals (T-T) and for all pairs of control individuals (C-C) the relatedness was estimated using the program MER

Table 2 Allele frequencies for the markers used in the simulations

Marker set	Marker	Allele frequencies
Aquaculture	A1	0.668, 0.077, 0.067, 0.054, 0.042, 0.040, 0.035, 0.012, 0.002, 0.002
	A2	0.116, 0.109, 0.094, 0.082, 0.072, 0.057, 0.050, 0.050, 0.047, 0.042, 0.037, 0.035, 0.035, 0.025, 0.025, 0.020, 0.020, 0.020, 0.015, 0.010, 0.010, 0.007, 0.007, 0.005, 0.005, 0.005, 0.002
	A3	0.161, 0.136, 0.116, 0.109, 0.067, 0.050, 0.040, 0.040, 0.035, 0.030, 0.027, 0.027, 0.025, 0.025, 0.022, 0.020, 0.017, 0.015, 0.012, 0.010, 0.010, 0.005, 0.002
	A4	0.183, 0.171, 0.136, 0.124, 0.077, 0.072, 0.047, 0.045, 0.042, 0.032, 0.025, 0.022, 0.007, 0.007, 0.007, 0.002
	A5	0.149, 0.139, 0.131, 0.116, 0.101, 0.052, 0.047, 0.035, 0.030, 0.030, 0.027, 0.025, 0.025, 0.022, 0.020, 0.017, 0.015, 0.007, 0.005, 0.002, 0.002, 0.002
	A6	0.295, 0.282, 0.134, 0.084, 0.074, 0.037, 0.035, 0.025, 0.017, 0.010, 0.007
	A7	0.238, 0.213, 0.166, 0.092, 0.084, 0.079, 0.064, 0.057, 0.007
	A8	0.243, 0.149, 0.126, 0.089, 0.079, 0.059, 0.054, 0.052, 0.042, 0.025, 0.025, 0.015, 0.015, 0.012, 0.007, 0.005, 0.002
	A9	0.302, 0.183, 0.087, 0.084, 0.079, 0.077, 0.064, 0.047, 0.027, 0.025, 0.017, 0.007
Livestock A	A1	0.476, 0.171, 0.124, 0.091, 0.079, 0.017, 0.017, 0.012, 0.010, 0.005
	A2	0.361, 0.302, 0.149, 0.085, 0.054, 0.050
	A3	0.247, 0.190, 0.190, 0.143, 0.060, 0.049, 0.042, 0.034, 0.023, 0.020, 0.010, 0.005, 0.003, 0.003, 0.003
	A4	0.296, 0.154, 0.142, 0.121, 0.102, 0.080, 0.040, 0.019, 0.017, 0.014, 0.007, 0.007, 0.002
	A5	0.428, 0.323, 0.079, 0.055, 0.036, 0.029, 0.022, 0.017, 0.007, 0.005
	A6	0.303, 0.265, 0.107, 0.092, 0.088, 0.062, 0.043, 0.031, 0.005, 0.002, 0.002
Livestock B	B1	0.321, 0.278, 0.135, 0.068, 0.065, 0.056, 0.038, 0.018, 0.012, 0.004, 0.004, 0.003, <0.001
	B2	0.204, 0.202, 0.140, 0.122, 0.098, 0.063, 0.050, 0.047, 0.037, 0.018, 0.016, 0.001, 0.001, <0.001, <0.001
	B3	0.554, 0.193, 0.095, 0.073, 0.020, 0.019, 0.018, 0.015, 0.010, 0.004
	B4	0.325, 0.277, 0.170, 0.081, 0.064, 0.033, 0.022, 0.020, 0.004, 0.004, <0.001
	B5	0.277, 0.227, 0.206, 0.092, 0.056, 0.043, 0.027, 0.027, 0.027, 0.017, 0.002, <0.001, <0.001
	B6	0.489, 0.194, 0.150, 0.096, 0.038, 0.028, 0.005, 0.001, <0.001
	B7	0.415, 0.399, 0.122, 0.029, 0.028, 0.006, 0.001, 0.001
	B8	0.300, 0.283, 0.154, 0.093, 0.045, 0.040, 0.032, 0.018, 0.013, 0.011, 0.008, 0.002, 0.001, 0.001, <0.001
	B9	0.444, 0.360, 0.101, 0.096
	B10	0.197, 0.160, 0.140, 0.133, 0.099, 0.080, 0.074, 0.072, 0.011, 0.011, 0.009, 0.008, 0.006, <0.001
	B11	0.290, 0.218, 0.106, 0.079, 0.077, 0.066, 0.055, 0.035, 0.031, 0.026, 0.009, 0.008, <0.001
	B12	0.340, 0.238, 0.132, 0.123, 0.100, 0.039, 0.016, 0.004, 0.003, 0.003, 0.001, 0.001, <0.001, <0.001, <0.001, <0.001

(Wang, 2002). The medians of the estimated relatedness values (r_{TT} and r_{CC} for test and control pairs respectively) were calculated along with their difference ($r_{TT}-r_{CC}$). Large values of this difference were taken as significant evidence of the trait being heritable. Medians, rather than means, were used to guard against possible skewness in the relatedness estimates. The significance of the difference was found using a randomization test (Manly, 1997) to accommodate non-independence amongst the pairs in each group. Two hundred replicates of each situation were simulated, and the power of the test was estimated as the proportion of significant results at the 5% level.

RESULTS

The calculated powers are shown in Table 3 for the aquaculture scenario and Table 4 for the livestock scenario. Also shown are the average values (over the simulations) for the median relatedness in the test groups. Powers were good (i.e., mainly above 0.8) except for the situation with the least powerful marker set (livestock with six markers) and smaller sample size. For the aquaculture scenario there was little difference in powers between the two family sizes, indicating that power should still be good with even larger families. Power was generally higher for the larger sample size and for more powerful marker sets.

Table 3 Power of the proposed method to detect a heritable trait for the aquaculture scenario

Sample size	Family size	Power	Average r_{TT}
20	200	0.87	0.22
40	200	0.92	0.22
20	400	0.83	0.18
40	400	0.93	0.19

Table 4 Power of the proposed method to detect a heritable trait for the livestock scenario

Sample size	Number of markers	Power	Average r_{TT}
20	6	0.55	0.14
40	6	0.74	0.14
20	12	0.78	0.13
40	12	0.90	0.14

DISCUSSION

Genetic markers offer the opportunity to elucidate the genetics of a phenotype of interest in unpedigreed populations. At one extreme, markers (e.g., tens of thousands of SNPs) can be used in genome-wide association studies to find genomic regions that influence the trait (Balding, 2006). Another approach is to use markers (e.g., hundreds of SNPs) to reconstruct a relationship matrix to use in the estimation of the trait heritability (Coltman, 2005; Thomas, 2005). Our approach is more akin to the latter than the former, as the markers are not used to specifically map a region influencing the trait. However, it differs in that, for the situations studied, it can indicate an inherited trait with fewer markers and fewer individuals than the relationship matrix approach, but does not give a precise estimate of heritability. Our method is likely to be used as a precursor to a more extensive study (either marker-based or using recorded pedigrees) to provide further information about the genetics of the trait.

This study has focused on animal breeding applications, but the method may work in other situations as well, such as plant breeding or in the study of natural populations. The resources required will depend on the family structures in such populations. Our method has not used supplementary information about the family structures, but this could aid the estimates of relatedness. For example, if it is known that there are only a few large full-sib families,

methods that reconstruct these groups are likely to give better relatedness estimates (Thomas and Hill, 2002).

In our simulations, the phenotype was defined as being due to a single dominant gene. The results would also pertain to other situations which would give rise to the same relationships within the test group (i.e., all individuals share a common grandsire). This would be the case for a recessive gene at a reasonably low frequency, possibly with incomplete penetrance. It might also be the case where specific combinations of gene variants are required for the phenotype to be expressed.

We have shown that it is possible to infer that a trait is heritable, by using markers to compare the relatedness within the test group against the relatedness within a control group. The power of the method depends on the family structures in the population, the sample size, the characteristics of the marker set and the actual genetic cause of the trait. The cases we have simulated, in which the test group shares a recent common ancestor, where a moderate number of individuals were sampled and where marker sets with 8~12 microsatellite markers were used, indicate good power for such an experiment.

References

- Ayres, K.L., Overall, A.D.J., 2004. Api-calc 1.0: a computer program for calculating the average probability of identity allowing for substructure, inbreeding and the presence of close relatives. *Molecular Ecology Notes*, **4**(2):315-318. [doi:10.1111/j.1471-8286.2004.00616.x]
- Balding, D.J., 2006. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, **7**(10):781-791. [doi:10.1038/nrg1916]
- Blouin, M.S., 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology and Evolution*, **18**(10):503-511. [doi:10.1016/S0169-5347(03)00225-8]
- Coltman, D.W., 2005. Testing marker-based estimates of heritability in the wild. *Molecular Ecology*, **14**(8): 2593-2599. [doi:10.1111/j.1365-294X.2005.02600.x]
- Manly, B.F.J., 1997. Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapman & Hall, London, p.1-23.
- Milligan, B.G., 2003. Maximum-likelihood estimation of relatedness. *Genetics*, **163**(3):1153-1167.
- Oliehoek, P.A., Windig, J.J., van Arendonk, J.A.M., Bijma, P., 2006. Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. *Genetics*, **173**(1):483-496. [doi:10.1534/genetics.105.049940]

- Queller, D.C., Goodnight, K.F., 1989. Estimating relatedness using genetic markers. *Evolution*, **43**(2):258-275. [doi:10.2307/2409206]
- Taberlet, P., Luikart, G., 1999. Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, **68**(1-2):41-55. [doi:10.1006/bjil.1999.0329]
- Thomas, S.C., 2005. The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1459):1457-1467. [doi:10.1098/rstb.2005.1675]
- Thomas, S.C., Hill, W.G., 2002. Sibship reconstruction in hierarchical population structures using Markov chain Monte Carlo techniques. *Genetical Research*, **79**(3): 227-234. [doi:10.1017/S0016672302005669]
- van de Casteele, T., Galbusera, P., Matthysen, E., 2001. A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology*, **10**(6):1539-1549. [doi:10.1046/j.1365-294X.2001.01288.x]
- Wang, J., 2002. An estimator for pairwise relatedness using molecular markers. *Genetics*, **160**(3):1203-1215.
- Weir, B.S., Anderson, A.D., Hepler, A.B., 2006. Genetic relatedness analysis: modern data and new challenges. *Nature Reviews Genetics*, **7**(10):771-780. [doi:10.1038/nrg1960]