

A GENOTYPING STRATEGY TO MINIMISE THE COST OF DNA TESTING.

B. P. Kinghorn

Beef CRC and Department of Animal Science, University of New England, Armidale, NSW 2351

SUMMARY

This paper describes a method for reducing the cost of DNA typing by processing or 'genotyping' only some of the individuals in the target population. The method is designed such that there is in fact good DNA information on the individuals that have not been genotyped. The method works in cycles, first genotyping the one individual that will contribute most information to the total population, through pedigree links. Following this, analysis is done to find which is the individual that can contribute most at this next stage, etc. After genotyping about 10 % of the population, about 50 % of the utility of genotyping all animals is reached, leading to potentially big savings in the cost of genotyping.

Keywords: DNA test, QTL, gene marker, cost reduction, segregation analysis

INTRODUCTION

DNA typing is becoming widely practiced, with applications using both genetic marker loci and known Quantitative Trait Loci (QTL). The cost of genotyping is generally high, such that inferring genotype from the known genotypes of relatives and/or linked loci has the potential to play a useful role in reducing costs.

This paper outlines a genotyping strategy that uses genotype probabilities from segregation analysis, numerator relationship, and potentially EBV's and other factors to help choose which individual(s) and loci to genotype in each cycle. A simple objective might be to maximise the ratio of utility of the resulting information across the whole population to the total cost of genotyping. A more comprehensive description of the method and its application is given in Kinghorn (1999), together with a more complete test of its performance.

MATERIALS AND METHODS

An outline of the genotyping strategy is shown in Table 1. This was tested by simulation, with the objective of gaining information about a single known major gene for the live individuals in a pedigreed data set. The data set consisted of 268 live individuals in a pedigree containing 1,260 individuals. A single biallelic gene was generated, with no effect on the traits in the selection criterion, and a base population allele frequency of 0.1. Genotype probabilities were calculated at each cycle using segregation analysis, as described by Kerr and Kinghorn (1996). For each individual in the population, this gives the probability of being each of the three genotypes, conditional on genotypes that are known in relatives.

Table 1. The genotyping strategy. See text for details

Step	Action
1	Choose the first individual* to genotype. In the example, this is the individual that is, on average, most related to live animals in the population.
2	Run a segregation analysis to get updated genotype probabilities for all individuals.
3	Calculate population mean genotype probability index as a descriptor of the utility of results to date. If utility and/or funds spent on genotyping are appropriate, then stop.
4	Calculate the ranking criterion for each individual ($100 \times \text{CON} - \text{GPI}$ in the example used), and then genotype the highest ranking individual*. Go to step 2.

* Groups of individuals can be tested in each cycle if this is logistically more appropriate, but at some cost in performance (Kinghorn, 1999).

The criterion used here for selecting individuals for genotyping at each cycle was $100 \times \text{CON} - \text{GPI}$. CON is the 'connection value' of the individual – the average numerator relationship to all live individuals in the population, ranging 0 to 1.0. For example, a sire that has already been widely used in the population would have a high CON value. GPI is the individual's genotype probability index (Kinghorn 1997), ranging 0 to 100 percent. As shown in Figure 1, this index has a value of zero for an individual with no direct information – such that it has genotype probabilities equal to Hardy-Weinberg frequencies. Individuals that have been genotyped with full confidence have GPI values of 100 percent. This is also true for individuals that are confidently genotyped by inference from relatives.

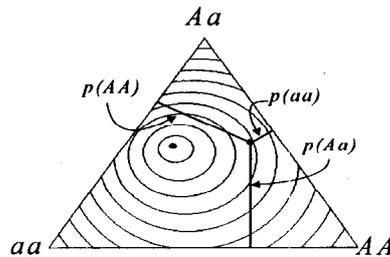


Figure 1. An illustration of the genotype probability index, GPI (Kinghorn 1997), for a biallelic locus (A, a). Probability $p(AA)$ of being genotype AA is the perpendicular distance from the point to the edge opposite the AA vertex. The dot to the left represents H-W genotype frequencies for $p(A) = 0.4$. Individuals at this location have a GPI value of zero. Individuals at the vertices have GPI values of 100 percent. Individuals on an edge of the triangle generally have positive GPI values, reflecting the value of being able to exclude the possibility of one genotype. The individual plotted has genotype probabilities $p(aa)=.1$, $p(Aa)=.5$, $p(AA)=.4$ and a GPI value of just under 40 percent, as can be seen from counting the contours.

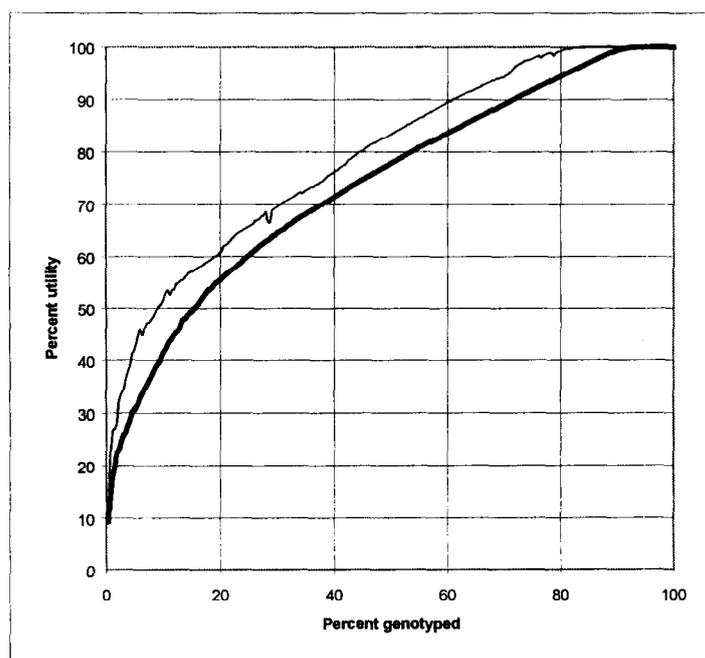


Figure 2. Plot of percent utility (population average GPI) versus percent of the population genotyped. The upper curve is for using $100 \cdot \text{CON} - \text{GPI}$ to rank and choose the individual to genotype at each cycle. The lower curve in bold is for random choice of individual to genotype at each cycle (average of 100 replicates, each with different random sampling).

The criterion $100 \cdot \text{CON} - \text{GPI}$ would favour individuals which currently have poor genotype information, and which are likely to provide more information to the rest of the population through genetic links. Other criteria incorporating positive emphasis on connection were tested, such as $(1 + \text{CON}) / (100 + \text{GPI})$, and similar results were found. Optimal weights in such criteria will depend on population structure.

A control strategy ('Random') was to select individuals to genotype at each cycle in a random manner. Results for this strategy are the average from 100 replicates, each with a different random sampling. Utility under each strategy was taken as the average GPI of live individuals after each genotyping.

RESULTS

Figure 2 shows that even the Random treatment gives a considerable increase in percent utility for the first genotypings made. This is expected, as with reasonably good pedigree information, a small number of randomly chosen individuals will be related to much of the population, giving genotype probabilities other than Hardy-Weinberg frequencies to much of the population. However, informed

choice of individual(s) to test at each cycle, using the criterion $100 \times \text{CON} - \text{GPI}$, can improve performance considerably (Figure 2).

DISCUSSION

This short paper has made a simple test of a cyclical approach to genotyping, under a single limited scenario. Kinghorn (1997) discusses extension of the genotype probability index to multiple alleles and multiple loci, such that inference about an individual's genotype at a given locus can be augmented by information about its known genotypes, and/or those of its relatives, at linked loci. The potential range of use of this approach is very wide, with many possible population structures and a range of applications, such that comprehensive testing to give broad conclusions is not possible in a short paper.

Utility of results has been taken here as average GPI across live individuals. This might be reasonable for some applications, for example those that use genotype probabilities as part of a selection index to exploit both a known QTL and polygenic breeding values for the same and/or other traits. However, in some other applications, utility should also consider dead individuals with appropriate phenotypic records, for example, where genotype probabilities are to be used to help estimate the effects of a major gene on phenotype (Kinghorn and Kerr 1995).

The ranking criterion for choosing individuals to genotype at each cycle should also be appropriate to the task. For example, with marker assisted selection there will be extra utility in gaining better information on individuals which are more likely to be selected – those with higher estimated breeding values – as genotyping them will contribute information to resulting descendants. In applications that aim to detect QTL, the index could usefully favour genotyping individuals of extreme phenotype, as this can give more detection power.

For many multi-locus applications the approach as described may be of limited value because of predefinition of multiplexed sets of loci to be tested. If dynamic construction and running of multiplexed sets is not possible, then there may be some value in nominating animals and sets of loci to be tested at each cycle.

ACKNOWLEDGEMENTS

I thank Richard Kerr, Gerard Davis, Jay Hetzel, Hugo Montaldo, and Mike Goddard for stimulating discussion, and a referee for useful comment. This work was carried out under project UNE038, funded by the Australian Meat Research Corporation. I thank the Twynam Pastoral Company, which funds my position.

REFERENCES

- Kerr, R.J. and Kinghorn B.P. (1996) *J. Anim. Breed. Genet.* **113**:457
Kinghorn, B.P. (1997) *Genetics.* **145**:479
Kinghorn, B.P. and Kerr, R.J. (1995) *Ass. Adv. Anim. Breed. Genet.* **12**:271
Kinghorn, B.P. (1999) *J. Anim. Breed. Genet.* In Press