Prediction Of Phenotypes Of Farm Animals: Interest Of Experimental Approaches

*E. Manfredi**

Introduction

The prediction of phenotypes by genetic means is a traditional research area in animal and plant breeding which has received increasing attention with the feasibility of genotyping a massive number of molecular markers. The topic is an active area of research in other contexts such as pharmacogenomics and individualized medicine.

Genomic prediction of phenotypes relies on the stability of marker (SNP) effects on phenotypes along time and across the population where prediction is to be performed. In the context of farm animal genetics, one major factor affects the reliability of prediction: the changes in the association between markers and phenotypes along generations and across structures created by directional selection.

A multi-generation experiment, using dense SNP chips of a species with sequenced genome, short generation interval and low maintenance costs (e.g., mice, rabbit) would provide unique data and results that are urgently needed in an international context where published results on multi-generation genomic prediction in selected animal populations are scarce and mainly based on simulated data. The observed changes of linkage disequilibrium (LD) patterns under carefully monitored selection would then be compared to theoretical expectations, and the existence of theoretical paradigms, such as the 'Bulmer effect', might be verified empirically.

Animal models, instead of simulated genomes, have been already used to test genetic models for prediction of phenotypes. As an example, recall that results of simulation studies by Meuwissen et al (2001) were somewhat misinterpreted and they generated a vague of optimism about the possibility to estimate marker effects in a sample of a population (training) and to use those estimates to select genotyped individuals in another sample (target) of the same population. Using real mice data Legarra et al. (2008) showed that reliability of prediction, when animals in the 'training' and 'target' populations were related, was higher than that observed in independent samples. In recent articles, Meuwissen (2009) and de Ross et al. (2009) confirmed by simulation the mice results and proposed approximated deterministic formulae for computing reliabilities of prediction for unrelated animals. This illustrates the value of model species for farm animals: results on true biological data can be used to verify or not the well accepted paradigms, and results can be general enough for extrapolation to other farm species. Selection experiments are complementary to simulations as well as to detailed work done on commercial populations.

* INRA UR631, SAGA, F31320 Castanet-Tolosan, France

Experiments on prediction of phenotypes

The new genomic tools available for animals allow to answer well known genetic questions, from the very general perspective on natural selection models (stabilizing, pleiotropic,... as reviewed by Johnson and Barton, 2005) or the 'objects' of selection (additivity, environment, partial epistasis...as reviewed by Hill, 2010) to more detailed aspects of artificial selection which we discuss here. For 'prediction' under artificial selection, several questions appear: what are the consequences, at the genomic level, of selection on phenotypes vs selection on markers? What decay of linkage disequilibrium under repeated cycles of artificial selection? What is the stability of marker effects along generations and across structures created by directional selection? What is the decay of reliability of prediction when genomic selection is performed without repeated phenotyping?

Some of these questions are being addressed fully with simulated data (e.g., Solberg et al. 2009) and partially with genomic data on commercial populations (e.g., Du et al., 2007 in pig; Wade et al., 2009 in horses; Abasht et al., 2009 in chicken; Gibbs and Taylor, 2009, Villa-Angulo et al., 2009 in cattle). From commercial populations, we are getting a general picture of genomes of breeds or strains under artificial selection: effective population size ranges from several hundreds to less than 30 (e.g. Abasht et al., 2009, in chicken lines) and most significant linkage disequilibrium is within 2 cM or within 5 cM according to authors who worked with varying marker densities and different measures of LD between markers. Associations between non syntenic markers were also reported (e.g. Farnir et al., 2000; Lipkin et al., 2009).

The question is whether or not experimental approaches can make complementary contributions to the original and useful results on animal genomes cited above. Three major features provided by experimental selection are: i) the control lines, ii) the possibility of designing simple selection schemes in order to get clear cut results, and iii) the control of key factors affecting genome architecture. As an example, consider the question on stability of marker effects along generations and its consequences on reliability of prediction. The experiment would have the following characteristics:

- Selection lines. Two lines might be compared in terms of LD decay and phenotype changes: no selection (control) vs. selection by the sum of marker effects, estimated before experimental selection starts. Complete data (phenotypes and genotypes) would be collected in both lines. The control line should give information on the combined effects of uncontrolled factors affecting the variability of measures, namely, recombination rate (overall and local), mutation, drift and natural selection.
- Selected trait. To simplify matters, univariate artificial directional selection might be performed on a single quantitative trait of moderate heritability, not subjected to maternal or imprinted effects. Genome regions governing the trait might be identified before and after selection. The behavior of selected and unselected regions would be compared along generations, and models for hitchhiking effects (e.g. Stephan et al., 2006) would be tested.
- Genotyping. A high marker density might be used to describe LD finely. Density is particularly critical to study persistence of accuracy of prediction along generations. Simulations suggest that marker densities should be high, up to 10*Ne/M (effective

population size per Morgan) when individuals for estimation of marker effects are not related to individuals whose phenotypes are being predicted (Meuwissen, 2009).
- Demography. The size of the experiment may be the limiting factor, as usual. Population size is determinant of the noise due to drift. For mice or rabbit, 100 to 200 parents per generation with many progeny providing phenotypes may be selected during 4/5 generations in 2/3 years. Also, in genomic prediction, it is necessary to warrant a minimum of phenotypes for estimating marker effects (e.g., at least 2000 phenotyped individuals in Meuwissen et al., 2001, to estimate 10000 marker effects).
- Selection. In some commercial data selection rates are not well monitored but they can be fully controlled in an experiment. In particular, having complete data on both sexes allows to explore different selection rates for each sex and give a more realistic picture of gametic phase disequilibrium along selection generations. A selection scheme with non-overlapping generations would simplify matters and this can be also governed in an experiment.
- Mating. For the questions cited above and for simplicity, mating should be at random. However, in other complementary experiments, many questions about mating would have immediate practical interest, e.g., inbreeding issues, or consequences of assortative vs. complementary mating (as simulated for major genes by Manfredi et al., 1998).
- Time and costs. The choice of a small experimental species includes gains in time (due to short generation interval) and reductions of costs (due to experimental facilities for small animals). But, genotyping remains an expensive operation specially when sample size must be kept high. Costs can seriously limit the possibilities to replicate these experiments.

Discussion - Conclusion

New genomic tools are progressing rapidly (e.g. Eid et al., 2008) and they can be regarded as modern "microscopes" in descriptive approaches to improve our statistical models for prediction and applied selection. A risk of a descriptive approach is the potential noise and high variance of observed trends together with the difficulties to warrant replicates of these expensive experiments. But these risks are even higher in the description of commercial populations. Further, in most simulation research, methods are compared under the assumption that the genetic model is known.

Besides evident practical implications of experimental results on LD decay under selection as commented above, the proposed studies may give insight on the adequacy of statistical models to capture genetic variation in farm animal populations. In particular, Population Association, which models the dependency of phenotypes on markers in a population in a simple and appealing way, influenced two important fields in genetics: the search of mutations responsible for phenotypic variation (Risch and Merikangas, 1996) and the prediction of phenotypes and genetic values (Meuwissen et al., 2001). Both articles generated new possibilities of research and successful applications in human genetics (e.g. Klein et al., 2005) and in animal breeding (e.g., VanRaden et al., 2009). However, results in both articles rely on simulated data where the association between phenotypes and genotypes is neat and probably simpler than in real life. Some disappointing results such as the poor explanation of polygenic variance of human height (as reviewed by Visscher, 2008) suggest that true polygenic effects are governing the trait, but also that simple marker-phenotype statistical models are unable to capture genetic variation of complex traits. Gianola et al.

(2009) discussed how elusive the relationships between marker variance and additive genetic variance can be. Experiments suggested above can give insight on the pertinence of our statistical models to improve the representation of genetic effects using genomic data. One might think that when the objective is prediction of phenotypes (and genetic evaluation), the need for a realistic representation of underlying genetic effects is less urgent, since the objective is to obtain good predictions as opposed to biological understanding. However, in genetic evaluation, the choice of the model (and of the statistical method) determines the ultimate scientific result: the selection rule, which should be robust along generations and across subpopulations of an animal population under artificial selection.

References

Abasht, B., Sandford, E., Arango, J. *et al*. (2009). *Bmc Genomics*, 10.

de Roos, A. P. W., Hayes, B. J., Spelman, R.J., *et al*. (2008). *Genetics*, 179(3): 1503-1512.

Du, F. X., Clutter, A. C., and Lohuis, M.M. (2007). *Int. J. of Biol. Sciences*, 3(3): 166-178.

Eid, J., Fehr, A., Gray *et al*. (2009*). Science*, 323(5910), 133-138.

Elston R.C., and Stewart J. (1971). *Human Heredity*, 21:523-542.

Farnir F., Coppieters W., Arranz J.J. *et al*. (2000). *Genome Research*, 10(2): 220-227.

Gianola, D., G. de los Campos, Hill, W.G. *et al*. (2009). *Genetics*, 183(1): 347-363.

Gibbs, R. A., J. F. Taylor, Van Tassel, C.P. *et al*. (2009). *Science*, 324(5926): 528-532.

Hill, W.G. (2010) *Phil. Transact. of the Royal Society*, B, 365, 73-85.

Johnson T., and Barton N. (2005) *Phil. Transact. of the Royal Society*, B, 360, 1411-1425.

Klein, R..J., Zeiss, C., Chew, E.Y. et al. (2005). *Science* 308, 385-389.

Legarra A., Robert-Granié C., Manfredi E. *et al*. (2008). *Genetics*,180(1), 611-618.

Lipkin, E., K. Straus, Stein, R.T. *et al*. (2009). *Genetics* 181(2): 691-699.

Manfredi, E., M. E. Barbieri, Fournet, F. *et al*. (1998*). Genet. Sel. Evol.* 30: 127-148.

Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (2001). *Genetics* 157(4), 1819–1829.

Meuwissen T.H.E. (2009). *Genet. Sel. Evol.*, 11, 41-35.

Risch, N. and Merikangas, K. (1996). *Science* 273(5281): 1516-1517.

Solberg, T.R., Sonesson, A.K., Woolliams J.A. *et al*. (2009). *Genet. Sel. Evol.*, 41:53.

Sonesson, A.K. and Meuwissen, T.H.E. (2009). *Genet. Sel. Evol.*, 41, article 37.

Stephan, W., Song, Y.S. and Langley, C.H. (2006). *Genetics*,172(4), 2647-63.

VanRaden, P. M., C. P. Van Tassell, Wiggans, G.R., *et al*. (2009*). J. Dairy Sci.* 92(1): 16-24.

Villa-Angulo, R., Matukumalli, L. K., Gill, C.A. *et al*. (2009). *Bmc Genetics* 10: 13.

Visscher, P. M. (2008). *Nature Genetics* 40(5): 489-490.

Wade, C., Giulotto, M.E., Sigurdsson, S. et al. (2009). *Science* 326(5954): 865-867.