WGS And MAS Expected Effects On Genome Architecture

***Ehud Lipkin*[*], Alessandro Bagnato[†] and Morris Soller[*]**

Introduction

Dairy cattle breeding has been very successful in the past five decades, due to the development of BLUP based estimation of breeding values for pedigree-selection of young sires for progeny testing, followed by progeny testing of large numbers of young sires, while choosing only the 10% most successful for widespread use. This has resulted in a genetic doubling of Holstein milk yield in 50 years. As a result, however, the Holstein genome is riddled with "selection signatures": small chromosomal regions that show a marked decrease in variability between cows, relative to adjoining chromosomal regions. This was shown by whole genome analysis using microarrays that can monitor tens of thousands of polymorphic sites (SNPs). The presence of selection signatures is attributed to selection favoring one particular allele in the region dragging the allele and its entire adjoining region to near fixation. For traits under selection, a "selection signature" is the signature of a QTL affecting the target trait, and hence might equally be termed a "QTL signature".

Concurrently, in last 20-30 years the Holstein breed has seen a noticeable deterioration in functional traits: Conception rate has gone from 60 to 30%, and embryo survival from 85 to 42%; there is increased incidence of mastitis in mature cows, and Holstein calves are notorious for their fragility (only 10% of Holstein calves in Israel reach slaughter without visible lung damage due to Bovine Respiratory Disease); average milking life of a cow has dropped from 3.4 to 2.8 lactations.

Selection signatures and functional trait deterioration

Selection signature formation and functional trait deterioration may be related. There have been a number of preliminary reports and personal communications at recent meetings indicating that selection signatures have positive effects on production traits and negative effects on functional traits (pers. comm., H. Schwarzenbacher, Technical University of Munich; Sonstegard *et al*., 2009). This may be due to direct pleiotropic effects of the alleles under selection for production traits, indirect effects generated by endo-environmental correlations, or hitchhike effects on linked genes affecting functional traits.

Increase in the weight of functional traits in current selection indexes has resulted in noticeable recent success in stopping further deterioration in conception rate and even reversing the

[*] *Department of Genetics, The Hebrew University of Jerusalem*

[†] *Department of Veterinary Sciences and Food Safety Technologies, University of Milan*

deterioration in milk somatic cell score (SCS, a marker for mastitis infection) and daughter conception rate (DCR) (Weller and Ezra, 2004; http://aipl.arsusda.gov/eval/summary/trend.cfm?R_Menu=HO.s#StartBody). However, there may be aspects of functional performance that are not captured by SCS or DCR. For example, the genetic correlation between SCS and clinical mastitis (CM) in only about 0.6, and it is well known that SCS is particularly sensitive to longstanding subclinical infection while CM presents as short-term acute infection that may be missed in the monthly SCS evaluation. Thus, some of the genes involved in resistance to CM may not be targeted by selection against SCS. More generally, natural selection molds a genome that is optimal with respect to functional traits. Artificial selection progressively modifies this genome to be optimal with respect to production traits. As the genome deviates from its functionally optimal architecture, functional traits can be expected to suffer. Although the deviation of genome architecture from the optimum for function as a result of selection for production traits may be linear on the selection applied, the response of the functional traits to this deviation may not be so. Thus, simple continuation of present selection practice may overwhelm the ability of the new index weightings to maintain functional performance of the dairy herd.

Potential deleterious effects of Whole Genome Selection (WGS)

This worrisome situation is about to be acerbated by the widespread introduction of Whole Genome Selection (WGS) to animal breeding in general and dairy cattle breeding in particular. WGS is a method for identifying animals of superior genetic merit based on analysis of the entire genome, using above-mentioned SNP microarrays to identifying SNPs that are in high correlation with trait breeding value. WGS will markedly increase the effectiveness of selection by increasing the accuracy and intensity of selection of bull mothers and of young bulls, and reducing the generation interval by eliminating the need for progeny testing. Thus, animal breeding in general and dairy cattle breeding in particular are on the verge of a quantum leap in the intensification of selection signatures and consequent intensification of the risks to functional performance that this entails.

Potential Positive contribution of Whole Genome Analysis (WGA)

Fortunately, the challenge presented by WGS may contain its own solution, at least in part. This lies in the ability of WGS to remedy a deficit in current index selection theory and practice and on the way reduce the impact of WGS on genome architecture. Currently, a selection index for an individual is evaluated by first obtaining estimates of the breeding value of the individual for all of the traits in the index. These are then weighted in the index according to their economic/functional importance, and also in accord with their biometrical genetic correlation with the other traits in the index. But, as will be seen from the example of milk yield and protein percent presented below, the actual pleiotropic effect of a locus on a pair of related traits can vary greatly between loci. Under current index selection, loci that carry a strong pleiotropic effect are not penalized for the reduction of functional traits to the extent that they should, so that if the overall genetic correlation is low, they may go to fixation in spite of the counter selection for functional traits. Conversely, production traits loci at which the pleiotropic effect is absent, are

penalized as if it were present, reducing potential progress in production traits, achievable at these loci. Thus, most of the selection for functional traits may be expended on loci that do not have such effects at all!

The advent of WGS will present opportunities for high resolution mapping and high resolution estimation of QTL allele effects. This will become possible through the large amounts of LD mapping data that are provided by the whole genome analysis (WGA) that is an essential prior step of the WGS approach. This data should allow direct and pleiotropic effects of chromosomal regions under selection (the "selection/QTL signature" regions) to be evaluated with unprecedented accuracy. By focusing selection on production QTL that have minimal deleterious functional effects, while limiting selection at production QTL having large associated functional effect, the impact of selection on genome architecture can be reduced, while still maintaining good progress in production traits.

An actual example: protein yield

Protein is now the major product of the dairy herd. Protein yield (PY) is the product of milk yield (MY) and milk protein percent (PP), i. e., PY = MY*PP. Consequently, selection for both MY or PP should increase PY. The genetic correlation of MY and PP is negative, r_G= -0.5 (reviewed in Lipkin *et al*, 2008b). Biometrical Selection for PY takes this into account when weighting MY and PP in the selection index. Current status of Israel Holstein: Mean MY = 10,000 kg; Mean PP = 3.3%; so that Mean PY = 10,000*0.033 = 330 kg.

QTL mapping for MY and PP (Bagnato *et al*., 2008; Lipkin *et al*., 2008a) shows that about 50-55 QTL affect these traits (see also recent review of milk production QTL by Ogolrevc *et al*., 2009). Of these QTL, about 25% (say, ~13) exclusively affect MY (excMY QTL), 25% (say, ~13) exclusively affect PP (excPP QTL) 50% (say, ~26) affect both traits (MY-PP QTL). By definition r_G = 0 at excMY and excPP loci; and consequently must equal -1.0 at MY-PP loci. By considering total variation in EBV for MY and PP, the above number of assumed loci affecting these traits, assuming that allele frequency at these QTL is not far from 0.4 and allele effects are the same whether the locus exclusively affects MY or PP or whether it affects both traits, we can estimate average allele substitution effect at all QTL affecting MY=250 kg, and average allele substitution effect at all QTL affecting PP = 0.04% (See Lipkin *et al*., 2008b for details). Furthermore, since genetic correlation at the loci affecting both traits is -1.0, the positive allele for MY at these loci will have effects +250 kg MY and -0.04% PP, so that new PY at fixation for an individual allele of this nature will be: 10,250*0.0326 = 334.15, increase of 4.15 kg. The positive allele for PP at these loci will have effect -250 kg MY and +0.04% PP, so that PY at fixation for an individual allele of this nature will be 9,750*0.0334 = 325.65, decrease of 4.35 kg (for further explanation of this seeming paradox see Lipkin *et al*., 2008b).

We now consider the effects on PY of selection to fixation for the various classes of loci.
Selection to fixation for excMY and excPP QTL only:

13 excMY QTL x 250 kg/locus = +3,250 kg

New mean MY: =10,000 + 3,250 = 13,250
13 excPP QTL x 0.0004/locus =+0.0052 PP
New Mean PP: 0.033 + 0.0052 = 0.0382
New Mean PY = 13,250*0.0382 = 506.15 kg; increase of **176.15** kg protein.

Selection to fixation at loci affecting MY and PP.
Selection for allele with positive effect on MY:
New mean MY = 10,000 + 26*250 = 16,500
New mean PP = 0.033 – 26*0.0004 = 0.023
New mean PY = 379.5 kg; increase of only **49.5** kg protein.
Thus, Selection for the positive allele at the MY-PP loci contributes very little to increase in PY, yet introduces another 26 selection signatures!

Selection to fixation at all classes of loci:
New mean MY = 10,000 + 39*250 = 19,750
New mean PP = 0.033 + 13*0.0004 – 26*0.0004 = 0.0278
New mean PY = 19,750*0.0278 = 549.05; increase of **219.05** kg protein.

Selection at all loci increases total response by only 42.9 kg (24.4%) compared to selection at excMY and excPP loci alone, but doubles the number of selection signatures (from 26 to 52). Thus, MAS targeted to specific loci, may provide a means of minimizing deleterious effects while continuing rapid increase in productivity. This will not be possible with GWS aimed at increasing overall index breeding value.

References

Sonstegard, T. S., Ma, L., Cole, J. B., *et al*. (2009). *In Proc PAG XVII*, pages 10-14.

Bagnato, A., Schiavini, F., Rossoni, A., *et al*. (2008). *J. Dairy Sci*., 91:767-783.

Lipkin, E., Tal, R., Friedmann, A. *et al*. (2008a). *J. Dairy Sci*., 91:1614–1627.

Lipkin, E., Bagnato, A. and Soller, M. (2008b). *J. Dairy Sci*., 91:2857–2863.

Ogorevc, J., Kunej, T., Razpet, A. *et al*. (2009). *Anim. Genet*. 40:832-851.

Weller, J. I. and Ezra, E. (2004). *J. Dairy Sci*., 87:1519-1527.