



REVIEW

## 10th WCGALP in beautiful Vancouver

R.J.C. Cantet<sup>1</sup>, O.F. Christensen<sup>2</sup>, M. Pérez-Enciso<sup>3</sup> & J.H.J. van der Werf<sup>4</sup>

<sup>1</sup> Universidad de Buenos Aires, Buenos Aires, Argentina

<sup>2</sup> Aarhus University, Aarhus, Denmark

<sup>3</sup> ICREA – Centre for Research in Agrigenomics (CSIC-IRTA-UAB-UB), Barcelona, Spain

<sup>4</sup> University of New England, Armidale, NSW, Australia

### Correspondence

Asko Mäki-Tanila,

University of Helsinki, Finland

E-mail: Asko.Maki-Tanila@helsinki.fi

Received: 28 August 2014;

accepted: 29 August 2014

The 10th World Congress was inaugurated by organizers Filippo Miglior and John Pollak in Vancouver at 8 pm on Sunday 17 Aug, preceded by a cocktail to warm up attendees' epigenomes. We return to these congresses each time in higher numbers, now over 1500 participants. The arrangements were very good and the weather cherished us all week, including the boat trip out to open sea among the small hydroplanes whirling up and down around us on the water. The new technology was adopted in presenting the posters (of rather dated outlay though) and the talks could now be easily found by author names and also re-listened to at the congress web site. It is not easy to itemise separate themes or avoid overlaps in reviewing the congress, where the sessions were thoroughly filled or hollowed by our extensive genome-wide studies.

### From sequence to presequence (Miguel Pérez-Enciso)

In the review on Leipzig's WCGALP, I predicted that the Vancouver meeting would be flooded by sequences: 'At the next World Congress, Vancouver 2014, complete genomes will be as popular as SNP microarrays were at the Leipzig venue' (Pérez-Enciso 2010, *J Anim Breed Gen* **127**, 338). I was wrong: Vancouver WCGALP was overwhelmingly dominated by genomic selection (GS) issues with QTL – GWAS studies being never more popular in animal sciences. This is partly due to the clear practical focus of WCGALP but does not explain the whole variance among the communication dataset from the meeting.

Among the oral presentations, 15 were on next-generation sequencing (NGS) data versus 77 on genomic selection, 23 on population genetics topics (selection footprint, variability), 47 with GWAS / QTL approaches and 9 on systems biology (networks, pathways). For the posters, the numbers were 6, 61, 30, 59, and 6, respectively. (These numbers are mainly based on reading the title of communications and are subjective to an extent). In any case, NGS talks were overrepresented suggesting that NGS was considered a hot topic while the animal NGS field has not exploded yet in its entirety. The same seems to be the case for systems biology. Genomic selection was actually under represented among oral contributions, possibly due to frequency dependent selection.

For the observations above I have counted only genome wide sequence: RNAseq has become a popular tool (20 papers in the congress) than genomic NGS. The likely reasons are the lower costs and richer information. For example, e.g., RNAseq can be used not only to measure overall expression, but also allele specific expression (Bill Muir) or to refine the annotation. If one extrapolates the prospective avenues from the on-going human research, it can be expected epigenome and metagenome to be quite popular targets in the near future. There were <10 papers on them in the meeting.

Most of the sequence data reported were on cattle, (over 2000 genomes). It seems that the most fruitful application of NGS was in detecting monogenic, deleterious mutations. Michel Georges and Richard Spelman used Hardy–Weinberg deviation to find lethal

mutations affecting embryo development. They used some 500 bull sequences (7X on average) with imputation for 10 million animals. Aurélien Capitan et al. at INRA further isolated several causal mutations affecting rare syndromes using the 1000 bull genome data, which consists of 1147 highly influential bulls from 27 breeds sequenced at on average of 11X. Using the same data Ben Hayes found that with BayesR the accuracy of genomic selection accuracy is only some 2% higher than with dense marker set. Imputation accuracy was 90% for minor allele frequency (MAF) > 0.1, and dropped dramatically for lower MAF. The issue, of course, is that low MAF variants are the most frequent ones in sequence data, and this could be one of the reasons why complete sequence did not help.

After all, the subdued appearance of NGS data in the meeting is possibly caused by the difficulties in their analyses, which are much greater than anticipated. Such data require costly computer resources and are noisy for the scale needed in animal breeding. Their analyses are also complicated due to limited sequencing depth (which introduces uncertainty in SNP calling). We have observed significant changes in the SNP calling process when using two different versions of the same software like *samtools mpileup* tool. Jerry Taylor confessed he had become a NGS addict, with 500 bulls sequenced. So did I, but I am in a detoxifying treatment for the given reasons. As penitence, I devote now most of my time developing analytical tools to make the most out of the data already available, and to optimize experimental designs before it is too late.

Several communications (e.g., Jerry Taylor, Vincent Ducrocq, Ben Hayes, Mike Goddard) revolved around the utilization of the causal mutations in selection for complex traits, which seems a bit counterintuitive in the genomic selection paradigm. Several decades of research and sequencing have proven how difficult this is, even if the causal mutations are present in your data. In WCGALP several authors, Peter Sørensen and Mike Goddard, among others, recognized the importance of using biological information for prediction purposes. I fully agree but an accepted or meaningful way to do so remains to be elucidated. Perhaps, one could start by recognizing that not all SNPs are born equal and introduce annotation in the model. Tools like variant effect predictor in *ensembl* classify SNPs according to their expected degree of severity. When using sequence, this information can be readily taken into account in the priors. However, this is not relevant for genotyping arrays because most chip SNPs will be intergenic and likely neutral *per se*.

In this paradigm, sequence data could make a difference. As you can guess, I do not dare doing any prediction for New Zealand's event, though.

### Genomic selection matures (Ole F. Christensen)

A very interesting symposium was about industry applications of GS. The session started from dairy cattle with Esa Mäntysaari's historical perspective about the enormous impact of GS on the sector. For poultry, Anna Wolc presented the results from a multi-generation GS and experiences about the GS implementation in broilers and layers. In poultry the applications arrived later than in dairy cattle, primarily because of the prohibitive expenses of genotyping relative to the value of individual selection candidates. Atlantic salmon is a very different species due to its much later domestication (only some ten generations ago). The present population is an admixture. Another feature is the high fecundity both in males and females. Jørgen Ødegaard praised the high potential of GS in aquaculture and compared GS models in a two-trait context (lice resistance and fillet colour). There are clearly species specific issues in the GS applications.

Several presentations were about methodology for single-step genomic evaluation (ssGBLUP) using a hybrid relationship matrix with a good overview given by Andres Legarra. Ismo Strandén presented an equivalent equation system for solving ssGBLUP without constructing and inverting the pedigree relationship matrix for genotyped animals. Dorian Garrick formulated ssGBLUP as a SNP effect model. From a conceptual point of view it is very important to have the two equivalent formulations. Zenting Liu said that SNP model allows excluding/including specific animals in the training data.

Prediction across breeds was a topic with many presentations and several groups are developing useful approaches. Mario Calus concluded that the predictions using information across breeds benefit from few closely related individuals while some individuals can deteriorate the predictions. Shared large effect QTL's improve the prediction across breeds (Mahdi Saatchi and Dorian Garrick), and they could be detected from imputed whole genome sequence (e.g. Rasmus Brøndum).

In addition to the many sessions focusing by name on GS, there were presentations about genotyping and phenotyping strategies, and presentations where GS (or ssGBLUP) was not of primary interest but a natural part of the genetic evaluation. GS is now more mature and over the highest uncontrolled enthusiasm. Many of the challenges seen with pedigree-based

genetic selection are still present. Because the cost of genotyping is steadily coming down, there will be many more marker based evaluation programs. By the next WCGALP, I would expect to see many studies where the main focus is not on GS but marker genotypes (or causal variants) are included in the genetic model.

### **Developments in quantitative genetics (Julius H. J. van der Werf)**

Quantitative genetics is the foundation of much of the work in animal breeding. At the conference it was covered by 'Breeding objectives, economics of selection schemes, and advances in selection theory' but appeared in many other topics. The amazing genomic toolbox requires sound quantitative genetic theory to underpin models of analysis while challenging assumptions. So genomics is causing a revolution similar to the one almost hundred years ago when Ronald Fisher and Sewall Wright proposed to use pedigree information to enhance genetic analysis of quantitative traits.

The genome wide association studies give us a clue about the size and distribution of gene effects that control quantitative genetic variation, and further about gene by gene and gene by environmental interactions. The analyses with dense markers give information about the level of heterozygosity, or absence thereof, genetic diversity, and signatures of selection. Molecular information also provides a tool to gain a greater insight about identity of descent at the level of a single locus, and from that we can derive coefficient of covariance for a range of genetic effects. For example, dominance variation can be estimated based on variation in genomic dominance relationships, and there is no longer a requirement of having full-sib families. Theo Meuwissen pointed out how the genomic prediction is contributed by information on pedigree, linkage and linkage disequilibrium. The relative importance depends on the true genetic model, with the linkage based approach being more important for large QTL effects. The veil of the underlying genetic model is slowly being lifted.

Some believe that the missing heritability problem is mainly due to non-additive genetic effects (e.g. Zuk et al, *PNAS* 109: 1193). However, Asko Mäki-Tanila and Bill Hill showed that epistatic effects rarely contribute much to the observed variation and taking them into account in either selection or GWAS strategies is unlikely to have a large impact. These studies are good attempts to reconcile the top-down and bottom up approaches into quantitative genetic analysis, as suggested by Eric Lander at the International

Quantitative Genetics Symposium in Edinburgh in 2012. With sequence data we are able to detect more causal variants, but we are far from explaining the observed (additive) genetic variance with detected QTL effects and terabytes of data grinding should be passed to resolve for the explanation. The resurgence in the hunt for QTL is a logical next step in the genomic prediction models. Hopefully, we can now make use of the lessons learned more than a decade ago. E.g. the selection on QTL is less optimal as the joint selection on QTL and polygenic background. This would be difficult to achieve if genomic selection was based on just a few QTL, suggested by Jerry Taylor for the use of sequence information.

The plain phenotypic variation could be analysed by new genetic models, e.g. in the analysis of maternal and social effects, traits measured on trajectories and genotype by environment interactions. These could reveal nonlinear relationships between traits as shown by Han Mulder, and these effects could also be selected upon. Epigenetic studies had not arrived in large numbers at the WCGALP in spite of being, for some years now, a hot topic in human genetic analyses. Only eight studies looked at epigenetics, with simple variance components or with gene expression or methylation patterns. I would expect that such studies will also become more common in animal genetics with interesting phenomena as an outcome. Yaodong Hu, Guilherme Rosa and Daniel Gianola showed that imprinting could lead to a significant reduction in the GWAS heritability.

It was good to see that optimal contribution selection has now become part of the regular animal breeding toolbox. John Woolliams considered cases where selection accuracy is equal to one and stated that 'if accuracy does not approach one with huge numbers, then the community needs to completely overhaul the basis of its most cherished models for genetic evaluation'. We'll have to see!

In the symposium on Breeding Objectives some excellent insights were presented relevant to achieving successful outcomes in animal improvement programmes. The debate continues because the assessment of utility that can vary between circumstances and people. The area lacks a comprehensive theory, and if anything, we were made aware that the existing framework of (linear) selection index principles are rarely found adequate when determining relative trait emphasis in multiple trait selection practices (Pieter Knap; Rob Banks). Jack Dekkers made it clear that the way in which breeding objectives are achieved largely depends on the information available per trait, and e.g. genomic information may change

the direction of genetic progress. It is somewhat ironic that the availability of genomic markers has sparked more interest in phenotyping, not less, with several sessions devoted to phenotyping for traits difficult to measure. At the end of the day, the genomic information just helps us to make better inferences but the phenotypic information remains the basis of genetic improvement.

### **Statistical animal breeding, genomics and prediction of breeding value (Rodolfo J. C. Cantet)**

Animal breeders have used a plethora of statistical methods to predict breeding values when adding genomic information to phenotypes and to pedigree data. At the Leipzig WCGALP two different models for prediction of breeding value entered the genomic arena: the independent multiple marker (SNP) model by Theo Meuwissen and co-workers (2001 *Genetics* 157, 1819), and the infinitesimal animal model with the covariance matrix of independent markers defined by Paul VanRaden (2008 *J Dairy Sci* 91, 4414), as worked out by Ignacy Misztal and co-workers (see Legarra et al. 2009 *J Dairy Sci* 92, 4656) ssGBLUP. In Vancouver we now witnessed (un)intentional efforts to converge from either side to a model that takes into account the genetic architecture of the trait and be consistent with the infinitesimal model that has served so well until now. Reflecting my thoughts as somebody dealing with genetic evaluation for beef cattle: (i) I try to avoid situations where I have to compare the genomic and conventional methods as the existence of two different predictions of breeding value affect the confidence on either method; (ii) If a bull or cow has no new phenotypic data but genomic information (or not), it is easy to explain a change in the predicted breeding value from ssGBLUP. However, this is not easy from the SNP model, in particular if imputation and Bayesian algorithms have been used.

The basic difficulty with both the models is that under linkage and linkage disequilibrium, genome segments rather than SNPs are transmitted over gen-

erations. Hence there are neither independent SNP effects nor real permutable markers to calculate genomic relationships (see Thompson 2013 *Genetics* 194, 301). Therefore VanRaden's predictor of the true genomic relationship partially captures the pattern of inheritance under Hardy–Weinberg, but it does not take into account the genetic architecture (variable gene effects over the genome) behind the trait variation.

The first session on Monday morning evidenced all of this. Dorian Garrick and Vincent Ducrocq presented models accommodating individual markers and polygenic effects. Gustavo de los Campos gave a clever presentation on the difficulties in equating the estimate of the additive genetic variance from the SNP model with the one from the infinitesimal model. The former works well for prediction, but it does not allow the estimation of additive variance. The search for the covariance matrix and the linear prediction model that takes into account genomic information, was presented by Gregor Gorjanc, and also by Zulma Vitezica on metafounders. I showed with her how genomic information is utilized by the classic regression approach of breeding values across generations to track the Mendelian sampling effects and thereby account for more additive variance than with the conventional animal model.

When the (true) underlying genetic model is undefined, the accuracy of prediction cannot be defined the usual way and we cannot compare the two methods of utilizing genomic information. We have been performing genetic evaluation over decades assuming the infinitesimal model, as it has proved to be consistent with the observations. Usually overlooked by animal breeders, asymptotic theory arguments formalizing the infinitesimal model have been given by Lange (1978, *J Math Biol* 6, 59) and by Abney et al. (2000, *Am J Hum Genet* 66, 629). Hopefully we are soon able to perform genetic evaluation (like Dr. Henderson taught us in the non-genomic era) having all the benefits of genomic information and taking into account the genetic architecture of the trait.