

**Wageningen  
The Netherlands**

**20-21 April  
2009**

**13<sup>th</sup> Quantitative Trait Locus and  
Marker Assisted Selection Workshop**

**QTL MAS**  
GGTCA  
ATCAAGTC



# CONTENTS

<b>CONTRIBUTIONS AND SUPPORT .....</b>	<b>2</b>
<b>PROGRAM .....</b>	<b>3</b>
<b>ABSTRACTS.....</b>	<b>7</b>

## **CONTRIBUTIONS AND SUPPORT**

The QTL-MAS 2009 workshop has been organized jointly by Biometris, the Animal Breeding and Genomics Centre, and Wageningen UR Plant Breeding, all part of Wageningen University and Research Centre.

Organizing committee:

John Bastiaansen

Marco Bink

Mario Calus

Chris Maliepaard

This book of abstracts has been edited by the organizing committee.

The following sponsors have provided **financial support** to the QTL-MAS 2009 workshop:

**EADGENE**

**LEB fonds**

**Pioneer Hi-Bred International**

**Illumina**

**CRV**

**TTI Groene Genetica**

**SABRE**

**CBSG (Centre for BioSystems Genomics)**

**IPG**

**Keygene**

**Roche**

# PROGRAM

**Monday April 20th**

**Morning**

**Location : Hof van Wageningen (WICC)**

*08:30 09:00 Registration / Coffee & Tea*

**Sponsor : SABRE**

**Chair : Johan van Arendonk**

---

<i>09:00</i>	<b>Session I</b>	<b>Introduction / understanding trait complexity</b>
09:00	LOC qtlmas2009	Welcome and Introduction
09:15	<b>Roel Ophoff</b>	<b>Genomic Perspectives of Schizophrenia (invited)</b>
09:45	Wilco Ligterink	Unraveling the complex trait of seed quality in tomato
10:00	Lars Ronnegard	Detecting variance-controlling QTL
10:15	Albart Coster	Simulation of the QTLMAS 2009 common dataset

*10:30 11:00 Coffee, Tea*

---

<i>11:00</i>	<b>Session II</b>	<b>Methods for the estimation of genomic breeding values</b>
11:00	Hans-Peter Piepho	Ridge regression and extensions for genome-wide selection in maize
11:15	David Habier	A mixture genetic model for whole-genome analyses
11:30	Hans Daetwyler	A comparison of genomic BLUP and BayesB at various population and trait genetic architectures
11:45	Pascal Croiseau	Use of the Elastic-Net algorithm for genomic selection in dairy cattle
12:00	Tomasz Suchocki	Detecting the Association of SNPs and Prediction of Future Yields for a Growth Trait Using a Mixed Model with Orthogonal Polynomials
12:15	Kacper Zukowski	Repeatability model and kriging as methods for the estimation of growth curves

**12:30 13:30 Lunch**

**Sponsor : Illumina**

**Monday April 20th**

**Afternoon**

**Location : Hof van Wageningen (WICC)**

**Sponsor : CRV**

**Chair : Roel Veerkamp**

<b>13:30</b>	<b>Session III</b>	<b>Integrating new tools and methods</b>
<b>13:30</b>	<b>Maarten Koornneef</b>	<b>Quantitative genetics in Arabidopsis: from QTL to QTN (invited)</b>
14:00	Ronny Joosen	Genes for seed quality: New tools and approaches
14:15	Saber Qanbari	Footprints of Recent Positive Selection in Holstein Cattle
14:30	Lucy Crooks	An improved method for calculating line origin probabilities for crosses between outbred lines
14:45	Stefan Marklund	Extraction of QTL-informative SNPs using genome sequencing data from founder line DNA pools in chicken
<b>15:00</b>	<b>15:30</b>	<b>Coffee, Tea, Cold drinks</b>
<b>16:00</b>	<b>18:00</b>	<b>Session IV</b>
		<b>Common dataset : Estimation of breeding values</b>
16:00	Ricardo Pong-Wong	A two-step approach combining the Gompertz growth model with genomic selection for longitudinal data
16:15	Klara Verbyla	Sensitivity of Genomic Selection to using different prior distributions
16:30	Zhe Zhang	The prediction of genomic breeding values for longitude traits using random regression
16:45	Matthew Cleveland	Genomic breeding value prediction using three Bayesian methods and application to reduced density marker panels
17:00	Sebastian Mucha	Comparison of methods for estimation of genetic covariance matrix from SNP or pedigree data utilised to predict breeding value
17:15	Nicola Macciotta	Prediction of Genomic Breeding Values for a longitudinal trait by using Principal Component Analysis
17:30	Torben Schulz-Streeck	Genome-wide selection by mixed model ridge regression and extensions based on geostatistical models
17:45	LOC qtlmas2009	Comparative analysis of submitted Breeding Values and applied methods
<b>18:00</b>	<b>19:00</b>	<b>Drinks</b>
		<b>Sports Bar Down Under</b>
<b>19:00</b>	<b>21:30</b>	<b>Workshop dinner</b>
		<b>Terraszaal</b>

**Tuesday April 21st**

**Morning**

**Location : Junushoff**

*08:30 09:00 Coffee & Tea*

**Sponsor : Pioneer Hi-Bred International      Chair : Fred van Eeuwijk**

*09:00 10:30*    **Session V**    **Methods for QTL detection and association analysis**

---

09:00	Andres Legarra	Joint association and linkage QTL mapping in half-sib families by regression
09:15	Xia Shen	Monte Carlo Full Likelihood Approach in Variance Component QTL Analysis
09:30	Jose Alvarez-Castro	On the estimation of genetic effects for multiallelic systems
09:45	Cajo ter Braak	Identity by descent probability matrix decomposition by a latent ancestor allele model and its application in QTL analysis
10:00	Joseph Powell	Factors influencing the optimum multiple marker model for analysis of association data

*10:15 10:45 Coffee, Tea*

*10:45 12:15*    **Session VI**    **Applications on LD and QTL detection**

---

10:45	Anna Johansson	Linkage disequilibrium in two divergently selected chicken lines
11:00	Katrin Mackenzie	Patterns of Linkage Disequilibrium in Cultivated Barley
11:15	Weronica Ek	QTL analysis of vitiligo in an outbred chicken intercross
11:30	Maria Siwek	A QTL on GGA14 – story of detection and validation
11:45	Dindo Tabanao	Genetic relatedness and population structure in analysis of linkage disequilibrium in barley
12:00	Filippo Biscarini	Across-line SNP association study of innate and adaptive immune response in laying hens

*12:15 13:30 Lunch*

Tuesday April 21st

Afternoon

Location : Junushoff

Sponsor : **TTI Groene Genetica**      Chair : Richard Visser

---

<i>13:30</i>	<i>15:30</i>	<b>Session VII</b>	<b>Common dataset : QTL detection</b>
13:30	Roel Veerkamp		Simultaneous QTL detection and genomic breeding value estimation using high density SNP chips
13:45	Olivier Demeure		QTL detection for a medium density SNP panel: comparison of different LD and LA methods
14:00	Georgia Hadjipavlou		Extensive QTL and Association Analyses of the QTLMAS2009 Data
14:30	Henri Heuven		Bayesian multi-QTL mapping for growth curve parameters
14:45	Hossein Yazdi		QTL analysis of QTLMAS2009 dataset through variance components including haplotype information and Identity-By-Descent probabilities (IBD)
15:00	LOC qtlmas2009		Comparative analysis of submitted QTL results and applied methods

*15:15*    *15:45*    **Coffee, Tea, Cold drinks**

---

<i>15:45</i>	<i>17:00</i>	<b>Session VIII</b>	<b>Selection with markers</b>
15:45	<b>Alfred de Vries</b>		<b>The application of Genomic Selection in dairy cattle (invited)</b>
16:15	Han Mulder		Marker-assisted breeding value estimation with missing marker genotypes
16:30	Jack Dekkers		Response and inbreeding from genomic selection
16:45			Closing remarks

# **ABSTRACTS**

# Genomic Perspectives of Schizophrenia

Roel A. Ophoff<sup>1,2</sup>

<sup>1</sup> Department of Medical Genetics, UMC Utrecht, The Netherlands

<sup>2</sup> Center for Neurobehavioral Genetics, UCLA, Los Angeles, California, USA

Schizophrenia is a complex disorder, caused by both genetic and environmental factors and their interactions. Research on pathogenesis has traditionally focused on neurotransmitter systems in the brain, particularly those involving dopamine. Schizophrenia has been considered a separate disease for over a century, but in the absence of clear biomarkers, diagnosis has historically been based on signs and symptoms. A fundamental message emerging from genome-wide association studies (GWASs) of copy number variations (CNVs) associated with the disease is that its genetic basis does not necessarily conform to classical nosological disease boundaries. Certain CNVs confer not only high relative risk of schizophrenia but also of other psychiatric disorders. The structural variations associated with schizophrenia can involve several genes and the phenotypic syndromes, or the “genomic disorders“, have not yet been characterized. Single nucleotide polymorphism (SNP)-based GWASs with the potential to implicate individual genes in complex diseases may reveal underlying biological pathways. However, the identification of common variants associated with neuropsychiatric traits has not been successful so far. The question emerges whether alternative (genomic) approaches are needed for the identification of the genetic susceptibility of schizophrenia.

# Unraveling the complex trait of seed quality in tomato

Wilco Ligterink<sup>1\*</sup>, Rashid Kazmi<sup>1</sup>, Noorullah Kahn<sup>1</sup>, Leo Willems<sup>1</sup>, Henk Hilhorst<sup>1</sup>

<sup>1</sup>Laboratory of Plant Physiology, WUR, Wageningen, Netherlands

\* wilco.ligterink@wur.nl

The yield and economic success of horticultural crops depends to a large degree on the quality of the seed used to grow these crops. Seed quality attributes include dormancy, germination, seed and seedling vigour, seedling dry weight, normal embryo- and seedling morphology, as well as the ability to develop into a normal plant. The molecular-genetic dissection of the processes that underlie these quality parameters and their relationship with seed and seedling phenotypes will identify the regulatory genes and signaling pathways involved and, thus, provide the means to predict and enhance seed quality.

Our aim is to elucidate the mechanisms involved in the acquisition of seed quality and to develop molecular markers to aid in marker assisted breeding. To reach this aim we make use of the natural variation found in tomato RIL and IL populations. We are in the process of phenotyping a broad range of seed quality attributes in these populations and have already identified QTLs for various traits. Besides extensive phenotyping, the RIL lines will also be used for a metabolomics study and transcriptional profiling. This combined use of physiology, genetics and genomics, followed by advanced data analysis and in combination with a 'likely candidate gene approach' will allow us to relatively quickly identify genes that are responsible for quality related traits of seed and seedling. Subsequent analysis of the relevant genes by reverse genetics, using knock-out and overexpression mutants, will be employed to unambiguously confirm their function.

# Detecting variance-controlling QTL

Lars Rönnegård<sup>1,2\*</sup>

<sup>1</sup> Statistics Unit, Dalarna University, Borlänge, Sweden

<sup>2</sup> Department of Animal Breeding & Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

\* lrn@du.se

In experimental intercrosses, large genetic differences between the means of the founder breeds are expected for the studied trait. The main focus in these studies has been on mean differences between allele effects. There is substantial evidence, however, that not only the mean but also the environmental variance may be under genetic control in most organisms. The aim of the present study is to develop the commonly used regression based QTL model and extend it for detection of variance-controlling QTL. The proposed model consists of a mean part and a dispersion part. The mean part of the model includes Haley-Knott regression coefficients as covariates, where the squared residuals are subsequently used to model the dispersion part. A generalized linear model with gamma distribution is used to fit the squared residuals and Haley-Knott coefficients are included in the linear predictor. The residual variance is thereby fitted at each tested position along the genome and we obtain a genome scan for QTL controlling the residual variance. Simulations from an  $F_2$  pedigree with 800 individuals in the last generation showed that the method is able to detect QTL controlling merely 2% of the variation in the model for the residual variance. However, the proposed QTL detection model can detect either variance-controlling QTL or mean-controlling QTL that are not fixed within founder lines. To distinguish the two from each other the previously presented FIA model is required. The proposed QTL detection model, together with FIA, gives us a powerful method to study the genetic control of environmental variance.

## QTLMAS 2009: Simulated Dataset

Albart Coster<sup>1</sup>, John Bastiaansen<sup>1</sup>, Mario Calus<sup>2</sup>, Chris Maliepaard<sup>3</sup>, Marco Bink<sup>4</sup>

<sup>1</sup> Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands,

<sup>2</sup> Animal Breeding and Genomics Centre, Animal Sciences Group, Wageningen University and Research Centre, Lelystad, The Netherlands,

<sup>3</sup> Plant Breeding, Wageningen University, Wageningen, The Netherlands

<sup>4</sup> Biometris, Plant Research International, Wageningen, The Netherlands

The simulation of the data for the QTLMAS 2009 workshop is described. The data consisted of markers, phenotypes and pedigree. Genotypes of 453 markers, distributed over 5 chromosomes of 1 Morgan each, were simulated for 2,025 individuals. From those, 25 (20 females and 5 males) were the parents of the other 2,000 individuals (100 full sib families). The 25 parents were genetically related due to 1,000 generations of random mating. Simulated phenotypic observations of a complex longitudinal trait were available for 1,000 of the 2,000 offspring individuals, at five time points. The phenotypes were simulated according to a logistic growth curve. The logistic growth curve was specified by three parameters. Each parameter was influenced by six QTL that were positioned at the five chromosomes. For each parameter one QTL had a large effect and five QTL had small effects. The simulation was performed with the R package HaploSim, available from the R repository CRAN at <http://cran.r-project.org/package=HaploSim>.

# **Ridge regression and extensions for genome-wide selection in maize**

Hans-Peter Piepho<sup>1\*</sup>, Torben Schulz-Streeck<sup>1</sup>

<sup>1</sup>Universität Hohenheim, Bioinformatics Unit, Fruwirthstrasse 23, 70599 Stuttgart, Germany

\* piepho@uni-hohenheim.de

This paper reviews properties of ridge regression for genome-wide (genomic) selection and establishes close relationships with other methods to model genetic correlation among relatives, including use of a kinship matrix and the simple matching coefficient as computed from marker data. A number of alternative models are then proposed exploiting ties between genetic correlation based on marker data and geostatistical concepts. A simple method for automatic marker selection is proposed. The methods are exemplified using a series of experiments with test-cross hybrids of maize conducted in five environments. Results underline the need to appropriately model genotype-environment interaction and to employ an independent estimate of error. It is also shown that accounting for genetic effects not captured by markers may be important. Finally, the methods are illustrated using the data of the QTL-MAS workshop.

Piepho HP 2009 Ridge regression and extensions for genome-wide selection in maize. *Crop Science* (May-June issue)

# **A mixture genetic model for whole-genome analyses**

D. Habier<sup>1</sup>, L. R. Totir<sup>2</sup> and R. L. Fernando<sup>3</sup>

<sup>1</sup>Institute of Animal Breeding, Christian-Albrechts University of Kiel, Germany

<sup>2</sup>Pioneer Hi-Bred International, Inc., Johnston, IA 50131, USA

<sup>3</sup>Department of Animal Science, Iowa State University, Ames, USA

presenting author D. Habier, dhabier@gmail.com

Accuracy of genome-assisted breeding values (GEBVs) can be improved and the decrease of accuracy with genetic relationships might be reduced by using information from both linkage disequilibrium (LD) and co-segregation. Mixture genetic models, in contrast to linear models, have the advantage that covariance matrices do not enter into the analysis and accommodating dominance, epistasis and imprinting is more straightforward. To make computations feasible for whole-genome analysis using dense marker data and large complex pedigrees, an approximate two-step approach is proposed that uses an overlapping-blocking Gibbs algorithm for sampling marker allele states and origins, but neglects information from trait phenotypes to sample those variables. The objective of this study was to compare this approach with an exact full-Bayesian analysis. To make the comparison computationally feasible, methods were applied to fine mapping of a QTL within a simulated chromosomal region. Realistic low marker-QTL LD and high LD was simulated. Estimates for location and effect of the QTL as well as LD parameters obtained by both methods were very similar. Only if LD could not be estimated sufficiently well to distinguish between putative QTL locations, posterior distributions of the QTL location parameter slightly differed between methods. Although the exact approach always performed better in estimating QTL genotypes, accuracy of GEBVs was the same for the two methods at realistic low LD. At high LD, however, the exact approach resulted in a slightly higher accuracy of GEBVs. Pedigree blocking used to sample SNP variables neither had an effect on estimates of QTL parameters nor on accuracy of GEBVs, whereas a smaller locus block size reduced the accuracy of GEBVs.

# A comparison of genomic BLUP and BayesB at various population and trait genetic architectures

Daetwyler, H.D.<sup>1,2\*</sup>, Pong-Wong, R.<sup>1</sup>, Villanueva, B.<sup>3</sup>, and J.A. Woolliams<sup>1</sup>

<sup>1</sup>The Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, UK, EH25 9PS;

<sup>2</sup>Wageningen Animal Breeding and Genomics Centre, Wageningen University, 6700 AH Wageningen, NL; <sup>3</sup>Scottish Agriculture College, West Mains Rd., Edinburgh, UK, EH9 3JG

\* presenting author, hans.daetwyler@roslin.ed.ac.uk

Genome-wide evaluation combines traditional breeding methods with genomic data to predict breeding values. In this study we compared the accuracy (correlation of true and predicted breeding values) of genomic best linear unbiased prediction (GBLUP) and Bayesian (BayesB) genome-wide evaluation at varying numbers of loci with effect ( $N_{qtl}$ ), ranging from 50 to 2800. In GBLUP a realised relationship matrix was calculated based on all loci. In BayesB, exact (i.e.,  $= N_{qtl}$ ) and low priors (i.e.,  $< N_{qtl}$ ) were investigated. Mutation drift balance was simulated in populations of effective population size ( $N_e$ ) 500 and 1000. The genome was 10M long with up to 3000 bi-allelic loci. A training population of equal size to  $N_e$  was generated to estimate loci effects and breeding values. GBLUP had a constant accuracy at each heritability ( $h^2$ ) regardless of  $N_{qtl}$ . BayesB had higher accuracy than GBLUP when  $N_{qtl}$  was low. However, this advantage diminished as  $N_{qtl}$  increased and the two methods' accuracy converged at 250  $N_{qtl}$  ( $h^2$  0.5,  $N_e$  500), 450  $N_{qtl}$  ( $h^2$  0.2,  $N_e$  500) and 750  $N_{qtl}$  ( $h^2$  0.5,  $N_e$  1000). When  $N_{qtl}$  became large GBLUP outperformed BayesB. These trends were similar for both  $N_e$ . Low priors increased the accuracy of BayesB in the intermediate  $N_{qtl}$  range because of the estimation of fewer effects. Low priors did not outperform exact priors at high  $N_{qtl}$  as less of the genetic variance was captured. Our results illustrate that the performance of genome-wide evaluation methods depends significantly on population and trait genetic architecture.

# Use of the Elastic-Net algorithm for genomic selection in dairy cattle

Croiseau P<sup>1\*</sup>, Ducrocq V<sup>1</sup>

<sup>1</sup>INRA – UMR 1313Génétique Animale et Biologie Intégrative, 78352 Jouy-en-Josas, France

\* Croiseau Pascal: pascal.croiseau@jouy.inra.fr

## **Background:**

The availability of the 54K SNP array in dairy cattle makes possible to envision the use of genomic prediction instead of classical genetic evaluations in selection programmes. However, this requires suitable statistical approaches as the number of observations is often much lower than the number of predictor variables.

In animal breeding, it is clear that for a given trait, not all SNP across the whole genome are close to QTLs for the trait, whatever their size. In this context, we applied the Elastic-Net method to simultaneously select among all SNP the ones which are related with the studied trait and estimate their effect. The Elastic-Net approach consists of a combination of Ridge Regression and of a LASSO algorithm. A parameter  $\lambda$  varying between 0 and 1 indicates the relative weight given to Ridge Regression in the Elastic-Net procedure (pure Ridge Regression algorithm if  $\lambda=1$ , pure LASSO algorithm if  $\lambda=0$ ).

## **Results:**

We present here an application of the Elastic-Net approach on a set of 694 bulls in Montbéliarde breed using data available in 2004. The phenotypes of these animals are Daughter Yield Deviations (DYD) calculated for the French Marker Assisted Selection program for milk, fat and protein yields and contents. To evaluate the performance of the Elastic-Net algorithm, we use a validation set of 227 bulls for the Montbéliarde breed with phenotypes (DYD) available in 2008. The Elastic-Net approach directly applied to the whole data set was very time consuming (more than 20 days for the Montbéliarde breed). We proposed an alternative two-step approach which allows performing the analysis in less than 1 day. In this two-step approach, relevant SNPs were first selected by looking at moving intervals of 500 SNP and then combined into a single analysis.

## **Conclusion:**

Using the Elastic-Net approach with the whole data set is probably the best solution to obtain the optimal parameters. But, if we want to take into account the time consuming problem, the optimal Elastic-Net parameters are found using our two-step approach. A comparison with other methods is under way.

# **Detecting the Association of Single Nucleotide Polymorphisms and Prediction of Future Yields for a Growth Trait Using a Mixed Model with Orthogonal Polynomials**

Tomasz Suchocki<sup>1\*</sup>, Joanna Szyda<sup>1,2</sup>

<sup>1</sup> Department of Animal Genetics, Wrocław University of Life Sciences, Wrocław, Poland

<sup>2</sup> Institute of Natural Sciences, Wrocław University of Life Sciences, Wrocław, Poland

\* presenting author, tomasz.suchocki@up.wroc.pl

The main aim of this study is the detection of association between SNPs and a quantitative growth trait in the situation where a trait value for each individual is measured at five different time points. In order to account for the correlation between the consecutive measurements a mixed model, which describes each effect using the 4<sup>th</sup> order Legendre orthogonal polynomials, is fitted. In this model, the SNPs are modeled as fixed effects while the environment as a random effect. The maximum likelihood estimates of all model parameters are obtained using the EM algorithm and the assessment of the significance of the additive effects of each SNP is based on the Likelihood Ratio Test with P values corrected for multiple testing. For each significant SNP, we calculate the percent of the total variance contributed by that SNP. Moreover, by using a fixed model with orthogonal polynomials, which simultaneously incorporates effects of all the SNPs we are able to predict future trait values.

# Repeatability model and kriging as methods for the estimation of growth curves

Żukowski Kacper<sup>1\*</sup>, Macierzyńska Anna<sup>1</sup>, Wierzbicki Heliodor<sup>1</sup>, Szyda Joanna<sup>1,2</sup>

<sup>1</sup> Institute of Animal Genetics, Wrocław University of Environmental and Life Sciences

<sup>2</sup> Institute of Natural Sciences, Wrocław University of Environmental and Life Sciences

\* kacper.zukowski@up.wroc.pl

The aim of the study was to predict breeding values of all 1000 non-phenotyped individuals on time600.

The data were modeled using the repeatability model. The repeatability model can be employed for the analysis of the data when multiple measurements on the same trait are recorded on an individual, such as litter size in successive pregnancies, milk yield in successive lactations or growth trait like in QTLMAS2009 data set [3]. With repeated measurements it is assumed that there is an additional covariance between records of an individual due to non-genetic permanent environmental effects. Thus, the between-individual variance is partly genetic and partly environmental. The repeatability model is usually of the form:

$$y = Xb + Za + Wpe + e,$$

where  $y$  = vector of observations,  $b$  = vector of fixed effects (generation1, generation2, time0, time132, time256, time397 and time530),  $a$  = vector of random animal effects,  $a \sim N(0, A \otimes G)$ ,  $pe$  = vector of permanent environmental effect and non-additive genetic effects,  $pe \sim N(0, I \otimes P)$ ,  $e$  = vector of random residual effects,  $e \sim N(0, I\sigma_e^2)$  and  $X$ ,  $Z$ ,  $W$  are incidence matrices of fixed, random animal and permanent environmental effects, respectively [1,4].

Additionally, a modification of the above model with the additive relationship matrix estimated from the pedigree information replaced by the kinship matrix estimated from genetic information (for 108 SNPs with  $r^2 > 0.5$ ) was used.

Kriging which is based on a minimum-mean-squared-error method of spatial prediction was applied to predict breeding values at time point 600 [2]. The applied model assumes:

$$Z(s) = \mu(s) + \varepsilon(s)$$

where,  $Z(s)$  the weight at time  $s$  averaged over all individuals,  $\mu(s)$  is trend of trait and  $\varepsilon(s)$  is zero-mean intrinsically stationary random process with variogram  $2\gamma(s)$ .

<sup>[1]</sup> Arango J.A., Cundiff L.V., Van Vleck L.D. (2004) Covariance functions and random regression models for cow weight in beef cattle; *Journal of Animal Science*, 82:54-67

<sup>[2]</sup> Cressie N. (1993) *Statistics for Spatial Data*, revised edition; Wiley, New York

<sup>[3]</sup> Interbull (2000) *Proceedings of the 2000 Interbull Meeting, May 14-15, BLED, SLOVENIA*

<sup>[4]</sup> Mrode R.A. (2005) *Linear Models for the Prediction of Animal Breeding Values*; 2<sup>nd</sup> Edition, CABI Publishing

## Quantitative genetics in Arabidopsis: from QTL to QTN

Maarten Koornneef<sup>1,3</sup>, Joost Keurentjes<sup>1,2</sup> and Matthieu Reymond<sup>3</sup>

<sup>1</sup>Laboratory of Genetics and <sup>2</sup>Laboratory of Plant Physiology, Wageningen University, Netherlands. <sup>3</sup>Max Planck Institute for Plant Breeding Research, Cologne, Germany.

A large natural genetic variation is present among Arabidopsis accessions and which is assumed to reflect part of the adaptation of the different accessions to their natural growth environment. The variety of populations developed and their use for mapping quantitative trait loci (QTL) will be discussed. In addition, the feasibility to reveal the gene responsible for the effect of a QTL by a combination of map based-cloning coupled with mutant approaches has now been demonstrated for several traits. For this purpose, the Arabidopsis genome resources represent powerful advantages. The full genome sequence and the resequencing of accessions allows the creation of high density polymorphism maps required and an efficient identification of candidate genes (QTGs) and even candidates nucleotide polymorphisms (QTN) . Up to now, cloned QTLs include those encoding genes for physiological traits such as flowering time, seed dormancy, frost tolerance etc. However, QTL analysis has now also been extended to molecular traits such as metabolites (mQTL), enzyme activities, proteins (pQTL) and gene expression (eQTL). Combining morphological or phenological traits with “omic” traits allows the construction of molecular genetic networks based on the co-regulation by the same genetic factors at different levels (transcription, transduction) and may assist the identification of the gene responsible for the effect of the QTL. The use of this approach will be demonstrated on the elucidation of the genetic regulation of glucosinolates, flowering time and primary metabolism. This integrative strategy allows the development of system biology approaches.

# Genes for seed quality: New tools and approaches

Ronny Joosen

Wageningen University, Plant Physiology

Ronny.Joosen@wur.nl

Seed performance is a very complex trait which comprises a large number of physiological principles related to important plant developmental processes. We use a physiological genetical genomics approach to survey these processes and resolve the underlying molecular mechanisms.

We are using the *Arabidopsis* Bay-0xSha RIL population and are in the process of locating phenotypic QTL's for diverse seed quality traits, like the germination performance under various environmental stresses. To be able to generate cumulative germination curves in a high throughput manner, we have developed an automated germination scoring system, in which we combined sophisticated image analysis with newly developed curve fitting software.

In parallel, we perform a comprehensive eQTL study on different developmental seed stadia using a generalized setup (Li et al., 2008). For this eQTL study we use the SNPtile array, a new Affymetrix tiling array which also harbors 250K SNP's. The combination of this transcriptomic and SNP data with the genetical power of a comprehensively phenotyped RIL population will open a new and exciting area in the field of physiological genetical genomics.

*Li, Y., Breitling, R. and Jansen, R.C. (2008) Generalizing genetical genomics: getting added value from environmental perturbation. Trends in Genetics. 24, 518-524.*

# Footprints of Recent Positive Selection in Holstein Cattle

S. Qanbari<sup>1\*</sup>, E. C. G. Pimentel<sup>1</sup>, J. Tetens<sup>2</sup>, G. Thaller<sup>2</sup>, A.R. Sharifi<sup>1</sup> and H. Simianer<sup>1</sup>

<sup>1</sup> Animal Breeding and Genetics Group, Department of Animal Sciences, George-August University, 37075 Göttingen, Germany

<sup>2</sup> Institute of Animal Breeding and Animal Husbandry, Christian-Albrechts-University, 24098 Kiel, Germany

\* Presenting author: [sqanbar@gwdg.de](mailto:sqanbar@gwdg.de)

The detection of signatures of selective breeding across the genome could facilitate identifying genes and regulatory elements involved in beneficial traits. To this purpose, we used the newly available 50K SNP chip data for tagging the genome wide footprints of positive selection in Holstein-Friesian cattle. We employed the recently described Extended Haplotype Homozygosity test, which detects selection by measuring the characteristics of haplotypes within a single population. To formally assess the significance of these results, we compared the combination of frequency and Relative Extended Haplotype Homozygosity value of each core haplotype with equally frequent haplotypes across the genome. A subset of the putative regions identified by high significance in the genome wide EHH tests was mapped. We annotated genes to identify possible influence they have in beneficial traits using the Gene Ontology (GO) database. A panel of genes, including FABP3, CLPN3, SPERT, HTR2A5, ABCE1, BMP4 and PTGER2 was detected which overlapped with the top P-values. This panel seems to comprise the most interesting candidates as they represent a broad range of economically important traits such as milk yield and composition as well as reproductive and behavioral traits. We also reported high values of LD and a slower decay of haplotype homozygosity for some candidate regions harboring major genes related to dairy quality. The results of this study provide a genome wide map of selection footprints in Holstein genome and could be used to identify genes of economic interest in dairy cattle.

# An improved method for calculating line origin probabilities for crosses between outbred lines

Lucy Crooks<sup>1\*</sup>, Carl Nettelblad<sup>2</sup>, Sverker Holmgren<sup>2</sup>, Örjan Carlborg<sup>1</sup>

<sup>1</sup>Department of Animal Breeding and Genetics, SLU, Sweden

<sup>2</sup>Department of Information Technology, Uppsala University, Sweden

\*presenting author, Lucy.Crooks@hgen.slu.se

Analyses of experimental crosses between divergent outbred lines have been successful in mapping QTL for a range of traits in a number of organisms. Central to such analyses is determining the probability of a chromosomal section originating from each line, given the marker genotypes. With outbred lines, the same marker alleles may be segregating in both lines, so that the allele does not directly identify the line origin (partially informative marker). This is especially likely with SNP markers because they are usually only biallelic. To calculate the line origin probabilities for a given genomic location, one should consider all combinations of possible line origins for partially informative markers between flanking fully informative markers. The original and still most widely used approach is to separately evaluate each combination and take the sum (Haley *et al.*, 1994). This requires a rapidly increasing number of calculations with the number of markers. To prevent this number becoming intractable, in the web-based tool GridQTL (Seaton *et al.*, 2006), no more than 15 partially informative markers are used for each location. Here, we present an improved method that is computationally more efficient and therefore allows information from many more markers to be included. The method is based on a Hidden Markov Model structure. We compare our method to the procedure in GridQTL and discuss the implications in relation to the increasing availability of SNP chips containing tens of thousands of markers.

Haley CS, Knott SA, Elsen JM. *Genetics* 136:1195-1207 (1994).

Seaton G, Hernandez J *et al. Proc. 8th WCGALP, Brazil* (2006).

# Extraction of QTL-informative SNPs using genome sequencing data from founder line DNA pools in chicken

Stefan Marklund<sup>1\*</sup>, Örjan Carlborg<sup>1</sup>

<sup>1</sup>Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, SE-750 07 Uppsala, Sweden

\* Stefan.Marklund@hgen.slu.se

Several quantitative trait loci (QTL) regulating growth were previously detected using an intercross between two chicken lines long-term (> 45 generations) selected for high and low body weight (at 56 days), respectively. We used the Applied Biosystems Solid method for genome resequencing with 4X coverage on line-specific pools of DNA from 11 founder individuals. More than 100 000 single nucleotide polymorphisms (SNPs) were detected within the QTL regions. To extract the SNPs with most variation between lines we only considered positions where one line differed from the reference genome sequence in all reads (score = 0), whereas no difference was listed for the other line. Other preferences included high sequencing coverage and low variation within lines around the SNP. Fixation could hypothetically be indicated by a high ratio between the number of SNPs between lines (with score = 0) and the number of SNPs showing polymorphism within lines (score > 0). Therefore, we computed a modified measurement of this ratio in 2 kb and 10 kb intervals across the QTL regions and tested its correlation with allele frequency difference between lines ( $f_D$ ) using 43 SNPs detected with sequencing score = 0 and previously used for genotyping of 15 founder birds from each line. The Spearman's rank-order correlations were 0.54 and 0.60 using 2 kb- and 10 kb-intervals, respectively. The described criteria and correlations guided our selection of SNPs for QTL fine mapping, which resulted in 37 SNPs with significantly higher ( $P = 0.001$ ) average  $f_D$  than the 43 SNPs used for the correlation tests.

# **A two-step approach combining the Gompertz growth model with genomic selection for longitudinal data.**

Pong-Wong R\*, Hadjipavlou G

The Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, Midlothian, EH25 9PS  
UK

\* presenting author: [Ricardo.pong-wong@roslin.ed.ac.uk](mailto:Ricardo.pong-wong@roslin.ed.ac.uk)

## **Background:**

We used the Gompertz growth curve to model a simulated longitudinal dataset provided by the QTLMAS2009 workshop and applied genomic evaluation to the derived model parameters and to a model-predicted trait value.

## **Results:**

The prediction of phenotypic performance assuming a non-linear function allowed us to obtain genomic breeding value estimates for time where not phenotypic records were available. Genomic breeding values calculated from predicted phenotypes were highly correlated with the breeding values obtained by directly using the observed phenotype. This suggests that the Gompertz curve provided a good fit of the data.

The results from the genomic selection provided strong evidences about the genetic architecture of the traits. When examining the parameters of the Gompertz curve, a high genetic correlation between the asymptotic final value and the maximum growth was found, where the same set of SNPs appear to be having a large effect in both traits.

In this study we estimated of the proportion of SNP affecting a given trait, contrasting with previously reported implementations of genomic selection, where this parameter is assumed to be known with error.

## **Conclusions:**

The results from this study showed that the two-step approach used here to combine curve fitting and genomic selection on longitudinal data provided a very simple way for combining these two complex tasks without any detrimental effect in the breeding value estimation.

# Sensitivity of Genomic Selection to using different prior distributions

Klara L Verbyla<sup>1,2,3,4\*</sup>, Phil Bowman<sup>2</sup>, Ben Hayes<sup>2</sup> and Mike Goddard<sup>2,3,4</sup>

<sup>1</sup> Animal Breeding and Genomics Centre, ASG Wageningen UR, PO Box 65, 8200 AB Lelystad, The Netherlands

<sup>2</sup> Biosciences Research Division, Department of Primary Industries Victoria, 1 Park Drive, Bundoora 3083, Australia

<sup>3</sup> Melbourne School of Land and Environment, The University of Melbourne, Parkville 3010, Australia

<sup>4</sup> The Cooperative Research Centre for Beef Genetic Technologies, University of New England, Armidale, NSW 2351, Australia

\*Corresponding author

Genomic selection describes a selection strategy based on genomic breeding values (GEBV) predicted from dense single nucleotide polymorphism (SNP) data. Different Bayesian models have been suggested, with the main difference centred around the specification of the prior distributions. The simulated dataset of the 13<sup>th</sup> QTL-MAS workshop was analysed using four Bayesian approaches to predict GEBV for animals without phenotypic information. Different prior distributions were assumed to assess their affect on the accuracy of the predicted GEBV. A BLUP approach, assuming equal variances across SNP, produced significantly worse results than the other methods. Models that assumed unequal variances for the SNP produced comparable results even when the prior distribution doesn't match the true distribution of QTL effects. However, the recommendation would be to include any prior knowledge available about a trait's QTL effect distribution when choosing prior distributions to increase the accuracy of the predicted GEBV.

# The prediction of genomic breeding values for longitude traits using random regression

Zhe Zhang<sup>1</sup>, Jianfeng Liu<sup>1</sup>, Qin Zhang<sup>1</sup>

<sup>1</sup>State Key Laboratory for Agrobiotechnology, Key Laboratory of Animal Genetics and Breeding of the Ministry of Agriculture of China, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

## Background

Genomic selection has become a popular tool for genetic improvement in livestock due to its distinct advantages over traditional selection based solely on pedigree. In the framework of genomic selection, a huge number of SNPs or marker haplotypes are utilized to predict the so-called genomic breeding values (gEBV). The objective of the present study is to verify the BLUP method for predicting genomic breeding values using the public simulated data from the website <http://www.qtlmas2009.wur.nl/UK/>.

## Method

To obtain the gEBVs of the 1,000 unphenotyped individuals at the time point 600, three different methods were employed including (1) gEBV were predicted based on phenotypic values of time point 600, which were estimated via the known phenotype values at time points 0 through 530; (2) predicting the SNP effects on time point 600 via the SNP effects of the previous 5 time points, which were separately predicted based on corresponding phenotypic observations; (3) predicting the gEBV on time point 600 using the predicted gEBV of the previous 5 time points. All predictions were achieved by fitting a 3 order Legendre polynomial in prediction equations. Variance components were estimated by REML. A BLUP-based algorithm was developed to calculate the effects of the 453 SNPs at different time points, with the assumption that all SNP effects have the equal variance.

In addition to the gEBVs, the breeding values of the 1,000 individuals with phenotypes at the first five time points and their parents were also predicted for time point 600 with traditional animal model BLUP, using predicted phenotypes as mentioned above.

## Results

The heritabilities were estimated to be 0.5 for the simulated trait at all the time points. The gEBVs of all 1000 phenotyped animal have no significance difference between the three methods. The correlations between the gEBVs and the traditional EBVs are 0.8628, 0.8908 and 0.8923 for the 1,000 phenotyped animals, their sires, and their dams, respectively.

# Genomic breeding value prediction using three Bayesian methods and application to reduced density marker panels

Matthew A Cleveland<sup>1\*</sup>, Selma Forni<sup>1</sup>, Nader Deeb<sup>1</sup>, Christian Maltecca<sup>2</sup>

<sup>1</sup>Genus plc., 100 Bluegrass Commons Blvd., Suite 2200, Hendersonville, TN, 37075, USA

<sup>2</sup>North Carolina State University, Department of Animal Science, Raleigh, NC, 27695-7627, USA

\* matthew.cleveland@pic.com

## Background

Bayesian approaches for predicting genomic breeding values (GEBV) have been proposed that allow for different variances for individual markers resulting in a shrinkage procedure that uses prior information to coerce negligible effects towards zero. These approaches have generally assumed application to high-density genotype data on all individuals. In this study, three approaches were compared for predicting GEBV when training at high marker density and predicting at high or low densities: the well-known *Bayes-A*, a generalization of *Bayes-A* where scale and degrees of freedom are estimated from the data (*Student-t*) and a Bayesian implementation of the *Lasso* method. Scenarios were evaluated for predicting GEBV using low-density marker subsets, including selection of SNP based on genome spacing or effect size and the inclusion of unknown genotype information as genotype probabilities.

## Results

The GEBV accuracy (calculated as correlation between GEBV and traditional breeding values) was highest for *Lasso*, followed by *Student-t* and then *Bayes-A*, though differences were small. In general the shrinkage applied by the *Lasso* approach was less conservative than *Bayes-A* or *Student-t*, indicating that *Lasso* may be more sensitive to QTL with small effects. In the reduced-density marker subsets the accuracy when using *Lasso* was reduced more than the other methods, relative to the high-density set, but the ranking of the methods was the same. Overall, low-density, evenly-spaced SNPs did a poor job of predicting GEBV, but SNPs selected based on additive effect size yielded accuracies similar to those at high density, even when coverage was low. The inclusion of genotype probabilities to the evenly-spaced subsets showed promising increases in accuracy that need more evaluation when true breeding values are available.

## Conclusions

In this dataset the *Lasso* approach slightly outperformed the other methods when predicting GEBV at both high and low density and may have particular advantages in situations where large QTL are not expected. When markers were selected at low density based on genome spacing, the inclusion of genotype probabilities increased GEBV accuracy which would allow a single low-density marker panel to be used across traits.

# Comparison of methods for estimation of genetic covariance matrix from SNP or pedigree data utilised to predict breeding value

Sebastian Mucha<sup>1§</sup>, Anna Wolc<sup>1</sup>, Tomasz Strabel<sup>1</sup>

<sup>1</sup>Department of Genetics and Animal Breeding, Poznan University of Life Sciences, Wolynska 33, 60-637 Poznan, Poland

## Background

The analysis was based on a dataset prepared for the 13<sup>th</sup> QTL-MAS Workshop in Wageningen (The Netherlands), which contained records of a trait observed in five time points on 1000 individuals from a simulated population of 2025 animals. The aim of this paper was to predict breeding value of the 1000 non-phenotyped animals in the 6<sup>th</sup> time point, using three different strategies based on similarity between individuals due to common ancestry (pedigree records), and two methods based on marker similarity

## Methods

Genetic co-variance matrix between all animals present in the dataset was estimated with three different methods: one using pedigree information only and two based on SNP markers. It was utilized in ASREML with single-trait animal model to estimate breeding value of animals separately in each of the five time points. Quadratic regression on breeding values in first five points was used to predict the unknown breeding values in the 6<sup>th</sup> time point of the 1000 non-phenotyped individuals.

## Results

Genetic and residual variances increased with time regardless of the method used to calculate genetic co-variances. Breeding values estimated with all three methods were increasing with increasing genetic variance. Correlations between breeding values estimated with the three methods were between 0.82 and 0.89 and were decreasing with time. The ranking of animals with respect to their breeding value differed between the methods. However, there was a 45% overlap in the individuals that were listed in the top 20 animals according to each of the three methods.

## Conclusions

Due to lack of validation methods it is difficult to compare results of the three methods and point out which one provides estimates of breeding values that are closest to the true values. Therefore comparison of results based on simulation studies can be of great value.

# Prediction of Genomic Breeding Values for a longitudinal trait by using Principal Component Analysis

Nicolò P.P. Macciotta<sup>1\*</sup>, Giustino Gaspa<sup>1</sup>, Roberto Steri<sup>1</sup>, Camillo Pieramati<sup>2</sup>, Maria Annunziata Pintus<sup>1</sup>, Corrado Dimauro<sup>1</sup>

<sup>1</sup>Dipartimento di Scienze Zootecniche, Università di Sassari, Italia

<sup>2</sup>Centro di Studio del Cavallo Sportivo, Università di Perugia, Italia

\* Presenting author, e-mail: macciott@uniss.it

## Background

Several traits of economic interest are expressed as repeated measurements along a trajectory of time. This is a further issue to be addressed in genomic selection, beside the problem of the large number of marker effects that have to be estimated. The five phenotypes available from the common data set of the QTL-MAS workshop were treated as repeated measures of the same trait. Polygenic EBVs were estimated for all the 2025 individuals and considered as “golden standard”. Two data sets were created: TRAINING, made by 1000 individuals with available phenotypes; PREDICTION, with the remaining 1025 individuals. Marker genotype effects were estimated on TRAINING by a mixed linear model, assuming an equal contribution of each marker (as  $\sigma_a^2/n$ ) and no interaction between marker loci. As an alternative,  $k$  principal component were extracted from the SNP data matrix and used in the estimation step instead of SNP genotypes. Estimated effects were then used to predict GEBVs in the PREDICTION individuals. Accuracy of GEBVs was evaluated as correlation with polygenic EBVs. Moreover, principal component analysis was also carried out on five phenotypes treated as different traits, to derive new variables able to represent the longitudinal trait under study. Also for these new variables, GEBVs were calculated using the above described approach

## Results

Seventy-six principal components were retained for GEBV estimation. Correlations between polygenic EBVs and GEBVs in the TRAINING data were 0.78 and 0.77 when SNP marker or Principal components were used as predictors, respectively. Values decrease to 0.52 and 0.50 in the PREDICTION data. The Principal component analysis carried out on the five phenotypes yielded a variable related to the “average level” and one related to the shape of the longitudinal trait under study. Correlations between polygenic EBVs and GEBVs for both the two principal component were of the same order of those estimated in the repeated measure approach, both in the TRAINING and PREDICTION individuals.

## Conclusions

Principal component analysis was able to reduce (<20%) the number of predictors in the calculation of GEBVs, keeping the same accuracy of estimation either when repeated records or new synthetic variables were modelled.

# Genome-wide selection by mixed model ridge regression and extensions based on geostatistical models

Torben Schulz-Streeck<sup>1</sup>, Hans-Peter Piepho<sup>1§</sup>

<sup>1</sup>Bioinformatics Unit, Institute for Crop Production and Grassland Research,  
Universität Hohenheim, Fruwirthstrasse 23, 70599 Stuttgart, Germany

<sup>§</sup>Corresponding author

Email addresses:

HPP: piepho@uni-hohenheim.de

TSS: torben.schulz-streeck@uni-hohenheim.de

## Background

The success of genome-wide selection (GS) approaches will depend crucially on the availability of efficient and easy-to-use computational tools. Therefore, approaches that can be implemented using mixed models hold particular promise and deserve detailed study. A particular class of mixed models suitable for GS is given by geostatistical mixed models, when genetic distance is treated analogously to spatial distance in geostatistics.

## Methods

We consider various spatial mixed models for use in GS. The analyses presented for the QTL-MAS 2009 dataset pay particular attention to the modelling of residual errors as well as of polygenetic effects.

## Results

It is shown that geostatistical models are viable alternatives to ridge regression, one of the common approaches to GS. In the example considered, we did not find a large effect of the residual error variance modelling, largely because error variances were very small. A variance components model reflecting the pedigree of the crosses did not provide an improved fit.

## Conclusions

We conclude that geostatistical models deserve further study as a tool to GS that is easily implemented in a mixed model package.

# Joint association and linkage QTL mapping in half-sib families by regression

Legarra A.<sup>1\*</sup>, Fernando R.L.<sup>2</sup>

<sup>1</sup>Institut National de la Recherche Agronomique, UR631 SAGA, BP 52627, 31326 Castanet Tolosan, France

<sup>2</sup>Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, IA, USA

\* presenting author, andres.legarra@toulouse.inra.fr

*Methods.* We propose to combine association mapping by regression and Haley-Knott half-sib regression for linkage (Knott *et al.*, 1996 Theor. Appl. Genet. 93:71) to map QTLs combining linkage disequilibrium (LD) and linkage analysis (LA).

In association mapping, at a given location, the average effect of each QTL of each sire depends on the allele state of the closest marker (haplotype), and is modeled by regression. In Haley-Knott linkage regression, the effect in the offspring of the paternally inherited QTL is modeled as a regression on the probability  $p$  of transmission of each of the paternal QTLs. This probability is computed using all available markers. The maternal QTL in the offspring can be modeled by regression on marker (haplotype) allele/state. Thus, for son  $j$  of sire  $i$ :

$$y_{ij} = pb_{i1} + (1-p)b_{i2} + pa_{i1} + (1-p)a_{i2} + b_{ij} + e_{ij}$$

Where  $b_{ik}$  are the effects of the marker observed at the  $k$  chromosome at the  $i$  sire,  $b_{ij}$  is the effect of the marker observed at the maternal chromosome of  $ij$ , and  $a_{ik}$  are the QTL effects in the sire not accounted for by the observed marker (i.e., if LD QTL – marker is incomplete). We term this model “LDLA”. Assuming LD is complete (all  $a_{ik} = 0$ ) leads to a more parsimonious model called “LDdecay”.

*Results and conclusions.* We tested this approach against Meuwissen *et al.* (2002 Genetics 161:373) method (IBD) by simulation in several scenarios. Performance in term of mean squared error is generally similar; however, IBD tends to be biased towards the center of the haplotype because it considers all available markers for LD. Computations in the regression approach are extremely simple. Extensions of the model to genomic selection and multi-QTL mapping are straightforward. The model is thus a simple and satisfactory approach for QTL mapping using LD and LA.

# Monte Carlo Full Likelihood Approach in Variance Component QTL Analysis

Xia Shen<sup>1,2\*</sup>, Lars Rönnegård<sup>2,3</sup>, Örjan Carlborg<sup>3</sup>

<sup>1</sup> Linnaeus Centre for Bioinformatics, Uppsala University, Uppsala, Sweden.

<sup>2</sup> Statistics Unit, Dalarna University, Borlänge, Sweden

<sup>3</sup> Department of Animal Breeding & Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.

\* xia.shen@lcb.uu.se

The identity-by-descent (IBD) matrix is the core of the variance component QTL model. The true IBD matrix comes from a distribution of IBD matrices given the marker information, but its expectation is usually used in the likelihood. This gives an incorrect likelihood value since the extra uncertainty in estimating the IBD matrix is not included. We therefore developed a Monte Carlo method for calculating the true likelihood (so-called full likelihood, Xu, S.; *Genetics* 1996; 144(4): 1951-60) incorporating the uncertainty of the estimated IBD matrix. The aim of the study was to compare the true likelihood with the likelihood based on the expected IBD matrix, and to study their computational requirements for large pedigrees with a small founder generation. Our simulation results show that the likelihood based on the expected IBD matrix approximates the true likelihood accurately. Our Monte Carlo full likelihood method can actually be computationally more efficient than the expectation method for large pedigrees with a small founder generation, because the rank of the true IBD matrix is much lower than the rank of the expected IBD matrix. Using the IBD matrices produced in our Monte Carlo Full Likelihood (MCFL) method we may also simplify the modeling of epistasis for linked QTL following Mao & Xu (*Heredity* 2005; 94(3):305-15).

# On the estimation of genetic effects for multiallelic systems

Álvarez-Castro JM<sup>1\*</sup>, Yang, R-C<sup>2</sup>, Nettelblad C<sup>3</sup>, Carlborg Ö<sup>4</sup>

<sup>1</sup>Department of Genetics, University of Santiago de Compostela, Lugo, Spain

<sup>2</sup>Department of Agricultural, Food and Nutritional Science, University of Alberta, Alberta, Canada

<sup>3</sup>Department of Information Technology, Uppsala University, Uppsala, Sweden

<sup>4</sup>Animal Breeding and Genetics Department, Swedish University of Agricultural Sciences, Uppsala, Sweden

\* jose.alvarez.castro@usc.es

**Background:** The traditional models of genetic effects have been extensively revised since the advent of QTL analysis. On one hand, statistical models of genetic effects, formulated upon average effects over populations, were improved to better handling model selection strategies and the decomposition of the genetic variance. On the other hand, functional models were built to formulate genetic effects regardless of the gene frequencies of any population. By providing a link between the two formulations, the NOIA model enables the study of valuable evolutionary properties of a trait that has been subject to QTL analysis.

**Results:** Here we present implementations of the functional and the statistical formulations of genetic effects to multiple alleles in the framework of the NOIA model. These implementations conclude the functional formulation, which gets to be completely general. We also show that this formulation fits to a common statistical testing procedure. Concerning the statistical formulations, our implementations improve previous models and point to further progress to be made in this direction. In particular, we provide expressions to transform between functional genetic effects and statistical, average effects under Hardy-Weinberg proportions. Further, we inspect the impact of deviations from the Hardy-Weinberg proportions in the differences between multiallelic average effects and average excesses.

**Conclusions:** We stress the need of further improving the models of genetic effects, which as kernel pieces of the QTL machinery can be implemented regardless of the statistical methods used for performing the estimates at the shell of that machinery. For instance, they can be implemented for using them in a Haley-Knott regression framework, although an imputations framework could be preferable.

# Identity by descent probability matrix decomposition by a latent ancestor allele model and its application in QTL analysis.

Cajo J.F. ter Braak<sup>1\*</sup>, Martin Boer<sup>1</sup>, Radu Totir<sup>2</sup>, Chris Winkler<sup>2</sup>, Howie Smith<sup>2</sup>, Marco Bink<sup>1</sup>

<sup>1</sup>Biometris, Wageningen University and Research Centre, Wageningen, the Netherlands, 6700 AC

<sup>2</sup>Pioneer Hi-Bred International, A Dupont Company, Johnston, Iowa 50131, USA

\* presenting author [cajo.terbraak@wur.nl](mailto:cajo.terbraak@wur.nl)

Elements of IBD probability matrices measure the probability that two individuals share the same allele of a common ancestor. The ancestral alleles and their number remain implicit. For human inspection and QTL analysis, an explicit representation in terms of ancestral allele origin and number of alleles may be desirable. To this purpose, we decompose the IBD matrix by a latent class model with  $K$  classes (latent ancestor alleles). We provide an efficient algorithm to fit the model. The algorithm correctly reconstructed the ancestry of 16 maize inbreds from their IBD matrix only. We also show that the model can help QTL detection from connected crosses.

# Factors influencing the optimum multiple marker model for analysis of association data

J. E. Powell<sup>1\*</sup>, A. Kranis<sup>2</sup>, C. S. Haley<sup>1</sup>

<sup>1</sup> Department of Genetics and Genomics, The Roslin Institute, University of Edinburgh, EH25 9PS

<sup>2</sup> Aviagen, Newbridge, Edinburgh, EH28 8SZ

\* J. E. Powell; josephpowell@roslin.ed.ac.uk

## Background

There is currently much interest in the comparison of single-locus, multi-locus and haplotype model performance for their ability to explain genetic variance, and map QTL in humans and livestock. Relative performance of these models is expected to be affected by range of localized genetic architecture, such as LD patterns and allele frequencies at the causal variants. A growing number of studies demonstrate that haplotype-based approaches may provide more power and accuracy in locating QTL and disease variants than single locus methods. However, a limitation of these comparisons is that they focus on extremes, using either the maximum of single-locus statistics or a global test for all haplotypes. Alternative approaches include the use of multiple markers, but only fitting low order interactions, such as a main order effects multiple-regression analysis. To date, there has been little focus on determining how models perform across a range of genetic architecture conditions.

## Methods

With the aim of determining the relative performance of four association models, we used the extent of LD in a broiler chicken data set, consisting of 13,000 genome-wide SNPs genotyped in 200 individuals, to assess their ability to explain genetic variance and provide statistical support for QTL with a range of minor allele frequencies.

The data were used to simulate LD between markers and a QTL by randomly selecting one SNP to act as a surrogate QTL (sQTL), and then using surrounding markers to test for association using four regression-based models. In total 6300 sQTL were tested against with the following models: single-locus, three marker main order effects and two haplotype models each with three markers. sQTL were divided into ten minor allele frequency (MAF) bins, with equal range, from 0 to 0.5, to determine the relative performance of models mapping for QTL with different MAF.

## Conclusions

When mapping against sQTL with MAF below 0.225 the haplotype model that accounted for uncertainty in phase performed better than all other models. Performance, relative to other models, increased as sQTL MAF dropped. There was little difference in performance of models for sQTL with intermediate MAF. Under the assumption of neutral mutation or stabilizing selection models, QTL allele frequencies are expected to resemble a U-shaped distribution, producing a high proportion of QTL with low MAF. Therefore, unless prior information of QTL frequencies is available, a haplotype-based approach, accounting for phase uncertainty, is preferable to low order interaction, or single locus models. Many genome-wide association studies, particularly in humans, have used single-locus approaches and struggled to explain much genetic variation with markers, despite high-density panels. Given the findings shown here, there may be some advantage in revisiting these datasets and reapplying haplotype based approaches.

# Linkage disequilibrium in two divergently selected chicken lines

Anna Johansson<sup>1\*</sup>, Örjan Carlborg<sup>1</sup>

<sup>1</sup>Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences

\* presenting author, [anna.johansson@hgen.slu.se](mailto:anna.johansson@hgen.slu.se)

The pattern of linkage disequilibrium in a population affects the possibility to map genes by association mapping. We investigate how directional selection in a population of finite size affects the pattern of fixation and linkage disequilibrium, around the selected loci. We started with simulations of single locus selection and then investigated various forms of epistasis between the two loci. Using an example with an experimental chicken population selected in both directions for body weight, we found that at most previously identified QTLs, selection is strong enough to lead to fixation at the selected locus. Thus the population size has been large enough to prevent genetic drift from overriding the effect of selection at these QTLs. When interactions between two loci are present, the probability to get fixation and the pattern of linkage disequilibrium is strongly dependent on the initial haplotype frequencies. The level of linkage disequilibrium in the replicates without fixation is rather low for the studied interacting pairs from the chicken lines.

# Patterns of Linkage Disequilibrium in Cultivated Barley

Katrin MacKenzie<sup>1\*</sup>, Jordi Comadran<sup>2</sup>, Robbie Waugh<sup>2</sup>, Christine Hackett<sup>1</sup>

<sup>1</sup>BioSS, Invergowrie, United Kingdom

<sup>2</sup> Scottish Crop Research Institute, Invergowrie, United Kingdom

\* presenting author Katrin@bioss.ac.uk

We carried out a genome wide analysis of association between 2132 SNP loci across 190 elite cultivated accessions chosen to represent the available genetic variation in current North West European and North American barley germplasm. Population sub-structure and linkage disequilibrium (LD) varied considerably across the seven barley chromosomes with gene-rich and rarely-recombining haplotype blocks extending across the genetic centromeres. We use this data to test measures of population structure (none, STRUCTURE<sup>#</sup>, Kinship matrix) with association between SNP loci across the genome. Only loci with a minor allele frequency greater than 0.1 and less than 0.1 missing data were used in the analyses. We show that using information about kinship in the form of a relationship matrix decreases the number of false associations between loci and highlights blocks of short-range linkage disequilibrium.

<sup>#</sup>Pritchard,J.K., Stephens,M., Rosenberg,N.A., & Donnelly,P. Association mapping in structured populations. *Am. J. Hum. Genet.* **67**, 170-181 (2000).

## QTL analysis of vitiligo in an outbred chicken intercross

Weronica Ek<sup>1\*</sup>, Susanne Kerje<sup>3</sup>, Anna-Stina Sahlkvist<sup>2</sup>, Olle Kempe<sup>2</sup>, Olov Ekwall<sup>2</sup>, Leif Andersson<sup>3</sup>

<sup>1</sup> Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden

<sup>2</sup> Department of Medical Sciences, Uppsala University, University Hospital, 751 85 Uppsala, Sweden

<sup>3</sup> Department of Medical Biochemistry and Microbiology, Uppsala University, 751 23 Uppsala, Sweden

\* presenting author, Weronica.ek@hgen.slu.se

Vitiligo is an autoimmune disease characterized by a spontaneous pigment loss due to death of melanocytes. Smyth Line chickens have several features of this pigmentation disorder that are similar to those found in humans and are therefore a valuable animal model for the study of autoimmune vitiligo in humans. We have studied an F<sub>2</sub> cross between Smyth Line and Brown Line (control) chickens. Vitiligo was phenotyped as a binary trait, healthy or sick, and a genome-wide QTL scan was performed to identify the loci underlying the disease. We used two approaches to analyze the data. Generalized linear regression on line-origin QTL genotype probabilities calculated using cnF2freq (Nettelblad *et al*, 2009) and Flexibel Intercross Analysis (FIA), a variance component model that allows for segregation (Rönnegård *et al*, 2007). The results obtained with the two methods were highly correlated. Both identified the same significant QTL and FIA did not show any evidence of this QTL segregating within the crossed lines.

Rönnegård, L., Besnier, F., Carlborg, Ö. 2007. An improved method for quantitative trait loci detection and identification of within-line segregation in F<sub>2</sub> intercross designs. *Genetics* 178: 2315-2326.

Nettelblad, C., Holmgren, S., Crooks, L., Carlborg, Ö. 2009. cnF2freq: Efficient determination of genotype and haplotype probabilities in outbred populations using Markov models. *In press*.

## **A QTL on GGA14 – story of detection and validation.**

Siwek M<sup>1\*</sup>, Sławińska A<sup>1</sup>, Egbert F. Knol<sup>2</sup>, Witkowski A<sup>3</sup>, Bednarczyk M<sup>1</sup>.

<sup>1</sup>University of Technology and Life Sciences Mazowiecka 28, 84-085 Bydgoszcz, Poland

<sup>2</sup>IPG, Institute for Pig Genetics, P.O. Box 43, 6640 AA Beuningen, The Netherlands

<sup>3</sup>University of Agriculture, Akademicka 13, 20-950 Lublin, Poland.

\*Maria Siwek [siwek@utp.edu.pl](mailto:siwek@utp.edu.pl)

Improvement of health of chickens by selection for enhanced general resistance to pathogens is an attractive alternative for veterinary health treatments. Therefore quite often non pathogenic antigens are applied as a selection criterion to increase overall immune response. QTL to non pathogenic antigens which represent different types of immune responses have already been detected in various experiments. To verify the existence of a QTL observed in an initial genome scan, confirmation is necessary, preferably in an independent population. Keyhole Lymphet Heamocyanin (KLH) represents a novel antigen for birds, which they never encounter during lifetime, and which results in a TH-2 dependent (antibody) immune response. QTL for an adaptive response to KLH antigen on GGA14 has been originally detected in a cross of two lines selected for high and low response to SRBC. After first detection of the QTL region, this QTL has been validated in a cross of two lines selected against feather pecking. Hereby we present a second validation of this QTL for an adaptive response to KLH in a F2 cross of a commercial layer (White Leghorn) and native polish breed (Green-Legged Partridgelike). The experimental population, which consisted of 559 individuals, was typed with 7 microsatellite markers evenly spaced on GGA14. Titers of antibodies binding KLH were measured for all individuals by ELISA. Three genetic models were applied: a half-sib model (sire common parent /dam common parent), a line cross model using the regression interval method and a combined LDLA analysis. This study confirms the QTL for an adaptive response to KLH to GGA14 and makes a case for localization of genes related to immune response to this antigen.

# Genetic relatedness and population structure in analysis of linkage disequilibrium in barley

Dindo Tabanao<sup>1\*</sup>, Marco Bink<sup>2</sup>, Marcos Malosetti<sup>2</sup>, Fred van Eeuwijk<sup>2</sup>

<sup>1</sup>Philippine Rice Research Institute, Maligaya, Muñoz 3119, Nueva Ecija, Philippines

<sup>2</sup>Biometris, Wageningen University and Research Center, Netherlands

\*presenting author (dindo.tabanao@wur.nl)

Association mapping promises higher precision because it makes use of historical recombinations among diverse individuals. It also provides results that are more relevant to plant breeding applications when the mapping population is comprised of breeding materials. These populations, however, are prone to genetic heterogeneity which might lead to false association between marker polymorphism and trait variation. Molecular markers have become useful in estimating relatedness between individuals and uncovering differentiation in a population, which are valuable information by which false association could be minimized. This study aimed to compare different methods of estimating relatedness and inferring population structure in linkage disequilibrium (LD) analysis in barley. The mapping population, comprised of diverse cultivars collected from the Mediterranean area and the rest of Europe, was previously genotyped with 811 mapped DArT markers.

Relatedness was estimated using pedigree (PD) and different marker-based measures. Among six marker estimates, Queller-Goodnight (QG) showed the highest Spearman correlation with PD, though only 0.26. Of the total 16,836 pairs of cultivars, only 687 had  $PD \geq 0.05$ . Of this number showing relationship by genealogy, 1/3 was not related based on QG. In the cultivar pairs (16,149) for which  $PD < 0.05$ , more than 70% had  $QG \geq 0.05$ . No discernable improvement in  $r$  was found when comparison was made from chromosome to chromosome. Marker estimates were expected to disagree with PD, which relies on accuracy of parentage information and assumes identity between test genotypes and entries bearing their names in breeding records. Population structure was inferred by Bayesian and eigenvector clustering. Bayesian classification was performed under admixture and linkage models, which suggested five subpopulations. By eigenvector-based clustering, nine subpopulations were found. Clustering from both methods was in modest agreement with hierarchical clustering based on similarity by proportion of shared alleles. The 95th percentile of the genome-wide  $r^2$  was 0.15, below which LD across and within chromosomes was found to decay within 10 cM. LD tended to increase within subgroups as they became smaller and more homogeneous. Dissecting LD patterns within subgroups and chromosomes uncovers fine details of population structure, thereby retaining only association due to physical linkage.

## **Across-line SNP association study of innate and adaptive immune response in laying hens**

F. Biscarini\*, H. Bovenhuis\*, J. A. M. van Arendonk\*, H. K. Parmentier#, A. P. Jungerius§ and J. J. van der Poel\*

\* Animal Breeding and Genomics Centre, Wageningen University P.O. Box 338, 6700 AH Wageningen, The Netherlands

# Adaptation Physiology Group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands

§ Hendrix Genetics, Research & Technology Centre, Spoorstraat 69, 5830 AC Boxmeer, The Netherlands

Corresponding author: Filippo Biscarini.

Fax. +31 317 483929

Tel. +31 317 484262

e-mail address: [filippo.biscarini@wur.nl](mailto:filippo.biscarini@wur.nl)

The present study was aimed at detecting QTLs for innate and adaptive immunity in laying hens. A set of 1534 SNP markers was used on about 600 hens from 9 different layers lines. A novel approach based on an across-line analysis and testing of the SNP-by-line interaction was adopted. The amount of LD conserved among lines has a shorter extent than in the individual lines and therefore SNPs significantly associated with immune traits across lines are expected to be close to the functional mutations. The analysis was carried out in two consecutive steps. In the first step all SNPs were tested ignoring the relationships among animals. In the second step only SNPs with a significant and relevant effect on immunity and no significant SNP by line interaction were analysed taking into account the relationships among animals. Eventually, 59 significant associations between SNPs and immune traits were detected. This research confirmed some QTLs in regions in which QTLs have been previously identified. Potential new QTLs for immunity were detected. We found evidence for a role of the IL 17F gene on chromosome 3 on natural and acquired antibody titres and on the classical and alternative pathways of complement activation. The MHC genes on chromosome 16 showed significant effects on natural and acquired antibodies titres and classical complement activity. The IL 12 gene on chromosome 13 showed an effect on natural antibody titres.

# Simultaneous QTL detection and genomic breeding value estimation using high density SNP chips

Roel F. Veerkamp<sup>1§</sup>, Klara Verbyla<sup>1,2</sup>, Han A. Mulder<sup>1</sup>, Mario P.L.Calus<sup>1</sup>

<sup>1</sup> Animal Breeding and Genomics Centre, ASG Wageningen UR, PO Box 65, 8200 AB Lelystad, The Netherlands

<sup>2</sup>Biosciences Research Division, Department of Primary Industries Victoria, 1 Park Drive, Bundoora 3083, Australia

<sup>§</sup>Presenting author, Roel.Veerkamp@wur.nl

The simulated dataset of the 13<sup>th</sup> QTL-MAS workshop was analysed to i) detect QTL and ii) predict breeding values for animals without phenotypic information. Several parameterisations considering all SNP simultaneously were applied using Gibbs sampling. Fourteen QTL were detected at the different time points. Genetic correlations were high between models when estimating breeding values, apart from when the model assumed that all SNP effects came from one distribution. Even a model including the polygenic effect and only SNP associated with the 14 QTL gave close to unity correlations with the full parameterisations. The best parsimonious model is likely to depend on the true genetic architecture, rather than on the generic model.

# QTL detection for a medium density SNP panel: comparison of different LD and LA methods

Olivier Demeure<sup>1,2,\*</sup>, Nicola Bacciu<sup>1,2</sup>, Olivier Filangi<sup>1,2</sup>, Pascale Le Roy<sup>1,2</sup>

<sup>1</sup>INRA, UMR 598 Génétique Animale, F-35000 Rennes, France

<sup>2</sup>Agrocampus-Ouest, UMR 598 Génétique Animale, F-35000 Rennes, France

\*Corresponding author\*: olivier.demeure@rennes.inra.fr

## Background

New molecular technologies allow high throughput genotyping for QTL mapping with dense genetic maps. Therefore, the interest of linkage analysis models against linkage disequilibrium could be questioned. As these two strategies are very sensible to marker density, experimental design structures, linkage disequilibrium extent and QTL effect, we propose to investigate these parameters effects on QTL detection.

## Methods

The XIIth QTLMAS workshop simulated dataset was analysed using three linkage disequilibrium models and a linkage analysis model. Interval mapping, multivariate and interaction between QTL analyses were performed using QTLMAP.

## Results

The linkage disequilibrium models identified a large QTL on chromosome 1 and few regions with a couple of significant markers. However, most of the QTLs identified by interval mapping analysis are not detected by any linkage disequilibrium model. In addition, QTL effects are evolving during the time which was not observed using the linkage disequilibrium models.

## Conclusions

Our results show that for such a marker density the interval mapping strategy is still better than using the linkage disequilibrium only. While the experimental design structure gives a lot of power to both approaches, the marker density and informativity clearly affect linkage disequilibrium efficiency for QTL detection.

# Extensive QTL and Association Analyses of the QTLMAS2009 Data

Georgia Hadjipavlou, Gib Hemani, Richard Leach\*, Bruno Louro\*, Javad Nadaf\*, Suzanne Rowe\*, Dirk-Jan de Koning

Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, EH25 9PS, United Kingdom

\* presenting authors; contact: Suzanne.Rowe@roslin.ed.ac.uk

**Background.** We applied a range of genome-wide association (GWA) methods to map quantitative trait loci (QTL) in the simulated dataset provided by the QTLMAS2009 workshop in order to derive a comprehensive set of results.

**Results.** A Gompertz curve was modelled on the yield data and showed very good predictive properties. QTL analyses were done on the raw measurements as well on the individual parameters of the Gompertz curve and its predicted growth for each interval. Half-sib and variance component linkage analysis revealed QTL with different modes of inheritance but with low resolution. This was complemented by association studies using either single markers or haplotypes, and additive, dominance, parent-of-origin and epistatic QTL effects. All association analyses were done on phenotypes pre-corrected for pedigree effects. These methods detected QTL positions with high concordance to each other and with greater refinement of the linkage signals. Two-locus interaction analysis detected no epistatic pairs of QTL. Overall, using stringent thresholds we identified QTL regions using linkage analyses, corroborated by 6 individual SNPs with significant effects as well as two putatively imprinted SNPs.

**Conclusions.** We obtained consistent results across a combination of intra and inter family based methods using flexible linear models to evaluate a variety of models. The Gompertz curve fitted the data really well, and analyses of model parameters and predictions provided complementary information on the detected QTL.

# Bayesian multi-QTL mapping for growth curve parameters

Henri C.M. Heuven<sup>1,2\*</sup>, Luc L.G. Janss<sup>3</sup>

<sup>1</sup> Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, P.O. box 80163, 3508 TD Utrecht, The Netherlands

<sup>2</sup> Animal Breeding and Genomics Centre, Wageningen University, P.O. box 338, 6700AH Wageningen, the Netherlands

<sup>3</sup> Aarhus University, DJF Department of Genetics and Biotechnology, P.O. Box 50, 8830 Tjele, Denmark.

\*Corresponding author: [h.c.m.heuven@uu.nl](mailto:h.c.m.heuven@uu.nl)

## Background

Identification of QTL affecting a phenotype which is measured multiple times on the same experimental unit is not a trivial task because the repeated measures are not independent and in most cases show a trend in time. A complicating factor is that in most cases the mean increases non-linear with time as well as the variance. A two-step approach was used to analyze a simulated data set containing 1000 individuals with 5 measurements each. First the measurements were summarized in latent variables and subsequently a genome wide analysis was performed of these latent variables to identify segregating QTL using a Bayesian algorithm.

## Results

For each individual a logistic growth curve was fitted and three latent variables: asymptote (ASYM), inflection point (XMID) and scaling factor (SCAL) were estimated per individual. Applying an ‘animal’ model showed heritabilities of approximately 48% for ASYM and SCAL while the heritability for XMID was approximately 24%. The genome wide scan revealed four QTLs affecting ASYM, one (or two) QTL affecting XMID and four QTLs affecting SCAL. The size of the QTL differed. QTL with a larger effect could be more precisely located compared to QTL with small effect. The locations of the QTLs for separate parameters were very close in some cases and probably caused the genetic correlation observed between ASYM and XMID and SCAL respectively. None of the QTL appeared on chromosome five.

## Conclusions

Repeated observations on individuals were affected by at least nine QTLs. For most QTL a precise location could be determined. The QTL for the inflection point (XMID) was difficult to pinpoint and might actually exist of two (or more) closely linked QTL on chromosome one.

## **QTL analysis of QTLMAS2009 dataset through variance components including haplotype information and Identity-By-Descent probabilities (IBD)**

M. Hossein. Yazdi <sup>\*</sup>, Theodorus H.E. Meuwissen <sup>†</sup>, Morten Rye <sup>\*</sup>

<sup>\*</sup> Akvaforsk Genetics Center AS, N-6600 Sunndalsøra, Norway

<sup>†</sup> Department of Animal & Aquacultural Sciences,  
Norwegian University of Life Sciences, N-1432 Ås, Norway

Method of variance component for QTL mapping including haplotype information and probabilities of Identity-By-Descent (IBD) between haplotypes was implemented on dataset simulated by the framework of 13<sup>th</sup> QTLMAS workshop. Data included information on 453 markers located across 5 chromosomes for 2025 diploid individuals and phenotypic information of 5 traits on 1000 individuals in the last generation. The QTL detection was based on the visual inspection, value of log likelihood, LRT along the whole genome. We found evidence of 7 QTLs in the intervals (77-78), (43-44), (41-42), (85-86), and (2-3, 14-15, 45-46) on chromosomes 1, 2, 3, 4 and 5, respectively, affected the 5 traits recorded in this dataset. The final model for prediction of breeding values included overall mean, animal as the source of additive genetic effects and the detected QTLs which applied for each trait separately.

## **The application of Genomic Selection in dairy cattle**

A.G. de Vries, CRV BV, PO Box 454, 6800 AL Arnhem, The Netherlands

For a long time, genetic improvement of dairy cattle has relied on progeny testing of bulls. Test data from daughters are used to calculate BLUP breeding values for a large number of productive, functional and type traits. Progeny testing gives high accuracies of selection, but is expensive and slow. Usually, bulls are already 5 years old when they can be promoted to a commercial bull. By that time the investment per bull is close to 25,000 euro.

The use of DNA-markers allow earlier selection steps, and thus can lead to a more efficient breeding programme. A lot of research has been done to find QTLs to be exploited in marker-assisted breeding programmes. However, the variation explained with the small sets of markers was too limited.

A new approach is the use of high density marker sets covering the entire genome. Early 2007, CRV and the University of Liege developed a custom made 60K SNP BeadChip (Illumina) using publicly available SNPs. The marker effects were derived from a reference panel of 3600 bulls with reliable breeding values. Recent validations show that the marker effects can predict breeding values of bulls with up to 70% reliability. Selection based on these genome-wide markers is referred to as Genomic Selection.

The relatively high reliabilities with Genomic Selection offer big opportunities for cattle breeding organizations. Genotyping costs per animal are very low compared to phenotypic testing of a bull. Moreover, genotyping can be done shortly after birth. CRV is now exploiting these new opportunities after a redesign of the breeding scheme. We have doubled our selection pool of young males and females. All selection candidates animals are genotyped, and only the top 20% enter the phenotypic testing programme. With this new selection scheme, genetic improvement can be accelerated with 30-40%. Moreover, it can reduce inbreeding, as more families get a chance to contribute to the next generation.

# Marker-assisted breeding value estimation with missing marker genotypes

H.A. Mulder<sup>1\*</sup>, M.P.L. Calus<sup>1</sup>, R.F. Veerkamp<sup>1</sup>

<sup>1</sup> Animal Breeding and Genomics Centre, Animal Sciences Group, Wageningen UR, PO Box 65, 8200 AB Lelystad, The Netherlands.

\* herman.mulder@wur.nl

## Background

In livestock populations, missing genotypes on a large proportion of animals is a major problem when implementing marker-assisted breeding value estimation for QTL with a known effect. The objective of this study was to develop a method to include missing marker genotypes in breeding value estimation by predicting the number of haplotype copies (nhc) for ungenotyped animals, using 1, 2 or 4 markers. For genotyped animals the nhc represents the number of copies an animal carries for a certain haplotype, i.e. 0, 1 or 2 copies.

## Results

In a mixed model framework, the nhc were predicted for ungenotyped animals using the additive genetic relationship matrix and the nhc of genotyped animals. The heritability of nhc is assumed to be 0.99, allowing for minor genotyping and haplotyping errors. The predicted nhc were subsequently used as covariables in marker-assisted breeding value estimation by applying a random regression on them. To evaluate the method, a population was simulated with one additive QTL and an additive polygenic genetic effect. The QTL was located in the middle of a haplotype based on SNP-markers. The accuracy of the total EBV increased for genotyped animals comparing marker-assisted breeding value estimation with conventional breeding value estimation, but for ungenotyped animals the increase was marginal unless the heritability was smaller than 0.1. Haplotypes based on 4 markers yielded the highest accuracies and using only the left nearest marker yielded the lowest accuracy. The accuracy of the method increased when the marker density increased. The accuracy of the total EBV approached the accuracy of gene-assisted BLUP when using 4-marker haplotypes with a distance of 0.1 cM between the markers.

## Conclusion

The proposed method is computationally very efficient and suitable to apply for marker-assisted breeding value estimation in large livestock and plant populations including effects of a number of known QTL.

## Response and inbreeding from genomic selection

Jack C. M. Dekkers<sup>1\*</sup>, Hong-hua Zhao<sup>1,2</sup>, Jennifer M. Young<sup>1</sup>, David Habier<sup>1,3</sup> and Rohan L. Fernando<sup>1</sup>

<sup>1</sup> Department of Animal Science, Iowa State University, Ames, USA

<sup>2</sup> Pioneer Hi-Bred Int., Johnston, Iowa, USA

<sup>3</sup> Institute of Animal Breeding and Husbandry, Christian-Albrechts University of Kiel, Germany

\* presenting author [jdekkers@iastate.edu](mailto:jdekkers@iastate.edu)

Genomic Selection (GS) using breeding values (GS-EBV) estimated from dense marker data is promising for genetic improvement. Our objective was to evaluate responses from GS to selection on traditional BLUP-EBV over multiple generations. A trait with 100 or 200 QTL with heritability 0.3 and phenotyping prior to selection was simulated. GS-EBV were estimated using Bayes-B using 1000 individuals from the training generation only (GS-1) or with updating using data from all generations (GS-all). BLUP-EBV used data from all generations (BLUP-all). Response for GS-1 was similar to BLUP-all in initial generations but then fell behind because of declines in accuracy. GS-all had greater response than BLUP-all. Doubling the number of QTL increased response, in particular for GS-all. Rates of inbreeding increased from GS-1 to GS-all and BLUP-all. GS-all resulted in the fastest loss in variance, followed by BLUP-all and GS-1. Lost variance was to a greater degree from drift and loss of favorable QTL alleles for BLUP-all than GS. Despite lower inbreeding, GS had greater variance of response than BLUP-all because of variation in the accuracy of GS-EBV. Deterministic predictions of response were similar to those observed from simulation in the initial generations. Deterministic predictions of rates of inbreeding were similar to observed rates, but observed rates were higher in generations immediately following training because of the impact of relationships on GS-EBV. In conclusion, GS offers great opportunities to further advance genetic improvement programs. GS, as simulated here, primarily capitalizes on historic linkage disequilibrium.

SPONSORS:



LEB fonds



Diagnostics



S A B R E CUTTING EDGE GENOMICS FOR SUSTAINABLE ANIMAL BREEDING 